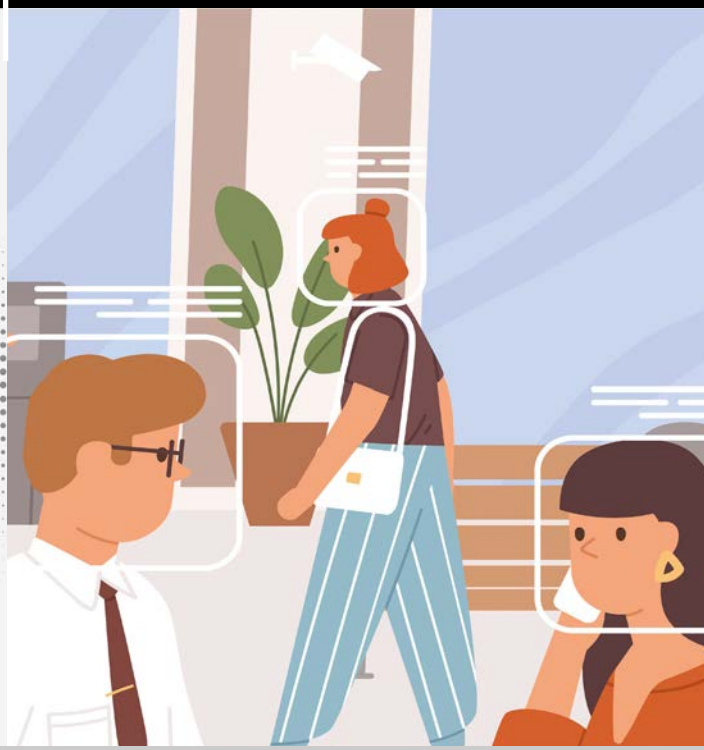

Murat Karaboga, Nula Frei, Frank Ebbers, Sophia Rovelli, Michael Friedewald, Greta Runge

Automatisierte Erkennung von Stimme, Sprache und Gesicht

Technische, rechtliche und gesellschaftliche Herausforderungen

v/d/f

TA-SWISS 79/2022

Brunngasse 36
CH-3011 Bern
www.ta-swiss.ch

Murat Karaboga, Nula Frei, Frank Ebbers, Sophia Rovelli, Michael Friedewald, Greta Runge

Automatisierte Erkennung von Stimme, Sprache und Gesicht

Technische, rechtliche und gesellschaftliche Herausforderungen

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.dnb.de abrufbar.

Zitiervorschlag
Karaboga M., Frei N., Ebbers F., Rovelli S., Friedewald M., Runge G. (2022): Automatisierte Erkennung von Stimme, Sprache und Gesicht. Technische, rechtliche und gesellschaftliche Herausforderungen. In TA-SWISS Publikationsreihe (Hrsg.): TA 79/2022. Zürich: vdf.

Coverabbildungen:
© Links: Adobe Stock / eriksvoboda
© Rechts: Adobe Stock / Good Studio

ISBN 978-3-7281-4140-8 (Printausgabe)

Download open access:
ISBN 978-3-7281-4141-5 / DOI 10.3218/4141-5

www.vdf.ethz.ch
verlag@vdf.ethz.ch

Dank

Die Autorinnen und Autoren danken TA-SWISS für die gewährte Förderung des Projekts «Automatisierte Erkennung von Stimme, Sprache und Gesicht: Technische, rechtliche und gesellschaftliche Herausforderungen». Ein ganz besonderer Dank gilt Prof. Dr. Christina Tobler, die uns vom Beginn des Projekts unter Pandemie-Bedingungen 2020 bis Ende 2021 als Projektverantwortliche bei TA-SWISS mit Rat und Unterstützung zur Seite stand. Dieser Dank gebührt auch Dr. Laetitia Ramelet, die diese Rolle nachfolgend seit Anfang 2022 ausfüllte und wesentlich bei der Finalisierung der Studie mitgewirkt hat.

Den Mitgliedern der Begleitgruppe danken wir für die durchgängige Unterstützung der Studie, deren Inhalte von der stets konstruktiv-kritischen inhaltlichen Auseinandersetzung profitiert haben. Dieser Dank gilt insbesondere Bruno Baeriswyl, dem Leiter der Begleitgruppe.

Des Weiteren möchten wir allen Personen und Institutionen danken, die mittels Interviews oder per schriftlicher Befragung an der Studie mitgewirkt haben, dies sind insbesondere die Kantonspolizei St. Gallen, die Kantonspolizei Aargau sowie PostFinance AG. Wir hoffen, dass die Studienergebnisse auch für sie Relevanz haben.

Die Bevölkerungsumfrage wurde seitens IPSOS SA durchgeführt. Wir danken unseren dortigen Ansprechpartnerinnen Hana Baronijan, Celine Perroud und Heloise Fortier, die uns stets kompetent unterstützend zur Seite standen und die erfolgreiche Durchführung der Umfrage gewährleisteten. Auch den Teilnehmerinnen und Teilnehmern der Fokusgruppen danken wir für ihre Bereitschaft, an den Diskussionen mitzuwirken.

Bei verschiedenen methodischen Schritten wurde das Projektteam von studentischen Mitarbeiterinnen und Mitarbeitern unterstützt. Diese waren Lelaina Seelbach und Kennedy Nwankwo. Auch Ihnen danken wir für die Unterstützung.

Abschliessend soll nicht unerwähnt bleiben, dass das grosse Engagement der Autorinnen und Autoren zur erfolgreichen Fertigstellung der Studie beigetragen hat.

Karlsruhe und Freiburg i. Ue., August 2022

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Zusammenfassung

Die Entwicklung von Stimm-, Sprach- und Gesichtserkennungstechnologien hat in den letzten Jahren bedeutende Fortschritte erlebt. Eine Vielzahl darauf basierender Anwendungen drängt vermehrt auf den Markt und wird sowohl von Unternehmen als auch Behörden verwendet. Einerseits sollen innovative Anwendungen Nutzen für den Einzelnen und die Gesellschaft mit sich bringen. Demnach können sie bspw. dabei helfen, Nutzerinnen und Nutzer bei alltäglichen Aufgaben zu unterstützen, Krankheiten frühzeitig zu diagnostizieren und bei polizeilichen Ermittlungen schnell grosse Bilddatenbanken nach Übereinstimmungen zu durchsuchen. Andererseits war in den letzten Jahren insb. der polizeiliche Einsatz von Gesichtserkennung aufgrund der Befürchtung einer anlasslosen Massenüberwachung mit zahlreichen ablehnenden zivilgesellschaftlichen Kampagnen konfrontiert. Aber auch andere Anwendungsmöglichkeiten sind umstritten, etwa weil deren Vorhersagen als technisch fehlerhaft oder gesellschaftlich unerwünscht gelten, so namentlich die Möglichkeit, von äusserlich wahrnehmbaren körperlichen Merkmalen auf andere Eigenschaften einer Person (Geschlecht, sexuelle Neigung, Gesundheitszustand, Emotionen usw.) zu schliessen.

Trotz der technologischen Fortschritte befinden sich die verschiedenen stimm-, sprach- und gesichtserkennungsbasierten Anwendungen jedoch noch in einem frühen Stadium: Häufig ist weder klar, wie technisch zuverlässig sie sind, noch ob sie rechtlichen Anforderungen genügen und welche gesellschaftlichen Herausforderungen sich durch ihren Einsatz ggf. stellen könnten.

Vor dem Hintergrund dieser Unübersichtlichkeit und der zunehmenden gesellschaftlichen Diskussionen ist es das Ziel der vorliegenden Studie, fundiertes Orientierungswissen zum Umgang mit den Technologien bereitzustellen, um den gesellschaftlichen und politischen Diskurs über Chancen und Herausforderungen von Stimm-, Sprach- und Gesichtserkennungstechnologien zu befördern.

Untersuchung von ausgewählten Anwendungsfeldern

Im Zentrum der Studie steht die Untersuchung von acht exemplarischen Anwendungsfeldern:

1. **Smarte Lautsprecher** sind auch in Schweizer Haushalten immer häufiger anzutreffen. Dadurch, dass smarte Lautsprecher als digitale Assistenten fungieren sollen, werden sie in aller Regel in Wohnzimmern und damit im Mittelpunkt des Privatlebens vieler Menschen aufgestellt.

2. **Gesichtserkennung durch polizeiliche Stellen** kommt in der Schweiz seit einigen Jahren zum Einsatz. Sie dient als technisches Unterstützungswerkzeug und wird ex post eingesetzt, um bei der Fahndung nach Personen Aufnahmen vom Tatort mit Fotos abzugleichen, die in Polizeidatenbanken gespeichert sind. In einigen Staaten kommt auch Gesichtserkennung in Echtzeit zum Einsatz. Dabei werden die Aufnahmen von

Überwachungskameras, die zumeist an öffentlichen Plätzen, Bahnhöfen usw. aufgebaut sind, laufend mit Polizei-Datenbanken abgeglichen, um nach Vermissten, Flüchtigen etc. zu fahnden.

3. **Authentifizierung via Stimme** wird in der Schweiz von einigen Banken eingesetzt, um beim Telefonbanking die Authentifizierung mittels Abfrage personenbezogener Daten durch die biometrische Erkennung der Stimme des Kunden oder der Kundin zu ersetzen.

4. **Gewaltprävention und -aufklärung in Sportstadien** mithilfe der Gesichtserkennung ist eines der seit Jahrzehnten besonders kontroversen Themen in Debatten über Gesichtserkennung. In der Schweiz scheiterten mehrere Anläufe zu ihrer Einführung, doch kommt sie inzwischen weltweit in vielen Staaten zum Einsatz.

5. **Erkennung physischer und psychischer Krankheiten** aus Aufnahmen der Stimme oder des Gesichts befindet sich derzeit noch grösstenteils in Entwicklung, doch erste Applikationen kommen bereits auf den Markt. Mit ihnen soll die (Früh-)Erkennung von verschiedenen Krankheiten oder Beeinträchtigungen, wie z.B. Parkinson, Alzheimer, Autismus oder Depressionen, möglich werden. Zudem sollen die Anwendungen auch unterstützend bei der Therapierung zum Einsatz kommen können.

6. **Emotionserkennung** auf Basis von Stimm-, Sprach- und Gesichtsaufnahmen wird über die Personenidentifikation bzw. -authentifikation hinaus vermehrt als zusätzliche Funktion angeboten. Insb. im Bereich Marketing und in Bewerbungsverfahren auf dem Arbeitsmarkt wird der Emotionserkennung grosses Potenzial beigemessen, weil sich die Betreiber Erkenntnisse über die «wahren» Wünsche, Fähigkeiten usw. der erfassten Menschen versprechen.

7. **Aufmerksamkeitsanalyse** ist ein weiteres Einsatzgebiet der Gesichtserkennung. Diese findet in der Schweiz zwar noch keine Anwendung, wird aber in anderen Ländern, etwa zur Erfassung und Bewertung der Aufmerksamkeit von Autofahrern oder von Schülerinnen und Schülern, bereits verwendet.

8. **Jedermann-Identifikation** bezeichnet die heute noch fiktive Möglichkeit, per Miniaturkameras, die bspw. in Datenbrillen eingebaut sind, alle Personen im Sichtfeld in Echtzeit zu identifizieren und zusätzliche (sensible) Informationen über sie bereitzustellen.

Diese Anwendungsfelder repräsentieren einen Querschnitt über heute **bereits im Einsatz befindliche Anwendungen** (*Smarte Lautsprecher*, *Gesichtserkennung durch Polizei*, *Authentifizierung via Stimme*), über solche, die **mittelfristig Anwendung finden könnten** (*Gewaltprävention und -aufklärung in Sportstadien*, *Erkennung physischer und psychischer Krankheiten* bzw. *Emotionserkennung*) bis hin zu aus heutiger Sicht **eher hypothetischen Anwendungsszenarien**, die jedoch weitreichende Implikationen mit sich brächten (*Aufmerksamkeitsanalyse* und *Jedermann-Identifikation*). Die Untersuchung der genannten Anwendungsfelder von Stimm-, Sprach- und Gesichtserkennung erfolgt im Hinblick auf **vier zentrale Fragekomplexe**:

- Analyse der technischen Grundlagen und Möglichkeiten
- Juristische Bewertung insb. der datenschutzrechtlichen Rahmenbedingungen

- Erörterung gesellschaftlicher und ethischer Herausforderungen:
- Untersuchung der gesellschaftlichen Wahrnehmung der Chancen, Risiken und Wünsche

Aktuelle technische Grundlagen und Möglichkeiten

Die Untersuchung der technischen Grundlagen und Möglichkeiten zeigt **ein weites Spektrum der technologischen Reife**: Heutige Anwendungen funktionieren entweder bereits technisch weitgehend zuverlässig (*Smarte Lautsprecher*, *Authentifizierung via Stimme*) oder könnten im Ergebnis der aktuellen Entwicklungsbemühungen in den nächsten Jahren eine **weitgehende technische Zuverlässigkeit erreichen** (*ex post Gesichtserkennung durch polizeiliche Stellen*, *Gewaltprävention und -aufklärung in Sportstadien*). Anwendungen in den Bereichen der *Erkennung von Krankheiten* sowie der *Emotionserkennung* sind hingegen mit teils grösseren Herausforderungen konfrontiert, **die einen zuverlässigen Einsatz behindern**. Erforderlich sind in diesen Bereichen intensive Bemühungen hins. der konzeptionellen Grundlagen sowie der Verbesserung der zugrunde liegenden Datenbestände.

Im Hinblick auf die **Erkennungsraten** von Stimm-, Sprach- und Gesichtserkennung ist v.a. der soziotechnische Kontext zu berücksichtigen, in dem eine Erkennung erfolgt: Beispielsweise verspricht fortschrittliche Kamerasensorik in einer kontrollierten Umgebung, die gut beleuchtet ist und in der sich eine Person kooperativ verhält, weitaus höhere Trefferraten als die Erkennung des Gesichts eines sich bewegenden Menschen, das bei schlechter Beleuchtung bzw. mittels schlechter Sensorik aus der Distanz aufgezeichnet wurde. Insofern kommt der Bewertung der technisch-organisatorischen Vertrauenswürdigkeit einer Anwendung eine zentrale Rolle zu. Hierzu wäre insb. eine **unabhängige Überprüfung der herstellerseitig verbreiteten Angaben zu Trefferraten der Software** erforderlich. Doch gerade diese Bewertung fällt v.a. in jenen Fällen schwer, in welchen es sich bei den für einen Technologieeinsatz Verantwortlichen um private Akteure handelt. Denn abgesehen von den instruktiven Studien der National Institute of Standards and Technology (NIST) sind lediglich vereinzelte wissenschaftliche Untersuchungen von Erkennungsalgorithmen bzw. ihren Trefferraten vorhanden. Diese werden allerdings in aller Regel dadurch begrenzt, dass die Untersuchungen in experimentellen Testumgebungen stattfinden und der tatsächliche soziotechnische Nutzungskontext unberücksichtigt bleibt, womit sie auch nur in beschränktem Masse auf den Praxiseinsatz übertragbar sind, der durch variierende Sensorik und Umweltbedingungen bestimmt ist. Insofern **verbergen sich die realen Trefferraten der Anwendungen hinter einer Gemengelage von Marketingversprechen, Betriebsgeheimnissen und Sicherheitserfordernissen**. Die technische Zuverlässigkeit der Systeme wird ausserdem durch die **Bias-Problematik** beeinträchtigt: Weil stimm-, sprach- und gesichtserkennungsbasierte Anwendungen mit von Menschen stammenden und aufbereiteten Daten sowie mit menschengemachten Algorithmen und Prozessen arbeiten, werden sie stets einen Bias beinhalten, der reduziert, aber voraussichtlich niemals vollständig ausgeräumt werden kann.

Ein grundsätzlicheres Problem liegt bei jenen Fällen vor, in denen wissenschaftlich fundierte Kategorien fehlen, was teils auf die Krankheitserkennung, aber insb. die Emotionserken-

nung zutrifft. Dadurch, dass Emotionen nicht klar voneinander abgrenzbar – und damit klar kategorisierbar – sind und stark in der Intensität variieren, ist die Berechnung von Trefferquoten problematisch, da es sich bei allen kategorialen Zuordnungen um Wahrscheinlichkeitswerte in Relation zur jeweiligen Kategorie handelt. Daher ist weitere Forschung hinsichltich der konzeptionellen Grundlagen in diesen Bereichen erforderlich.

Im Ergebnis der Analyse der technischen Grundlagen und Möglichkeiten ist festzuhalten, dass auch in der Zukunft selbst die zuverlässigsten auf Stimm-, Sprach- und Gesichtserkennungstechnologien basierenden Systeme **keine 100-prozentige technische Zuverlässigkeit** erzielen werden.

Rechtliche Rahmenbedingungen

Das Gesicht und die Stimme einer Person stellen in den allermeisten der hier untersuchten Fälle **biometrische Daten** und damit **besonders schützenswerte Personendaten** dar. Es handelt sich um eng und dauerhaft mit der Person verbundene Merkmale, die einzigartige Aspekte der Persönlichkeit darstellen. Die Bearbeitung dieser Daten stellt in jedem Fall eine Persönlichkeitsverletzung bzw. einen schweren Eingriff in Grundrechte Einzelner dar.

Bei der Nutzung von Stimm-, Sprach- oder Gesichtserkennungstechnologien durch **Behörden** sind das **Recht auf Privatsphäre sowie der Schutz vor Missbrauch der persönlichen Daten** (Art. 13 Abs. 1 und 2 BV) betroffen. Auf die Ausübung verschiedener Kommunikationsgrundrechte (Meinungsäusserungsfreiheit, Versammlungs- und Vereinigungsfreiheit, Art. 16, 22, 23 BV) kann die Verwendung dieser Technologie einen abschreckenden Effekt (*chilling effect*) haben. Ebenfalls betroffen sein können die Verfahrensgrundrechte (Art. 29, 29a, 30 BV) sowie das Verbot der Diskriminierung (Art. 8 Abs. 2 BV). Ein grundrechtskonformer Einsatz dieser Technologien bedingt eine sorgfältige vorgängige Prüfung durch den Gesetzgeber, der gehalten ist, die für den Einsatz notwendige **gesetzliche Grundlage**, in der Regel in einem **Gesetz im formellen Sinne**, zu schaffen und diese **präzise genug auszugestalten**, sodass daraus nicht nur ein eindeutiger Zweck erkennbar ist, sondern u.a. auch eine **Beschränkung** der Datenbearbeitung auf das absolut Notwendige, sowohl in zeitlicher wie umfangmässiger Hinsicht. Die **Verhältnismässigkeit** des Einsatzes dieser Technologien muss sowohl bei der Schaffung einer gesetzlichen Grundlage wie auch vor dem konkreten Einsatz durch die rechtsanwendenden Behörden geprüft werden. Dabei ist nach dem aktuellen Stand der Technik die **Eignung** des Einsatzes dieser Technologien häufig infrage zu stellen, denn wie die vorliegende Studie gezeigt hat, ist die **Zuverlässigkeit der Technologie sehr häufig nicht gegeben** (insb. bei der Emotionserkennung). Häufig dürfte auch die **Erforderlichkeit** nicht gegeben sein, da mildere Mittel vorhanden sind, um die mit dem Einsatz von Stimm-, Sprach- oder Gesichtserkennungstechnologien verfolgten öffentlichen Interessen zu erreichen. Gewisse Anwendungen würden zudem den absolut geschützten **Kernbereich** mehrerer Grundrechte verletzen und sind somit immer unzulässig. Dazu gehört insb. der Einsatz von Gesichtserkennungstechnologie zur **flächendeckenden Echtzeitüberwachung** sowie die Durchführung von **Emotionserkennung** durch den Staat.

Wird Stimm-, Sprach- oder Gesichtserkennungstechnologie durch **Private** genutzt, werden in der Regel mehrere Datenschutzgrundsätze verletzt, was die Datenbearbeitung gemäss

Art. 31 des neuen Datenschutzgesetzes (nDSG) rechtfertigungsbedürftig macht. Ein überwiegendes Interesse des Bearbeiters als Rechtfertigungsgrund dürfte in den meisten Fällen aufgrund der besonderen Schutzbedürftigkeit der bearbeiteten biometrischen Daten ausser Betracht fallen. Die **Einwilligung** der Betroffenen ist aber im Bereich der Stimm-, Sprach- und Gesichtserkennung häufig problematisch, denn sie ist entweder **praktisch unmöglich** zu erteilen bzw. einzuholen (etwa bei smarten Lautsprechern oder Datenbrillen und anderen Jedermann-Anwendungen), aufgrund bestehender **Machtasymmetrien** (in Bewerbungsverfahren, bei Versicherungen oder in Privatschulen), oder mangels datensparsamer Angebote (beim Einsatz in Sportstadien oder durch Banken) **nicht freiwillig** oder aber nicht nach **angemessener Information** erfolgt.

Da der Einsatz durch private Akteure zahlreiche Risiken für die Betroffenen birgt und in den untersuchten Anwendungsfeldern von den Anwendern resp. Betreibern vielfach nicht in rechtskonformer Weise agiert wird, entsteht aus grundrechtlicher Perspektive eine **staatliche Schutzpflicht vor Grundrechtsverletzungen durch Private** (Art. 35 BV). Diese Schutzpflicht sorgt dafür, dass der Staat die **Verwendung dieser Technologie durch Private regulieren und die Einhaltung der Regelungen überwachen** sollte. Das Projektteam legt entsprechende Empfehlungen vor.

Aus datenschutzrechtlicher Sicht ist abschliessend auch der Grundsatz der **Datensicherheit** zu nennen, der sowohl von Behörden wie von Privaten beachtet werden muss. Die Gefahren bei einem Verlust, Missbrauch oder Zweckentfremdung biometrischer Daten durch unberechtigte Personen sind gravierend, entsprechend hohe Anforderungen sind an die Massnahmen zur Sicherstellung der Datensicherheit zu legen.

Ethische und gesellschaftliche Herausforderungen

Selbst wenn alle (datenschutz-)rechtlichen Anforderungen erfüllt sind, können Anwendungen der Stimm-, Sprach- und Gesichtserkennung erhebliche weitere ethische Herausforderungen mit sich bringen. Dazu zählen **Abschreckungseffekte der Überwachung**, durch die die Wahrnehmung von Grundrechten beeinträchtigt wird. Und selbst wenn bspw. der Betrieb polizeilicher Gesichtserkennung rechtskonform erfolgt, besteht immer noch die Möglichkeit einer **schleichenden Einführung einer umfassenden staatlichen Überwachung** in immer mehr Bereichen des Lebens. Verhaltensänderungen bzw. internalisierter Konformismus könnten aber auch in anderen Bereichen resultieren. Insb. bei bestimmten Formen der Emotionserkennung wären Betroffene mit der Herausforderung konfrontiert, sich entweder dem Erkennungsdruck zu beugen und sich möglichst regelkonform bzw. vorteilhaft zu verhalten – hilft ein aufgesetztes Lächeln? – oder Nachteile in Kauf zu nehmen.

Weil davon auszugehen ist, dass eine 100-prozentige technische Zuverlässigkeit nicht möglich sein wird, ist zu erwarten, dass trotz verbesserter Erkennungsraten auch in der Zukunft **falsch-positive und falsch-negative Treffer** eintreten und **unerwünschte Konsequenzen** nach sich ziehen werden. Diese reichen von der ungerechtfertigten Verhaftung im Falle polizeilicher Gesichtserkennung über die Gefahr des illegalen Kontozugriffs beim Telefonbanking bis hin zu falsch-negativen (Selbst-)Diagnosen im medizinischen Bereich, die zu lebensgefährlichen Folgen führen könnten.

Schon heute verfügen staatliche und wirtschaftliche Stellen über weitaus mehr Informationen über die Bevölkerung als noch vor wenigen Jahren und Jahrzehnten. In der Diskussion der ethischen Herausforderungen wurde bei mehreren Anwendungsfeldern deutlich, dass sich dieses **Wissens- und** damit auch **Machtgefälle** mit dem zunehmenden Einsatz von Stimm-, Sprach- und Gesichtserkennungstechnologien voraussichtlich weiter zugunsten der Datenbearbeiter verschieben wird. (Datenschutz-)Gesetze bieten zwar Schutz vor allzu invasiven Missbrauchsmöglichkeiten dieser zunehmenden Macht im Einzelfall. Allerdings vermögen Datenschutzgesetze die Entwicklungen an sich nicht aufzuhalten. Zudem bleiben **zahlreiche Möglichkeiten der Manipulation individuellen Verhaltens** trotzdem erhalten: Prädiktive Modelle (Menschen, die A und B tun, tun/denken/fühlen mit hoher Wahrscheinlichkeit auch X) basieren bspw. auf der Auswertung anonym(isiert)er Daten anderer Menschen, können aber unter Hinzuziehung personenbezogener Daten einer bestimmten Person dazu verwendet werden, sensible Schlüsse über diese zu ziehen.

Im Rahmen des geltenden Rechts stellt gerade die **datenschutzrechtliche Einwilligung** eine grosse Herausforderung dar, weil sie in Situationen extremer Machtdifferenz die Illusion von Kontrolle erzeugen kann. Während im Falle der Erkennung eines Gesichts, der Interpretation von Emotionen und Verhalten usw. durch einen Menschen recht klar ist, wie diese vom Gegenüber aufgefasst werden, erlauben Stimm-, Sprach- und Gesichtserkennungstechnologien die **Gewinnung sehr weitreichender Informationen über die Betroffenen**. Menschen, die sich nicht im Klaren über dieses Potenzial sind, könnten **leichtfertig in die Bearbeitung ihrer Daten einwilligen**. Im Falle der Emotionserkennung könnten die Betroffenen im Glauben an eine fehlerbehaftete Technik in die Nutzung einwilligen und versuchen, diese durch Verhaltensanpassung zu den eigenen Gunsten zu überlisten, obwohl dies durchschaut und in Nachteilen für die Betroffenen resultieren würde. Zugleich verdeutlicht das Beispiel der Emotionserkennung: Ebenso, wie eine technisch fehlerhafte Technologie Gefahren mit sich bringt, drohen **ernst zu nehmende Herausforderungen auch im Falle eines technisch weitgehend fehlerfreien Betriebs** der Technologie. Gefahren wie eine staatliche Massenüberwachung oder die als Jedermann-Identifikation diskutierte Möglichkeit der Identifikation von fremden Menschen «im Vorbeigehen» würden sich unter den Bedingungen einer perfektionierten Technik mit besonderer Dringlichkeit ergeben.

Mittel- bis langfristig planen Technologiehersteller zudem die weitere Überführung von Stimm, Sprach- und Gesichtserkennungstechnologien in **konvergierende soziotechnische Systeme**, etwa den Rückgriff auf Emotionserkennung beim Telefonbanking und bei Bewerbungsgesprächen oder deren Einbettung in smarte Lautsprecher. Im Ergebnis dieser Entwicklung würde die Quantität und v.a. die Qualität der über das Verhalten der Betroffenen gesammelten Daten stark zunehmen. Tendenzen der Quantifizierung und Steuerung menschlichen Verhaltens und Lebens würden so bis in die letzten noch nicht digitalisierten Bereiche vordringen. Als unterstützendes Mittel würden derartige fortgeschrittene insbesondere Assistenz- oder Diagnosesysteme einen grossen alltäglichen Nutzen mit sich bringen. Andererseits würde das menschliche Leben im Ergebnis technologischer Feedback-Schleifen zunehmend von den Berechnungen und Routinen der Technologie gesteuert. Sofern eine demokratische Kontrolle der zugrunde liegenden Algorithmen vorhanden wäre, könnte dies die gesellschaftliche und individuelle Selbstbestimmung durchaus steigern. Falls eine intransparente algorithmische Steuerung erfolgte, könnte Selbstbestimmung hingegen zu

einer Farce werden, weil Entscheidungen von den grossen Fragen bis ins Privateste von diesen Algorithmen dominiert würden.

Eine weitere zentrale Herausforderung hängt mit den vielfach als grosser Vorteil diskutierten Effizienzgewinnen durch automatisierte Stimm-, Sprach- und Gesichtserkennung zusammen. Hier stellt sich die Frage, ob die möglich werdenden Effizienzgewinne in einer gesteigerten Qualität polizeilicher Arbeit, vergünstigten Tickets in Sportstadien oder der verbesserten Patientenversorgung münden oder ob ökonomische Rationalisierungseffekte zu **Personaleinsparungen und der einseitigen Verteilung etwaiger Gewinne** führen werden.

Die Erörterungen der ethischen Herausforderungen verweisen insb. auf einen **dringenden gesellschaftlichen Diskussionsbedarf**, der sich schlussendlich um die Frage dreht, wie die möglichen Potenziale der Technologieentwicklung mit einem demokratischen Politsystem unter Berücksichtigung der Grundrechte zusammengebracht werden können.

Öffentliche Wahrnehmung

Im Rahmen der Analyse der öffentlichen Wahrnehmung wurde klar, dass die Bürgerinnen und Bürger die **Chancen der Technologien** durchaus erkennen und zu schätzen wissen. Als Vorteil erkannten sie den erwarteten **Gewinn an Sicherheit, Gesundheit** und **Komfort**. Sie erwarten aber auch **Effizienzsteigerungen**, sodass freiwerdende personelle oder finanzielle Ressourcen anderweitig sinnvoll eingesetzt werden könnten. Eine nicht unerhebliche Rolle für die positive Bewertung spielt auch die nicht unbedingt zutreffende Wahrnehmung, dass der Einsatz der Technologien menschengemachte Diskriminierung reduzieren könnte und dass es sich bei den für den Einsatz verantwortlichen Akteuren um vertrauenswürdige Stellen handle.

Zugleich wird den Technologien und insb. deren Betreibern **misstraut**: Dabei überwiegt die **Furcht vor einem intransparenten Technologieeinsatz** und dem **Missbrauch personenbezogener Daten** wie dem Abgreifen von Kontozugangsdaten oder der anlasslosen Massenüberwachung. Daneben befürchten die Befragten, dass Stimm-, Sprach- und Gesichtserkennungstechnologien nicht zuverlässig genug funktionieren könnten und dadurch unerwünschte Folgen nach sich ziehen könnten, wie etwa eine Kontrolle oder Verhaftung infolge einer **Fehlerkennung** oder eine fatale **Fehldiagnose** bei medizinischen Anwendungen. Schliesslich zeigten sich viele Befragte darüber besorgt, dass die Technologien auch gegen ihre Interessen (unerwünschte Werbung, Preisdiskriminierung etc.) verwendet werden können.

Die Untersuchung der Bevölkerungsmeinung zeigt aber auch erhebliche **Wissenslücken** und einen damit zusammenhängenden **Informationsbedarf**. Dies spiegelte sich auch in der Befürwortung bzw. Ablehnung unterschiedlicher Anwendungen wider: Die Jedermann-Identifikation und die Aufmerksamkeitsanalyse in Schulen werden klar abgelehnt, der Einsatz von Gesichtserkennung in Sportstadien klar befürwortet. Bei allen anderen Anwendungen ist hingegen der **hohe Anteil der Unentschlossenen** auffällig und dies unterstreicht die Wichtigkeit einer öffentlichen Debatte zu den diskutierten Themen.

Zur Gewährleistung eines vertrauenswürdigen Einsatzes von Stimm-, Sprach- und Gesichtserkennungstechnologien sehen Bürgerinnen und Bürger den Staat und die Dienstebetreiber in der Pflicht. Verwiesen wurde dabei auf die **Notwendigkeit von weitgehenden Transparenzregelungen, Verboten, Aufklärungskampagnen, die Begleitung und Evaluation der Technologieeinsätze** durch unabhängige Experten, Vorabtestungen auf Wirksamkeit, eine Zertifizierungspflicht und dass der staatliche Einsatz auf möglichst konkreten gesetzlichen Grundlagen fussen müsse.

Empfehlungen

Die Studienergebnisse machen deutlich, dass **Handlungsbedarf** besteht. Wir formulieren Möglichkeiten, wie Entscheidungstragende aus Politik, Wirtschaft, Datenschutzbehörden, Medien und Zivilgesellschaft den unterschiedlichen Herausforderungen der Stimm-, Sprach- und Gesichtserkennungstechnologien begegnen könnten. Auf Basis der Erkenntnisse der vorliegenden Studie scheint insb. ein **Verbot** in drei Bereichen und ein Moratorium in einem Bereich angebracht:

- **Verbot automatisierter staatlicher Echtzeitüberwachung und staatlichen Social-Scorings mittels Stimm-, Sprach- oder Gesichtserkennung.**
- **Verbot vollständig automatisierter Entscheidungen gestützt auf Stimm-, Sprach- und Gesichtserkennungstechnologien in wichtigen Lebensbereichen** (z.B. in den Bereichen Gesundheit, Strafverfolgung, Finanzen und Kreditvergabe, Versicherungen, Arbeitsumfeld). Die Ergebnisse von teilautomatisierten Entscheidungsunterstützungssystemen sollten stattdessen von geschultem Personal kritisch überprüft und freigegeben werden müssen.
- **Verbot der Nutzung von Datenbrillen** und anderen nicht direkt erkennbaren Technologien, die sich zur Überwachung eignen, wie kleinen Kameras, sofern sie mit Gesichtserkennungstechnologie verknüpft sind, **in der Öffentlichkeit.**
- **Moratorium für Emotions- und Krankheitserkennung in wichtigen Lebensbereichen** (z.B. in den Bereichen Strafverfolgung, Finanzen und Kreditvergabe, Versicherungen, Schul- und Arbeitsumfeld), solange die genügende technische und organisatorische Zuverlässigkeit und Fairness nicht erwiesen ist. Insb. Verbot des Einsatzes von Emotions- und Krankheitserkennung bei der Stimmauthentifizierung und Bewerberauswahl durch Private sowie Verbot der Aufmerksamkeitserkennung an Schulen. Bestimmte erwünschte Anwendungen könnten auch als Hochrisikoanwendung klassifiziert und damit unter Auflagen erlaubt sein.

Zudem weisen wir darauf hin, dass **biometrische Merkmale und insb. die Stimme nicht als (alleiniger) Authentifizierungsfaktor verwendet werden sollten**. Denn anders als veränderbare Authentifizierungsmerkmale (etwa Passwörter) können körperliche Merkmale nicht geändert werden, wenn sie einmal kompromittiert wurden.

Die weiteren Empfehlungen unterteilen sich in anwendungsfeldspezifische sowie allgemeine, anwendungsfeldübergreifende Empfehlungen. Letztere sind:

- **Die Regulierung von Hochrisikoanwendungen**, insb. in den Bereichen Gesundheit, Strafverfolgung, Kreditvergabe, Versicherungen, Arbeitsumfeld, ist geboten, weil sie ein hohes Risiko für Grundrechte und das gesellschaftliche Zusammenleben bergen. Die Regulierung sollte ein Risikomanagement, darunter insb. die unabhängige Evaluation der technisch-organisatorischen Zuverlässigkeit, Dokumentationspflichten, eine Pflicht zur menschlichen Beaufsichtigung, Datensicherheitsvorgaben sowie unabhängige Kontrollmechanismen beinhalten.
- **Ausdrückliche gesetzliche Grundlagen für den Einsatz durch öffentliche Stellen** sind erforderlich, da besonders schützenswerte Personendaten bearbeitet werden oder ein Profiling stattfindet. Unbenommen von der Pflicht rechtsstaatlicher Sicherungsmechanismen bleibt die Überprüfung der Notwendigkeit des Technologieeinsatzes, die im Gesetzgebungsverfahren erfolgen sollte.
- **Angemessene Aus- und Weiterbildung** des Personals, das Anwendungen der Stimm-, Sprach- und Gesichtserkennung bedient, für deren Betrieb verantwortlich oder für die Überprüfung und Freigabe von automatisierten Entscheidungen zuständig ist, sodass ein kritischer und verantwortungsbewusster Umgang mit den Technologien möglich ist.
- **Handreichungen für Betreiber von Stimm-, Sprach- und Gesichtserkennungsanwendungen** zur Unterstützung beim datenschutzrechtskonformen Einsatz ihrer Systeme.
- **Unterstützung für Betroffene und Verbesserung der Rechtsdurchsetzung in ihrem Sinne**, um sie bei der Wahrnehmung ihrer Betroffenenrechte und ggf. auch der Nutzung von Anwendungen der Stimm, Sprach- und Gesichtserkennung zu unterstützen und besser vor deren negativen Konsequenzen zu schützen.
- **Gesellschaftliche Debatte über Vor- und Nachteile** von Stimm-, Sprach- und Gesichtserkennungstechnologien und den Umgang mit ihnen. Darunter die Gefahr einer staatlichen Massenüberwachung, die Regulierung der Aussendung politischer Werbung, die Notwendigkeit der Aushandlung sozialer Normen im Kontext der Nutzung smarter Lautsprecher oder der gesellschaftliche Umgang mit den Herausforderungen der Jedermann-Identifikation.
- **Bereitstellung ausreichender Ressourcen** für vertrauenswürdige Dritte und Medienschaffende, damit sie ihren Unterstützungs- und Informationsauftrag für Betroffene bzw. die Gesellschaft wahrnehmen können.

Das Autorenteam hofft, mit der vorliegenden Studie einen Anstoss für die dringend notwendige öffentliche und politische Debatte zu geben und damit zu einem verantwortungsbewussten Einsatz von Stimm-, Sprach- und Gesichtserkennungstechnologien beizutragen. Wichtig erscheint uns dabei, **weder die Chancen noch die Herausforderungen dieser Technologien zu überhöhen oder zu verkennen**. Da erste Anwendungen bereits eingesetzt und weitere schon bald folgen werden, scheint es dabei insb. notwendig zu sein, über **zu ziehende Grenzen zu sprechen**, die unseres Erachtens überall dort nötig sind, wo unüberwindbare Machtasymmetrien entstehen, ungerechtfertigte Eingriffe in Grundrechte stattfinden oder von der Gesellschaft als verwerflich angesehene Zwecke mit der Nutzung der Technologie verfolgt werden.

Executive Summary

The past few years have seen significant technological advances in the area of speech, speaker and facial recognition. An increasing number of detection systems based on these technologies are now entering the market, and both private companies and public authorities have introduced related infrastructures in their operations and activities. On the one hand, these innovative tools have the potential to bring benefits to both individuals and society: they help with everyday tasks, support early medical diagnoses and facilitate police investigations by enabling fast searches for matches in large image databases. On the other hand, fears of unjustified mass surveillance have recently given rise to numerous campaigns in civil society opposing such practices, first and foremost police use of facial recognition systems. Controversy also surrounds other potential uses for a variety of reasons, including the fact that the predictions made by these technologies are currently error-prone or are considered to be socially undesirable. The latter point specifically addresses using external physical traits to draw conclusions about other personal characteristics such as gender, sexual orientation, health or emotional state.

Despite the rapid technological advances, the various speech, speaker and facial recognition systems available are still in an early stage of development: it is unclear how reliably they function, their legal standing is uncertain and the various challenges they may pose to society are difficult to gauge.

Against the backdrop of these uncertainties and in response to the growing awareness of the topic in society, this study was designed to prepare an initial stock of well-researched information about these technologies with the overall aim of promoting social and political discourse on the opportunities and challenges surrounding speech, speaker and facial recognition technologies.

Technologies examined in the study

Eight representative detection technologies formed the focus of the study:

1. As in other countries, **smart speakers** are becoming more common in Swiss homes. Because these speakers are designed to function as digital assistants, they are generally set up in living rooms – and thus often at the heart of an individual's or family's private life.

2. In Switzerland, the **police** began using **facial recognition systems** several years ago. The technology is an aid in seeking suspects and is used ex post to compare crime scene images with images stored in police databases. In some countries, the police also use real-time facial recognition: when searching for missing persons or fugitives, recordings from surveillance cameras at train stations or other public spaces are continuously searched for matches with images in the police databases.

3. In telephone banking, some Swiss banks have introduced **speaker authentication** to replace verification processes in which clients are asked to provide personal data such as their date of birth to verify their identity. Speaker authentication technology is based on the biometric detection of an individual's voice.

4. Using facial recognition systems in the area of **violence prevention in sports stadiums** is one of the most contentious topics surrounding this specific detection tool. In Switzerland, several attempts to introduce the technology have failed; however, many other countries across the globe have already implemented facial recognition infrastructures in sporting venues.

5. The **detection of physical and mental illnesses** using data from a voice recording or a photograph of a person's face is, for the most part, currently in a developmental phase; nevertheless, initial prototypes have already entered the market. These technologies claim to facilitate the (early) detection of various illnesses or impairments such as Parkinson's, Alzheimer's, autism or depression. Moreover, it is hoped that they will find application as an aid in therapies and treatments.

6. **Emotion detection systems** that draw on voice, speech and facial recordings are increasingly used as an optional function in identification or authentication processes. It is believed that emotion detection holds particularly great potential for the field of marketing and for job application procedures on the labour market, as developers promise the tools will deliver precise information about what people «truly» want and what their abilities are.

7. Another area of application for facial recognition technology is **attention monitoring**. Although not yet used in Switzerland, attention monitoring technology has already been implemented in other countries for recording and assessing driver fatigue or the attention levels of schoolchildren.

8. **Universal identification** describes the still fictitious possibility of using miniature cameras that, for instance, are mounted on smartglasses and that not only identify but also supply additional (sensitive) information in real time about every single person in view.

These areas of application represent a cross-section of the **technologies currently in use** (*smart speakers, police use of facial recognition, speaker authentication technologies*), those that have the **potential to be implemented in the near future** (*violence prevention in sports stadiums, detection of physical and mental illnesses, emotion detection*) and those detection systems that, on the basis of current knowledge, will likely remain **hypothetical** but that would have far-reaching implications (*attention monitoring and universal identification*). The study addresses **four main questions** concerning speaker and facial recognition technologies:

- Analysis of the technological foundations and capabilities
- Legal assessment, particularly in the context of data protection laws
- Consideration of societal and ethical challenges
- Survey of perceptions, opportunities, risks and desires in society

Current technological foundations and capabilities

The findings from the analysis of technological foundations and capabilities reveal that the various tools have an **uneven range of technological maturity**. At present, some already function with a high degree of reliability (*smart speakers*, *speaker authentication*) or it is expected that ongoing developments will result in a **high degree of reliability** in the coming years (*ex post facial recognition by the police*, *violence prevention in sports stadiums*). In the area of *disease and emotion detection* systems, by contrast, greater challenges arise that **complicate reliable implementation**. To overcome these obstacles, concentrated efforts on the basic concept of a technology as well as improvements in the underlying datasets are needed. Moreover, the socio-technical context in which a technology is implemented must also be considered when assessing the **detection rate** of speech, speaker and facial recognition tools: for instance, advanced camera sensors in a controlled, well-lit environment where the recorded person is cooperating will yield much higher hit rates than a recording of a moving person made in a poorly lit area from a distance or with low-quality sensors.

For this reason, the assessment of a system's technical-organisational trustworthiness plays a key role. Here, an **independent review of the information on hit rates given by a software developer** would be of particular value. However, this type of assessment is difficult, particularly in those cases in which private entities are responsible for using a technology. Indeed, other than the informative studies by the National Institute of Standards and Technology (NIST), relatively few scientific findings on recognition algorithms and their hit rates are available. Moreover, these studies are generally constrained by the fact that they were conducted in experimental test environments in which a tool's actual socio-technical context was not accounted for; they therefore have limited value for real-world applications that utilise a range of sensor technologies and that are implemented in different environments.

All this implies that the **actual hit rates of a monitoring infrastructure are concealed beneath a mixture of marketing pitches, trade secrets and security requirements**. A system's technical reliability is also compromised by the **bias problem**: because speech, speaker and facial recognition systems work on the basis of datasets prepared and edited by humans, and because they operate using human-designed processes and algorithms, they always contain a bias. While it is possible to reduce this bias, it most likely will never be completely eliminated.

A more fundamental problem concerns cases in which scientifically based categories are lacking; to an extent, this pertains to disease detection, but the main issue lies in the field of emotion detection. Emotions are impossible to delineate clearly – and thus impossible to categorise – and they vary greatly in intensity, which makes an accurate calculation of hit rates problematic, as all categorisations will remain probability values based on a trait's relation to a given category. Further research is therefore needed on the fundamental concepts of these technologies.

It should also be noted that the results of the analysis of technological foundations and capabilities suggest that in future, too, even those speech, speaker and facial recognition systems that are based on the most advanced technologies will **fail to operate with 100-percent technical reliability**.

Legal bases

In the vast majority of the cases examined, the **biometric data** stored are essentially records of an individual's face and voice. This means they are **personal data particularly worthy of protection**, as these data depict characteristics that are closely and permanently associated with a specific person and that constitute unique aspects of an individual's personality. Processing these data always represents a breach of privacy or a serious infringement of an individual's fundamental rights.

Use of speech, speaker and facial recognition technologies by **public authorities** therefore has bearing on the **right to privacy and protection against the misuse of personal data** (Federal Constitution, Art. 13, para. 1 and 2). Moreover, these technologies can have a so-called chilling effect on how individuals exercise various fundamental rights such as freedom of expression, freedom of assembly and freedom of association (Federal Constitution, Art. 16, 22, 23). Procedural guarantees (Federal Constitution, Art. 29, 29a, 30) and anti-discrimination guarantees (Art. 8 para. 2) may also be affected. To ensure that fundamental rights are respected when such technologies are implemented, the legislative body must conduct a careful examination of the tools before they are put into operation in addition to creating the necessary **legal basis** for their use. As a rule, this will be an **act in a formal sense** that is **framed in such a way** that not only the purpose of data processing is **unmistakably clear** but also that actual processing of data is **limited** to an absolute minimum, both in terms of the time frame of data availability and the scope of use.

In addition, the public authority responsible for applying the law must assess a technology's **proportionality** when creating the legal basis for its use and before a tool finds actual application. Here, it is important to ask whether using a tool is **appropriate**, especially as the current state of development often means the **reliability of a technology** cannot be guaranteed, as the study findings demonstrate (this particularly concerns emotion detection systems). Also not automatically given is a tool's **necessity**: we already have less invasive technologies capable of achieving the same public interests as speech, speaker and facial recognition technologies. Moreover, certain uses should remain proscribed because they violate the constitutionally protected **principles** of several fundamental rights. This in particular includes facial recognition infrastructures for **blanket real-time surveillance** as well as state-run **emotion detection** systems.

As a rule, **private** use of speech, speaker and facial recognition technologies violates several data protection principles, which means that the data processing must be **justified** in accordance with Article 31 of the new Federal Act on Data Protection. Due to the fact that biometric data are particularly sensitive, a data processor's claim to having an overriding interest is unlikely to provide sufficient justification in most cases. However, obtaining **consent** from data subjects in the case of speech, speaker and facial recognition is often problematic, as this is **practically impossible** (for example, in the case of *smart speakers* or *smartglasses* and other *universal identification tools*). Moreover, **power asymmetries** (in job applications, with insurance companies or at private schools, for instance) or the fact that there are no pared-down data collection systems (at sports stadiums or banks), means that consent is **not voluntary** or that individuals are only **inadequately informed** about what they are agreeing to.

Because numerous risks arise for data subjects when these tools are used by private actors, and because both users and operators often fail to comply with the legal requirements when using the technologies examined in this study, the state is under a constitutional **duty to protect against violations of fundamental rights by private persons** (Federal Constitution, Art. 35). This duty to protect lays the groundwork for the state **to regulate private use of these technologies and to monitor compliance with the rules**. The project team have proposed corresponding recommendations.

From the perspective of data protection, it is also important to point to the principle of **data security**, which must be observed by public authorities and private entities alike. The consequences of lost, misused or misappropriated biometric data are serious; measures to guarantee data security must therefore meet correspondingly stringent requirements.

Ethical and societal challenges

Speech, speaker and facial recognition infrastructures can pose serious ethical challenges even when all (data protection-relevant) legal requirements are met. Negative impacts include the «**chilling effects**» **of surveillance** systems, which limit an individual's ability to exercise fundamental rights. And even when facial recognition tools – those used by the police, for instance – comply fully with the law, there is still the possibility that **wholesale state-run surveillance will gradually creep** into ever more areas of our lives. Moreover, monitoring systems also have the potential to bring about behavioural changes or internalised conformism in other domains. Of particular mention is that certain kinds of emotion detection would confront people with a stark choice: either bow to the pressure of being monitored and adapt their behaviour to be as conformist or advantageous as possible – will an artificial smile do the trick? – or accept the consequences. Operating on the assumption that 100-percent technical reliability will remain elusive, it must be expected that **false positive and false negative matches** will continue to occur in future – despite improved recognition rates – and that **undesirable effects** will be unavoidable. Such effects include unjust arrests due to erroneous police facial recognition matches, the danger of a bank account being breached in a telephone banking session, and false negatives in medical (self) diagnoses that potentially have life-threatening consequences.

Already today, businesses and public authorities have access to much more information about the population than they did just a few years or decades ago. The discussion on ethical challenges has made it clear that, in several areas of application, the **knowledge and thus power imbalances** arising with the increasing use of speech, speaker and facial recognition systems will most likely continue to shift to the advantage of those entities that process data. Although in individual cases, (data protection) laws can help to protect against all-too invasive misuse of this increasing power, they are unable to stop the trend on the whole. Moreover, **numerous possibilities to manipulate individual behaviours** remain: predictive models (people who do A and B are also highly likely to do/think/feel X) based on the analysis of anonymous or anonymised data of a large group of people may enable sensitive conclusions to be drawn about a specific person if this individual's personal data is entered into the system.

In the scope of prevailing data protection law, a particularly difficult challenge concerns **lawful consent**, as the act of granting consent can create the illusion of control in situations where extreme power asymmetries exist. With human recognition of a face – or interpretation of emotions and behaviours – it is fairly clear how and to what extent this takes place. By contrast, speech, speaker and facial recognition systems permit the **collection of highly detailed information about an individual**. Consequently, people who are unaware of this capacity may **recklessly consent to the processing of their personal data**. And in the case of emotion detection tools, people may grant consent only because they believe they can outsmart an error-prone technology by adjusting their behaviour to their own advantage – although the system could ultimately see through these attempts to an individual's detriment. At the same time, the example of emotion detection illustrates that, just as an error-prone technology bears certain risks, **a technology that operates largely error-free also presents serious challenges**. Should a technology be perfected, dangers arising through state-operated mass surveillance systems or the previously discussed possibility of *universal identification* tools in which passers-by are identified would be particularly acute.

In the medium to long term, developers also plan to transfer speech, speaker and facial recognition technologies to **convergent socio-technical systems** and to realise applications such as emotion detection in telephone banking and job interviews, or recognition technologies embedded in smart speakers. With this development, the quantity – and even more so the quality – of the data collected about a data subject's behaviour would increase drastically, and the existing trend towards quantifying and steering human behaviours and experiences would permeate the last few life domains that have not yet been digitised. Used as auxiliary tools, advanced assistive or diagnostic technologies in particular would have the potential to bring great benefits to everyday life. However, the resulting data feedback loops also imply that human existence would be further steered by calculations and rote algorithms. On the one hand, if the underlying algorithms were subject to democratic monitoring, this could actually boost self-determination, both in individuals and on a societal level. On the other hand, however, poor transparency in algorithm monitoring could degrade the idea of self-determination to a farce, as these algorithms would influence decision-making in general, from major life choices on to the most intimate matters.

Another core challenge is related to what is frequently considered one of the great advantages of detection technologies: increased efficiency due to the use of speech, speaker and facial recognition systems. Here, the question arises as to whether increased efficiency will bring about better police work, cheaper tickets to sporting events and improved patient care, or whether economic rationalisation will simply lead to **staff cuts and a one-sided distribution of profits**.

The range of ethical challenges underscores the **urgent need for societal discourse** on the matter – discourse that must focus on the question as to how potential capacities of detection technologies can be brought into alignment with a democratic political system and the fundamental rights of citizens.

Public perception

In the context of analysing public perception of detection technologies, it became clear that Swiss citizens understand and appreciate the opportunities they offer. **Gains in the areas of safety, health and convenience** number among the advantages the persons surveyed anticipate. They also expect to see increased efficiency, so that freed-up human or financial resources could be put to good use elsewhere. However, the positive assessment on the part of citizens is to a significant extent founded in a possibly misguided understanding that detection tools have the potential to reduce human-made discrimination and that they are operated by trustworthy entities.

At the same time, however, a certain degree of public **mistrust** was voiced about these technologies and, in particular, about their operators. Here, the most serious worries concern a **lack of transparency** and **the misuse of personal data** in, for instance, cases of misappropriated access data or unjustified mass surveillance. The persons surveyed also expressed concerns that the reliability of speech, speaker and facial recognition technologies is not yet adequate and could thus result in undesirable consequences such as fatal medical diagnoses, for instance, or being arrested or placed under observation due to a **false match** in a police database. Moreover, many respondents worried that the technologies could be used against their interests (unwanted advertising, price discrimination).

The survey of the general population also revealed some considerable **knowledge gaps** and thus a corresponding **need for information**, which was reflected in their support for, or rejection of, various uses of a detection technology: attention monitoring at schools and universal identification were clearly rejected, whereas the use of facial recognition in sports stadiums was clearly approved. Moreover, the **high percentage of undecided persons** regarding all other uses is striking, a finding that underscores how important it is to promote public discourse on these topics.

In general, the citizens surveyed believe the state and the service providers bear responsibility for the trustworthy use of speech, speaker and facial recognition systems. Nevertheless, they also saw a **need for far-reaching rules on transparency** as well as **bans and information campaigns**, and they called for independent expert **monitoring and evaluation of the technology implementations**, prior testing for efficiency and mandatory certification; they also pointed out that state-run systems must be based on legal foundations that are as specific as possible.

Recommendations

The outcome of the study reveals a clear **need for action**, and the project team have formulated possible ways in which decision-makers in the political sphere, the economy, data protection authorities, the media and civil society can address the various challenges posed by speech, speaker and facial recognition systems. On the basis of the findings, **a ban in three areas and a moratorium in one appear justified:**

- **Ban on automated state-operated real-time surveillance systems and state-operated social credit systems using speech, speaker or facial recognition tools.**
- **Ban on fully automated decision-making based on speech, speaker and facial recognition systems in important life domains** (such as health, law enforcement, finance and lending, insurance and the world of work). Rather, the results of semi-automated systems to support decision-making should be closely monitored by trained staff, who are responsible for authorising automated decisions.
- **Ban on the use of smartglasses in public spaces** and on the use of other not immediately recognisable technologies that can be used for surveillance in public spaces; this includes small cameras that are linked to facial recognition technology.
- **Moratorium on emotion and disease detection tools in sensitive life domains** (for instance, law enforcement, finance and lending, insurance, school settings and the world of work), until it is proven that these systems are sufficiently reliable, both from a technical and organisational standpoint. This particularly includes a ban on the private use of emotion and disease detection technologies for speaker authentication and job applicant selection as well as a ban on attention monitoring at schools. Specific desirable tools could be classified as high-risk and thus permitted under certain conditions.

Moreover, **biometric data, in particular speaker detection data, should not be used as the (sole) means of authentication**. The reason is that, unlike variable authentication criteria (passwords, for instance), people cannot change their physical features because their biometric data have been compromised.

Additional recommendations are divided into general recommendations and recommendations for specific purposes. General recommendations include the following:

- The **regulation of high-risk tools** – in particular those used in the fields of health, law enforcement, lending, insurance and work – is imperative, as they pose a threat to our fundamental rights and to how we live together as a society. All regulatory measures should encompass a risk management strategy that contains an independent evaluation of a tool's technical and organisational reliability, mandatory supervision by a human being, data security standards and independent control mechanisms.
- **Explicit legal bases for use in the public sector** must be introduced, as particularly sensitive personal data may otherwise be processed or used for profiling. In all cases, the necessity of a technology must be assessed before a legal basis is created.
- **Targeted education and further training** of staff that use speech, speaker and facial recognition systems in order to ensure a responsible and discerning approach to using these tools; this also pertains to staff in charge of operating a tool or of assessing and approving automated decisions.
- **Guidelines for operators of speech, speaker and facial recognition systems** to aid them in complying with data protection laws.
- **Support for data subjects and improvement of legal enforcement on their behalf** to aid individuals in exercising their data-subject rights and, if required, in using speech,

speaker and facial recognition systems, and to better protect them against potential negative consequences of using these systems.

- **Public discourse on the advantages and disadvantages** of speech, speaker and facial recognition technologies and how they should be used. Topics include the danger posed by state-operated mass surveillance, the regulation and broadcasting of political advertisements, the necessity of negotiating social norms for using smart speakers and how societies can deal with the challenges of universal identification.
- **Provision of sufficient resources** for trusted third parties and media professionals to aid them in realising their work to support and inform individuals and the broader community.

The project team hope this study will serve as a starting point for the urgently needed public and political discourse on this topic and thus make a valuable contribution to a responsible use of speech, speaker and facial recognition systems. When holding these discussions, it is important to **neither exaggerate nor underestimate the opportunities and challenges surrounding these technologies**. Because the first tools are already in use and others will soon follow, it is also essential to **discuss where boundaries must be drawn**; the authors believe such boundaries are necessary wherever entrenched power asymmetries arise, where unjustified infringements of fundamental rights occur or when a technology is used to pursue socially unacceptable interests.

Résumé

Le développement des technologies de reconnaissance de la voix, de la parole et du visage a connu des avancées significatives ces dernières années. Dans ce domaine, une multitude d'applications envahissent déjà le marché et sont utilisées aussi bien par les entreprises privées que par les autorités. Sans doute, ces applications innovantes revêtent une dimension utile pour les individus et la société – par exemple, elles apportent une aide dans les tâches quotidiennes, diagnostiquent des maladies à un stade précoce et recherchent rapidement des concordances dans de grandes bases de données d'images dans le cadre d'enquêtes policières. Mais, ces dernières années, on assiste aussi à de nombreuses campagnes de rejet de la part de la société civile, notamment contre le recours à la reconnaissance faciale par la police par crainte d'une surveillance de masse sans motif. D'autres applications sont aussi sujettes à controverse, par exemple lorsque leurs prédictions sont considérées comme techniquement erronées ou socialement indésirables, et notamment lorsqu'elles permettent de déduire certaines caractéristiques d'une personne (genre, orientation sexuelle, état de santé, émotions, etc.) à partir de particularités physiques visibles.

Il n'en reste pas moins que, malgré les progrès de cette technologie, les différentes applications basées sur la reconnaissance de la voix, de la parole et du visage en sont encore à un stade précoce : on ignore souvent si elles sont techniquement fiables, si elles répondent aux exigences légales et quels sont les défis que leur utilisation pourrait poser à la société.

Pour pallier ce manque de clarté et répondre à l'intensification des débats de société, la présente étude vise à fournir une base de connaissances fondées en matière d'utilisation de ces technologies afin d'encourager et d'orienter le discours sociétal et politique sur les opportunités et les défis de la reconnaissance du visage, de la voix et de la parole.

Examen de domaines d'application choisis

L'étude examine huit domaines d'application exemplaires :

1. Les **haut-parleurs intelligents** sont de plus en plus présents, y compris dans les foyers suisses. Censés faire office d'assistants numériques, ces appareils sont généralement installés dans les salons et se trouvent donc au cœur de la vie privée de beaucoup de gens.
2. La **reconnaissance faciale par la police** existe en Suisse depuis quelques années. Cet outil d'assistance technique est utilisé ex post pour comparer les enregistrements de la scène de crime avec les photos stockées dans les bases de données de la police en cas de recherches d'individus. Certains États ont aussi recours à la reconnaissance faciale en temps réel : les enregistrements des caméras de surveillance, installées pour la plupart dans les lieux publics, les gares, etc., sont comparés en permanence avec les bases de données de la police pour rechercher des personnes disparues, en fuite, etc.

3. En Suisse, certaines banques ont recours à l'**authentification par la voix** pour remplacer, lors d'opérations bancaires par téléphone, l'authentification via la demande de données personnelles par la reconnaissance biométrique de la voix du client ou de la cliente.

4. La reconnaissance faciale aux fins de **prévention et d'éducation contre la violence dans les stades de sport** est particulièrement controversée depuis des décennies. Tandis qu'en Suisse, plusieurs tentatives pour l'introduire ont échoué, elle est déjà utilisée dans de nombreux pays du monde entier.

5. La technologie de **détection de maladies physiques et psychiques** à partir d'enregistrements de la voix ou du visage est encore en phase de développement, mais ses premières applications arrivent déjà sur le marché. Elles devraient permettre de détecter (à un stade précoce) différentes pathologies ou déficiences, comme la maladie de Parkinson, la maladie d'Alzheimer, l'autisme ou la dépression. En outre, ces applications pourraient aussi apporter un soutien dans le cadre d'une thérapie.

6. La fonction de **reconnaissance des émotions** sur la base d'enregistrements de la voix, de la parole et du visage est une option à choix toujours plus fréquente, en plus de l'identification ou de l'authentification des personnes. Dans le domaine du marketing et des procédures de recrutement sur le marché du travail notamment, la reconnaissance des émotions se voit attribuer un grand potentiel : elle soulève l'espoir d'obtenir des informations sur les « véritables » souhaits, capacités, etc. des personnes concernées.

7. L'**analyse de l'attention** est un autre domaine où la reconnaissance faciale se développe. Si l'on n'y recourt pas encore en Suisse, des applications existent déjà dans d'autres pays, par exemple pour enregistrer et évaluer l'attention des automobilistes ou des élèves à l'école.

8. Encore au stade de fiction à ce jour, l'**identification de toute personne** désigne la possibilité, comme son nom l'indique, d'identifier en temps réel toutes les personnes se trouvant dans son champ de vision grâce à des caméras miniatures intégrées – par exemple, dans des lunettes connectées – et de fournir des informations supplémentaires (sensibles) à leur sujet.

Les domaines ci-dessus représentent une vue d'ensemble des **applications déjà existantes et utilisées** aujourd'hui (*haut-parleurs intelligents, reconnaissance faciale par la police, authentification par la voix*), des **applications qui pourraient être utilisées à moyen terme** (*prévention et éducation contre la violence dans les stades de sport, détection de maladies et de troubles physiques et psychiques ou reconnaissance des émotions*) et de scénarios d'application encore plutôt hypothétiques à ce jour, mais qui ont potentiellement des implications importantes (*analyse de l'attention et identification de toute personne*). L'étude des divers champs d'application de la reconnaissance du visage, de la voix et de la parole s'articule autour des **quatre axes principaux** suivants :

- Analyse des fondements et des possibilités techniques
- Évaluation juridique, notamment du cadre juridique en matière de protection des données

- Débat sur les défis sociaux et éthiques
- Étude de la perception par la société des opportunités, des risques et des souhaits

Fondements et possibilités techniques à ce jour

L'étude des fondements et des possibilités techniques révèle **un large éventail de maturité des technologies** : la **fiabilité technique** des applications actuelles est déjà **très élevée** (*haut-parleurs intelligents*, *authentification par la voix*) ou pourrait le devenir suite aux efforts de développement en cours (*reconnaissance faciale ex post par la police*, *prévention et éducation contre la violence dans les stades de sport*). Les applications dans le domaine de la *détection des maladies* et de la *reconnaissance des émotions* sont en revanche confrontées à des défis parfois plus importants, qui constituent un **obstacle à une utilisation fiable**. Dans ces domaines, un effort intensif doit être fourni pour établir les fondements conceptuels et améliorer les données sous-jacentes.

En ce qui concerne les **taux** de reconnaissance de la voix, de la parole et du visage, il faut surtout tenir compte du contexte sociotechnique dans lequel cette reconnaissance a lieu : ainsi, dans un environnement contrôlé, bien éclairé et dans lequel une personne se comporte de manière coopérative, les capteurs de caméra avancés promettent des taux de réussite bien plus élevés que lorsque le visage d'une personne en mouvement est enregistré à distance dans des conditions de faible éclairage ou au moyen de capteurs de mauvaise qualité. C'est pourquoi l'évaluation de la fiabilité technique et organisationnelle d'une application joue un rôle central. Dans cette optique, il faudrait notamment procéder à une **vérification indépendante des données diffusées par les fabricants sur les taux de réussite de leurs logiciels**. Or, cette évaluation est particulièrement difficile à réaliser lorsque les responsables de l'utilisation de la technologie sont des acteurs privés. En effet, hormis les études instructives du National Institute of Standards and Technology (NIST), il n'existe que quelques études scientifiques isolées sur les algorithmes de reconnaissance ou leurs taux de réussite. Celles-ci sont par ailleurs généralement limitées par le fait qu'elles ont lieu dans des environnements de test expérimentaux et que le contexte d'utilisation sociotechnique réel n'est pas pris en compte. Par conséquent, au vu de l'influence sur leurs résultats du système de capteurs et des conditions environnementales variables, ces études ne peuvent être transposées à la pratique que dans une mesure limitée.

En ce sens, **les taux de réussite réels des applications sont masqués par une combinaison de promesses marketing, de secrets d'entreprise et d'exigences en matière de sécurité**. La fiabilité technique de ces systèmes est d'ailleurs affectée par le **problème des biais** : parce que les applications basées sur la reconnaissance de la voix, de la parole et du visage s'appuient sur des données d'origine humaine traitées par l'humain, ainsi que sur des algorithmes et des processus créés par l'humain, elles contiendront toujours un biais qui, s'il peut être réduit, ne pourra probablement jamais être complètement éliminé.

Un problème plus fondamental se pose dans les cas où il manque des catégories scientifiquement fondées, ce qui s'applique en partie à la détection des maladies, mais surtout à la reconnaissance des émotions. En effet, les émotions ne sont pas toujours clairement

différenciées – c'est-à-dire qu'elles ne peuvent pas être clairement catégorisées – et varient fortement en intensité. Le calcul de taux de réussite est donc problématique puisqu'il s'agit de valeurs de probabilité par rapport à la catégorie concernée, pour toutes les classifications catégorielles. Il est donc nécessaire de poursuivre les recherches sur les fondements conceptuels dans ces domaines.

En conclusion de l'analyse des fondements et des possibilités techniques, il faut retenir que même les systèmes les plus fiables basés sur les technologies de reconnaissance de la voix, de la parole et du visage ne pourront **pas** atteindre **une fiabilité technique de 100 %** à l'avenir.

Cadre juridique

Dans la grande majorité des cas étudiés ici, le visage et la voix d'une personne constituent des **données biométriques** et donc **des données personnelles sensibles**. Il s'agit de caractéristiques étroitement et durablement liées à l'individu, qui constituent des aspects uniques de sa personnalité. Le traitement de ces données constitue dans tous les cas une atteinte à la personnalité, c'est-à-dire une atteinte grave aux droits fondamentaux des individus.

Le **droit à la vie privée et la protection contre l'emploi abusif des données personnelles** (art. 13, al. 1 et 2 Cst.) sont concernés lorsque les **autorités** recourent aux technologies de reconnaissance de la voix, de la parole et du visage. En matière de communication (liberté d'opinion, liberté de réunion et d'association, art. 16, 22, 23 Cst.), le recours à cette technologie peut avoir un effet dissuasif (*chilling effect*) sur l'exercice de différents droits fondamentaux. Les droits fondamentaux de procédure (art. 29, 29a, 30 Cst.) ainsi que l'interdiction de discrimination (art. 8, al. 2 Cst.) peuvent également être concernés. Pour être conforme aux droits fondamentaux, l'utilisation de ces technologies nécessite un examen préalable minutieux de la part du législateur. Celui-ci est tenu de créer la **base légale** nécessaire à leur utilisation, en règle générale dans une **loi au sens formel**, et de la **concevoir de manière suffisamment précise** pour qu'il en ressorte non seulement un objectif clair, mais aussi, notamment, une **limitation** du traitement des données au strict nécessaire, tant en termes de temps que d'étendue.

La **proportionnalité** du recours à ces technologies doit être contrôlée au moment où la base légale est créée, mais aussi avant que les autorités chargées d'appliquer le droit ne les mettent en pratique. Ainsi, en fonction de l'état actuel de la technique, le **caractère approprié** du recours à ces technologies est en général remis en question car, **très souvent, la fiabilité de la technologie n'est pas garantie** – comme l'a révélé la présente étude (en particulier en matière de reconnaissance des émotions). Souvent, le **caractère nécessaire** n'est pas non plus avéré, car il existe des moyens moins contraignants que ces technologies pour atteindre les intérêts publics visés. De plus, certaines applications susceptibles de violer le **noyau** absolument protégé de plusieurs droits fondamentaux demeurent par conséquent illicites. Il s'agit notamment de l'utilisation de la technologie de reconnaissance faciale pour une **surveillance généralisée en temps réel** et de la mise en œuvre de la **reconnaissance des émotions** par l'État.

Le recours à la technologie de reconnaissance de la voix, de la parole et du visage par des **particuliers** entraîne en général la violation de plusieurs principes de protection des données, ce qui implique que tout traitement des données soit **justifié**, selon l'article 31 de la nouvelle loi sur la protection des données (nLPD). Dans la plupart des cas, l'intérêt prépondérant de la personne qui traite les données ne devrait pas entrer en ligne de compte comme motif justificatif en raison du caractère particulièrement sensible des données biométriques traitées. Toutefois, obtenir le **consentement** des personnes concernées dans le domaine de la reconnaissance de la voix, de la parole et du visage, est souvent problématique. En effet, soit ce consentement est **impossible à donner ou à obtenir en pratique** (par exemple dans le cas des *haut-parleurs intelligents* ou des *lunettes connectées* et d'autres *applications grand public*), soit il est **non volontaire en raison d'une asymétrie de pouvoir** (dans les procédures de candidature, les assurances ou les écoles privées) ou faute d'offres économes en données (dans le cas de l'utilisation dans les stades ou par les banques), soit enfin il est obtenu sur la base d'une **information insuffisante ou inappropriée**.

Comme son utilisation par des acteurs privés comporte de nombreux risques pour les personnes concernées et que, dans les domaines d'application étudiés, les personnes qui y recourent ou l'exploitent n'agissent souvent pas de manière conforme au droit, il en résulte un **devoir de protection de l'État contre les violations des droits fondamentaux par des particuliers** (art. 35 Cst.). Ce devoir de protection exige de l'État qu'il **réglemente** l'**utilisation de cette technologie par les particuliers et contrôle le respect des règles**. L'équipe de projet présente des recommandations en ce sens.

Du point de vue de la protection des données, il convient de mentionner enfin le principe de la **sécurité des données**, qui doit être respecté tant par les autorités que par les particuliers. Les risques de perte, d'utilisation abusive ou de détournement des données biométriques par des personnes non autorisées sont sérieux : les mesures visant à garantir la sécurité des données doivent donc répondre à des exigences élevées.

Défis éthiques et sociétaux

Même si toutes les exigences légales (en matière de protection des données) sont remplies, les applications de reconnaissance du visage, de la voix et de la parole peuvent poser d'autres défis éthiques considérables. Il s'agit notamment **de l'effet dissuasif de la surveillance** qui entrave l'exercice des droits fondamentaux. Et même si, par exemple, l'exploitation de la reconnaissance faciale par la police est conforme au droit, il existe toujours la possibilité qu'une **surveillance étatique globale s'immisce insidieusement** dans un nombre croissant de domaines de la vie. Mais des changements de comportement ou une forme de conformisme internalisé pourraient également apparaître dans d'autres domaines. En particulier, il se peut que certains types de reconnaissance des émotions confrontent les personnes concernées à une alternative délicate : se plier à la pression et se comporter de la manière la plus conforme aux règles ou la plus avantageuse possible – un sourire de façade pourrait-il être utile ? – ou s'accommoder des inconvénients.

Comme il faut partir du principe qu'il sera impossible d'obtenir une fiabilité technique à 100 %, il est à prévoir que, malgré la progression en termes de taux de détection, des **faux positifs et des faux négatifs** continueront de se produire et d'entraîner des **conséquences indésirables** : une arrestation injustifiée – en cas de reconnaissance faciale par la police –, l'exposition au risque d'accès illégal à des comptes lors d'opérations bancaires par téléphone, voire des (auto)diagnostics faussement négatifs dans le domaine médical aux conséquences potentiellement mortelles.

Aujourd'hui déjà, les services publics et économiques disposent d'une somme d'informations sur la population bien plus vaste qu'il y a encore quelques années ou décennies. La discussion sur les défis éthiques a clairement révélé que, dans plusieurs domaines d'application, le recours croissant aux technologies de reconnaissance de la voix, de la parole et du visage fera vraisemblablement en sorte que ce **déséquilibre des connaissances**, et donc **du pouvoir**, continuera à se déplacer en faveur des personnes qui traitent les données. Si les lois (sur la protection des données) offrent une protection contre les abus invasifs de ce pouvoir croissant au cas par cas, elles ne peuvent pas stopper cette évolution. De plus, **de nombreuses possibilités de manipulation du comportement individuel** persistent encore : par exemple, les modèles prédictifs (tout individu qui fait A et B a de fortes chances de faire/penser/ressentir X), qui s'appuyent sur l'analyse de données anonymes d'autres individus, peuvent aussi servir à tirer des conclusions sensibles sur une personne en particulier, sur la base des données propres à cette dernière.

Dans le cadre du droit en vigueur, le **consentement en matière de protection des données** représente justement un grand défi car, dans des situations d'extrême déséquilibre des forces, il peut créer l'illusion d'un contrôle. Tandis que, dans le cas de la reconnaissance par un être humain d'un visage – ou de l'interprétation d'émotions, de comportements etc. –, l'ampleur et la façon dont ces derniers sont perçus par autrui est relativement claire, les technologies de reconnaissance de la voix, de la parole et du visage permettent quant à elles d'**obtenir des informations très poussées sur les personnes concernées**. Toute personne qui n'est pas consciente de ce potentiel est susceptible de **consentir à la légère au traitement de ses données**. Dans le cas de la reconnaissance des émotions, la personne concernée pourrait consentir à son utilisation et, sous-estimant la technologie, tenter de s'en jouer en adaptant son comportement à son avantage, bien que cela soit perceptible et se traduise par l'effet inverse pour elle. L'exemple de la reconnaissance des émotions illustre bien l'ambivalence de ce phénomène : si une technologie présente des risques lorsqu'elle est techniquement défaillante, elle présente aussi des **défis à prendre au sérieux lorsqu'elle fonctionne sans problème technique**. Des dangers comme la surveillance de masse par l'État ou la possibilité d'identifier les individus « en passant » – dont il est question avec l'*identification de toute personne* – pourraient acquérir un caractère urgent si la technologie se perfectionne.

À moyen et long terme, les fournisseurs de technologies prévoient également de poursuivre l'intégration des technologies de reconnaissance de la voix, de la parole et du visage dans **des systèmes sociotechniques convergents**, comme de recourir à la reconnaissance des émotions pour les opérations bancaires par téléphone et les entretiens d'embauche, ou de les intégrer dans des haut-parleurs intelligents. Cette évolution aurait pour conséquence d'augmenter fortement la quantité et surtout la qualité des données collectées sur le com-

portement des personnes concernées. Les tendances à la quantification et au contrôle du comportement et de la vie humaine s'étendraient ainsi jusqu'aux derniers domaines non encore numérisés. En tant que moyen d'assistance, de tels systèmes avancés, notamment d'assistance ou de diagnostic, seraient d'une grande utilité au quotidien. Toutefois, à cause des boucles de rétroaction technologiques, la vie humaine risque d'être de plus en plus guidée par les calculs et les routines de la technologie. Pour peu qu'un contrôle démocratique des algorithmes sous-jacents soit disponible, cela pourrait tout à fait accroître l'autodétermination sociale et individuelle. Si, au contraire, le contrôle algorithmique mis en place est opaque, l'autodétermination pourrait se résumer à une farce : les décisions concernant les grandes questions seraient dominées par ces algorithmes jusque dans la sphère privée.

Un autre défi central est lié au gain d'efficacité, souvent considéré comme un grand avantage, qui découle de la reconnaissance automatisée de la voix, de la parole et du visage. La question est de savoir si ce gain d'efficacité potentiel se traduira par une amélioration de la qualité du travail de la police, par des billets à prix réduit dans les stades ou par une amélioration des soins de santé – ou si les effets de rationalisation économique entraîneront **des économies de personnel et une répartition unilatérale des éventuels bénéfices**.

La discussion sur les défis éthiques révèle notamment le **besoin urgent d'un débat de société**. Il s'agit, en fin de compte, de savoir comment concilier les potentiels du développement technologique avec un système politique démocratique respectant les droits fondamentaux.

Perception du public

L'analyse de la perception du public révèle que la population reconnait et apprécie pleinement les **opportunités offertes par les technologies**. Les citoyennes et citoyens perçoivent comme un avantage le **gain** visé **en termes de sécurité, de santé** et **de confort**, tout en espérant **un gain d'efficacité** dans le but de libérer à bon escient des ressources humaines ou financières. Dans cette évaluation positive, un facteur contestable a joué un rôle non négligeable, à savoir la croyance que le recours à ces technologies puisse réduire la discrimination d'origine humaine et que les acteurs responsables de cette utilisation soient des organismes de confiance.

En parallèle, ces technologies, et en particulier les personnes qui y recourent, **suscitent la méfiance** : la **peur d'une utilisation non transparente de la technologie** et d'une **utilisation abusive des données personnelles** prédomine, comme dans le cas de la saisie des données d'accès aux comptes bancaires ou la surveillance de masse sans motif. En outre, la crainte a été exprimée que les technologies de reconnaissance de la voix, de la parole et du visage ne soient pas suffisamment fiables, ce qui pourrait avoir des conséquences indésirables, telles qu'un contrôle ou une arrestation suite à une **reconnaissance erronée** ou encore le **diagnostic erroné** et fatal d'une application médicale. Enfin, le fait que ces technologies puissent également être utilisées à l'encontre de leurs intérêts (publicité non sollicitée, discrimination par les prix, etc.) a été jugé comme préoccupant par une bonne partie de personnes interrogées.

Le sondage d'opinion révèle toutefois aussi que la population présente d'importantes **lacunes de connaissances** en la matière et éprouve le **besoin d'être informée**. Cela se reflète également dans le soutien ou le rejet de différentes applications dans l'opinion publique : l'identification de toute personne et l'analyse de l'attention dans les écoles sont clairement rejetées, tandis que l'utilisation de la reconnaissance faciale dans les stades est clairement approuvée. En revanche, pour toutes les autres applications, le **pourcentage élevé de personnes indécises** est frappant et souligne l'importance d'un débat public sur les sujets discutés.

Afin de garantir une utilisation fiable des technologies de reconnaissance vocale et faciale, les citoyennes et citoyens considèrent que l'État et les opérateurs de services ont un devoir à remplir. Il a été fait référence au **besoin d'un cadre réglementaire étendu en matière de transparence, d'interdictions, de campagnes d'information, de suivi et d'évaluation du recours aux technologies** par des expertes ou experts indépendants, de tests d'efficacité préalables, d'une obligation de certification et du fait que l'utilisation par l'État doit reposer sur des bases légales aussi concrètes que possible.

Recommandations

Les résultats de l'étude montrent clairement la **nécessité d'agir**. Nous formulons ci-dessous des propositions visant à aider les décisionnaires politiques et économiques, les autorités de protection des données, les médias et la société civile à faire face aux différents défis posés par les technologies de reconnaissance de la voix, de la parole et du visage. Sur la base des conclusions de la présente étude, une **interdiction** semble particulièrement indiquée dans trois domaines, ainsi qu**'un moratoire** dans un autre domaine:

- **Interdiction de la surveillance automatisée en temps réel et du scoring social par l'État au moyen de la reconnaissance de la voix, de la parole ou du visage.**
- **Interdiction des décisions entièrement automatisées basées sur les technologies de reconnaissance de la voix, de la parole et du visage dans des domaines clés de la vie** (par ex. dans le domaine de la santé, des enquêtes criminelles, de la finance et du crédit, des assurances, de l'environnement de travail). Les résultats des systèmes d'aide à la décision semi-automatisés devraient plutôt être soumis pour examen critique et validation à un personnel formé.
- **Interdiction de l'utilisation dans les lieux publics de lunettes connectées** et d'autres technologies non directement identifiables qui se prêtent à la surveillance, comme les petites caméras associées à la technologie de reconnaissance faciale.
- **Moratoire sur la reconnaissance des émotions et la détection des maladies dans des domaines importants de la vie** (par ex. dans le domaine des enquêtes criminelles, des finances et de l'octroi de crédits, des assurances, de l'environnement scolaire et professionnel), tant que leur équité et fiabilité technique et organisationnelle ne sont pas démontrées. En particulier, il s'agit d'interdire l'utilisation de la reconnaissance des émotions et de la détection des maladies dans le cadre de l'authentification par la voix et pour la sélection des candidatures par des particuliers, et d'interdire l'analyse de l'atten-

tion dans les écoles. Certaines applications souhaitables, classées comme applications à haut risque, pourraient donc également être autorisées sous condition.

En outre, soulignons **que les caractéristiques biométriques, et notamment la voix, ne doivent pas être utilisées comme (seul) facteur d'authentification**. En effet, contrairement aux formes d'authentification modifiables par nature (comme les mots de passe), les caractéristiques physiques ne peuvent plus être modifiées une fois qu'elles ont été compromises.

Les autres recommandations se subdivisent entre recommandations spécifiques à chaque domaine d'application concerné et recommandations générales concernant l'ensemble des domaines d'application. Ces dernières sont les suivantes :

- **Réglementation des applications à haut risque** requise en raison du risque élevé qu'elles représentent pour les droits fondamentaux et la cohabitation sociale, notamment dans le domaine de la santé, des enquêtes criminelles, du crédit, des assurances et de l'environnement de travail. Cette réglementation devrait inclure une gestion des risques, notamment une évaluation indépendante de la fiabilité technique et organisationnelle, des obligations de documentation, une obligation de supervision humaine, des exigences en matière de sécurité des données et des mécanismes de contrôle indépendants.
- **Bases légales explicites** requises **avant toute utilisation par des services publics** – car soit les données personnelles traitées sont sensibles, soit elles donnent lieu à un profilage. Le caractère nécessaire du recours à ces technologies doit dans tous les cas être vérifié avant l'élaboration de bases légales.
- **Formation adéquate et formation continue** du personnel qui utilise des applications de reconnaissance de la voix, de la parole et du visage, qui est responsable de leur fonctionnement ou qui est chargé de vérifier et de valider les décisions automatisées, ce de manière à permettre une utilisation critique et responsable de ces technologies.
- **Guides pour les exploitants d'applications de reconnaissance de la voix, de la parole et du visage** afin de les aider à utiliser leurs systèmes en conformité avec la législation sur la protection des données.
- **Aide aux personnes concernées et amélioration de l'application de la loi dans leur intérêt** afin de les aider à exercer leurs droits en tant que personnes concernées – y compris, le cas échéant, à utiliser les applications de reconnaissance de la voix, de la parole et du visage – et à mieux les protéger contre toute conséquence négative.
- **Débat de société sur les avantages et les inconvénients** des technologies de reconnaissance de la voix, de la parole et du visage et sur la manière de les utiliser. Doivent notamment faire l'objet d'un débat : le risque d'une surveillance de masse par l'État, la réglementation de l'émission de publicités politiques, la nécessité de négocier des normes sociales dans le contexte de l'utilisation de haut-parleurs intelligents ou la gestion sociale des défis de l'identification de toute personne.

- **Ressources suffisantes mises à la disposition** de tiers de confiance et de journalistes afin de leur permettre de remplir leur mission de soutien et d'information des personnes concernées et de la société en général.

L'équipe de projet espère que la présente étude donnera une impulsion au débat public et politique qui s'impose d'urgence et contribuera ainsi à une utilisation responsable des technologies de reconnaissance de la voix, de la parole et du visage. Il nous semble important de **ne pas exagérer ni méconnaître les opportunités et les défis de ces technologies**. Comme les premières applications sont déjà en place et que d'autres suivront bientôt, il semble nécessaire de **parler des limites** à fixer, qui, selon nous, sont nécessaires partout où des asymétries de pouvoir insurmontables apparaissent, où des atteintes injustifiées aux droits fondamentaux ont lieu, et où la technologie sert à poursuivre des objectifs considérés comme répréhensibles par la société.

Sintesi

Negli ultimi anni lo sviluppo di tecnologie di riconoscimento vocale (della voce e della parola) e facciale ha fatto grandi progressi. Sempre più spesso, nuove applicazioni basate su tali tecnologie fanno la loro comparsa sul mercato suscitando l'interesse sia delle imprese sia delle autorità. Da un lato, applicazioni innovative dovrebbero comportare benefici per il singolo e per la società. Dovrebbero contribuire ad esempio a facilitare compiti quotidiani degli utenti, a diagnosticare precocemente malattie e a cercare rapidamente corrispondenze in grandi banche dati di immagini nell'ambito delle indagini di polizia. Dall'altro, negli ultimi anni in particolare l'impiego del riconoscimento facciale da parte della polizia è stato preso di mira da numerose campagne di protesta della società civile a causa del timore di una sorveglianza di massa senza motivo. Vi sono però anche altre possibilità d'impiego controverse, perché le loro previsioni sono considerate tecnicamente soggette a errore o socialmente indesiderate: si pensi in particolare alla possibilità di dedurre caratteristiche di una persona (sesso, orientamento sessuale, stato di salute, emozioni ecc.) partendo da aspetti fisici esteriori.

Malgrado i progressi tecnologici, le applicazioni basate sul riconoscimento vocale e facciale si trovano ancora in uno stadio precoce: spesso non è chiaro quanto siano tecnicamente affidabili né se soddisfino i requisiti giuridici e quali sfide sociali possa eventualmente comportare il loro impiego.

Nel contesto di questa confusione e delle crescenti discussioni sociali, il presente studio si prefigge di mettere a disposizione conoscenze fondate sull'uso delle tecnologie al fine di promuovere il dibattito sociale e politico sulle opportunità e sulle sfide delle tecnologie di riconoscimento vocale e facciale.

Analisi di determinati campi di applicazione

Lo studio è imperniato sull'analisi di otto esempi di campi di applicazione.

1. Gli **smart speaker (o altoparlanti intelligenti)** sono sempre più presenti anche nelle case degli svizzeri. Siccome dovrebbero fungere da assistenti digitali, in genere gli smart speaker sono piazzati in soggiorno e quindi al centro della vita privata di molte persone.

2. Da qualche anno in Svizzera la **polizia** sfrutta il **riconoscimento facciale**. Quest'ultimo serve quale strumento di supporto tecnico ed è impiegato ex-post nell'ambito della ricerca di persone per confrontare immagini scattate sul luogo del delitto con fotografie salvate nelle banche dati della polizia. In alcuni Stati il riconoscimento facciale è utilizzato anche in tempo reale: le immagini delle videocamere di sorveglianza, posizionate in genere in luoghi pubblici, nelle stazioni ecc., sono confrontate continuamente con le banche dati della polizia per cercare persone scomparse, persone in fuga ecc.

3. L'**autenticazione vocale** è impiegata in Svizzera da alcune banche nell'ambito del banking telefonico per sostituire l'autenticazione mediante la richiesta di dati personali con il riconoscimento biometrico della voce del cliente.

4. Da decenni, la **prevenzione e il contrasto della violenza negli stadi** mediante il riconoscimento facciale sono uno dei punti particolarmente controversi nei dibattiti sul riconoscimento facciale. In Svizzera, vari tentativi di una loro introduzione sono falliti, nel frattempo trovano però impiego in molti Stati del mondo.

5. Il **riconoscimento di malattie fisici e mentali** partendo da registrazioni della voce o del volto è ancora perlopiù in fase di sviluppo, ma sul mercato compaiono già prime applicazioni, che dovrebbero consentire il riconoscimento (precoce) di diversi disturbi o malattie, come ad esempio il Parkinson, l'Alzheimer, l'autismo o la depressione. È previsto un uso di queste applicazioni anche per sostenere le terapie.

6. Il **riconoscimento delle emozioni** basato su registrazioni della voce, delle parole e del volto è offerto sempre più spesso quale funzione supplementare, oltre all'identificazione e all'autenticazione delle persone. Al riconoscimento delle emozioni è attribuito un grande potenziale in particolare nell'ambito del marketing e nei processi di candidatura sul mercato del lavoro: gli utenti si aspettano infatti conoscenze sui «veri» desideri, sulle «vere» capacità, ecc. delle persone esaminate.

7. L'**analisi dell'attenzione** è un altro campo di applicazione del riconoscimento facciale. Benché in Svizzera non trovi ancora impiego, in altri Paesi è già utilizzato ad esempio per registrare e valutare l'attenzione dei conducenti di auto o degli allievi a scuola.

8. L'**identificazione a tappeto** designa la possibilità, ancora fittizia, di identificare, mediante fotocamere in miniatura, montate ad esempio sugli smart glasses (occhiali intelligenti), tutte le persone presenti nel campo visivo e ottenere ulteriori informazioni (sensibili) su di esse.

Questi campi di applicazione rappresentano uno spaccato delle **applicazioni già in uso** oggi (*smart speaker*, *riconoscimento facciale da parte della polizia*, *autenticazione vocale*), di quelle che potrebbero trovare **impiego a medio termine** (*prevenzione e contrasto della violenza negli stadi*, *riconoscimento di malattie fisici e mentali* o *riconoscimento delle emozioni*) e di quelle che s'iscrivono in **scenari per ora ipotetici**, che tuttavia avrebbero ampie implicazioni (*analisi dell'attenzione* e *identificazione a tappeto*). L'analisi dei campi di applicazione del riconoscimento vocale e facciale menzionati si articola su **quattro grandi tematiche**:

- analisi delle basi e delle possibilità tecniche,
- valutazione giuridica, in particolare delle condizioni quadro in materia di protezione dei dati,
- sfide etiche e sociali,
- percezione sociale delle opportunità e dei rischi e auspici.

Basi e possibilità tecniche attuali

L'analisi delle basi e delle possibilità tecniche evidenzia un **ampio spettro di maturità tecnologica**: alcune applicazioni funzionano già in modo sostanzialmente affidabile sul piano tecnico (*smart speaker*, *autenticazione vocale*) o potrebbero raggiungere una **buona affidabilità tecnica** nei prossimi anni quale risultato degli attuali sforzi di sviluppo (*riconoscimento facciale ex-post da parte della polizia*, *prevenzione e contrasto della violenza negli stadi*). Le applicazioni nei settori del *riconoscimento delle malattie* e del *riconoscimento delle emozioni* sono invece confrontate a sfide in parte più complesse, che ne **ostacolano un uso affidabile**. In questi settori bisognerà intensificare gli sforzi a livello delle basi concettuali e del miglioramento delle basi di dati. Per quanto riguarda i **tassi di riconoscimento** del riconoscimento vocale e facciale occorre tener conto soprattutto del contesto tecnico-sociale in cui avviene il riconoscimento: i sensori di una videocamera all'avanguardia collocata in un ambiente controllato, ben illuminato e in cui le persone si comportano in modo cooperativo promettono ad esempio tassi di successo nettamente superiori rispetto al riconoscimento del volto di una persona in movimento, registrato da lontano con una cattiva luce o sensori di scarsa qualità. La valutazione dell'affidabilità tecnico-organizzativa di un'applicazione svolge quindi un ruolo fondamentale. A tal fine sarebbe necessaria in particolare una **verifica indipendente dei dati forniti dal fabbricante sui tassi di successo del software**. Proprio questa valutazione è tuttavia difficile soprattutto nei casi in cui i responsabili dell'uso di una tecnologia sono attori privati. A parte gli interessanti studi del National Institute of Standards and Technology (NIST), le indagini scientifiche sugli algoritmi di riconoscimento o sui tassi di successo sono infatti rare. Sono inoltre di norma soggette a limitazioni per il fatto che sono condotte in ambienti sperimentali, trascurando l'effettivo contesto tecnico-sociale di utilizzazione, e di conseguenza sono poco rappresentative per le applicazioni pratiche, caratterizzate da sensori e condizioni ambientali variabili. I **veri tassi di successo delle applicazioni si nascondono quindi dietro una giungla di promesse di marketing, segreti aziendali e requisiti di sicurezza**. L'affidabilità tecnica dei sistemi è compromessa anche dalla **problematica dei** bias: siccome lavorano con dati e processi forniti ed elaborati da esseri umani, le applicazioni basate sul riconoscimento vocale e facciale conterranno sempre un *bias*, che potrà essere ridotto, ma che probabilmente non potrà mai essere eliminato completamente.

Vi è poi un problema di fondo per i casi in cui mancano categorie scientificamente fondate, com'è il caso in parte per il riconoscimento delle malattie, ma soprattutto per il riconoscimento delle emozioni. Siccome le emozioni non possono essere delimitate – e quindi neanche categorizzate – chiaramente e la loro intensità è molto variabile, il calcolo di tassi di successo è problematico: qualsiasi attribuzione categoriale si basa infatti su valori di probabilità in relazione alla categoria corrispondente. In questi campi saranno quindi necessari ulteriori lavori di ricerca sulle basi concettuali.

L'analisi delle basi e delle possibilità tecniche evidenzia pertanto che, anche in futuro, anche i sistemi più affidabili basati sulle tecnologie di riconoscimento vocale e facciale **non raggiungeranno mai un'affidabilità tecnica del 100 per cento**.

Condizioni quadro giuridiche

Nella maggior parte dei casi considerati in questa sede, il volto e la voce di una persona rappresentano **dati biometrici** e quindi **dati personali degni di particolare protezione**. Si tratta di caratteristiche legate strettamente e durevolmente alla persona, che rappresentano aspetti unici della personalità. Il trattamento di questi dati costituisce in ogni caso una lesione della personalità o una grave ingerenza nei diritti fondamentali del singolo.

L'uso di tecnologie di riconoscimento vocale o facciale da parte delle **autorità** tange il **diritto alla sfera privata** e la **protezione da un impiego abusivo dei dati personali** (art. 13 cpv. 1 e 2 Cost.). Il ricorso a questa tecnologia può avere un effetto dissuasivo (*chilling effect*) sull'esercizio di diversi diritti fondamentali in materia di comunicazione (libertà di opinione, di riunione e di associazione, art. 16, 22 e 23 Cost.). Possono risentirne anche i diritti fondamentali procedurali (art. 29, 29*a* e 30 Cost.) nonché il divieto della discriminazione (art. 8 cpv. 2 Cost.). Un impiego di queste tecnologie nel rispetto dei diritti fondamentali presuppone un accurato esame preliminare da parte del legislatore, tenuto a creare le **basi legali** necessarie, di norma in una **legge formale**, **impostandole in modo sufficientemente preciso** per potervi riconoscere non solo uno scopo inequivocabile, ma anche tra l'altro una **limitazione** del trattamento dei dati allo stretto necessario, sia a livello di tempi sia a livello di portata. La **proporzionalità** dell'uso di queste tecnologie deve essere esaminata dalle autorità che applicano il diritto sia nel quadro dell'elaborazione delle basi legali sia prima dell'impiego concreto, mettendo spesso in dubbio, in base allo stato attuale della tecnica, l'**idoneità** dell'impiego di queste tecnologie: come ha infatti rivelato il presente studio, **molto spesso l'affidabilità della tecnologia non è garantita** (in particolare nell'ambito del riconoscimento delle emozioni). Spesso potrebbe non essere garantita neanche la **necessità**, essendo disponibili soluzioni più «morbide» per raggiungere gli interessi pubblici perseguiti con l'impiego di tecnologie di riconoscimento vocale o facciale. Determinate applicazioni violano inoltre l'**essenza**, assolutamente protetta, di vari diritti fondamentali e sono quindi sempre inammissibili. Vi rientrano in particolare l'uso della tecnologia di riconoscimento facciale per la **sorveglianza a tappeto in tempo reale** nonché il ricorso al **riconoscimento delle emozioni** da parte dello Stato.

Se le tecnologie di riconoscimento vocale o facciale sono utilizzate da **privati**, di norma vengono violati diversi principi della protezione dei dati, imponendo una giustificazione del trattamento dei dati secondo l'articolo 31 della nuova legge sulla protezione dei dati (LPD). Nella maggior parte dei casi, un interesse preponderante al trattamento quale giustificazione dovrebbe essere escluso a causa del carattere degno di particolare protezione dei dati biometrici trattati. Nell'ambito del riconoscimento vocale e facciale, il **consenso** dei diretti interessati è però spesso problematico, essendo **praticamente impossibile** da dare o richiedere (p. es. nel caso degli smart speaker o degli smart glasses e di altre applicazioni a tappeto), **non essendo facoltativo** a causa delle **asimmetrie di potere** (nelle procedure di candidatura, nell'ambito delle assicurazioni o nelle scuole private), non essendoci alternative che richiedono meno dati (in caso di uso negli stadi o da parte delle banche) oppure non potendo essere basato su un'**adeguata informazione** preliminare.

Siccome l'impiego da parte di attori privati cela numerosi rischi per i diretti interessati e siccome nei campi di applicazione analizzati spesso gli utilizzatori o gli operatori non agiscono

in modo conforme alla legge, nell'ottica dei diritti fondamentali lo **Stato** è tenuto a garantire una **protezione contro le violazioni dei diritti fondamentali da parte di privati** (art. 35 Cost.). Questo obbligo di protezione presuppone che lo Stato **disciplini l'uso di queste tecnologie da parte dei privati e vigili sul rispetto delle disposizioni**. Il team di progetto ha formulato raccomandazioni in proposito.

Dal punto di vista del diritto sulla protezione dei dati va infine menzionato anche il principio della **sicurezza dei dati**, che deve essere rispettato sia dalle autorità sia dai privati. Siccome la perdita, l'impiego abusivo o cambiamenti di destinazione dei dati biometrici da parte di persone non autorizzate comportano gravi pericoli, occorre fissare requisiti severi per le misure volte a garantire la sicurezza dei dati.

Sfide etiche e sociali

Anche se sono soddisfatti tutti i requisiti giuridici (compresi quelli in materia di protezione dei dati), le applicazioni del riconoscimento vocale e facciale possono comportare altre sfide etiche rilevanti. Tra di esse figurano gli **effetti dissuasivi della sorveglianza**, che compromettono l'esercizio di diritti fondamentali. E anche se ad esempio l'impiego del riconoscimento facciale da parte della polizia è conforme alla legge, resta sempre la possibilità dell'**introduzione strisciante di una sorveglianza statale completa** in un numero crescente di ambiti della vita. Potrebbero però verificarsi modifiche dei comportamenti o un conformismo internalizzato anche in altri ambiti. In particolare per determinate forme di riconoscimento delle emozioni, i diretti interessati sarebbero chiamati a scegliere tra piegarsi alla pressione del riconoscimento e comportarsi in modo il più possibile conforme alla norma e quindi vantaggioso – può servire un sorriso di circostanza? – oppure accettare degli svantaggi. Siccome occorre partire dal presupposto che non sarà possibile raggiungere un'affidabilità tecnica del 100 per cento, è presumibile che, anche in futuro, malgrado i migliori tassi di riconoscimento, non mancheranno i **risultati falsi positivi e falsi negativi**, con **conseguenze indesiderate**, da un arresto ingiustificato nel caso del riconoscimento facciale da parte della polizia al pericolo di accesso illegale al conto nel caso del banking telefonico o ancora a diagnosi (o autodiagnosi) false negative in ambito medico, che potrebbero avere conseguenze letali.

Già oggi i servizi statali ed economici dispongono di nettamente più informazioni sulla popolazione rispetto a fino pochi anni o decenni fa. Nella discussione sulle sfide etiche, per molti campi di applicazione è emerso che, con il crescente ricorso a tecnologie di riconoscimento vocale e facciale, presumibilmente questo **divario di conoscenze** e quindi anche **di potere** si allargherà ulteriormente a favore di chi tratta i dati. Benché nel singolo caso le leggi (comprese quelle sulla protezione dei dati) offrano una protezione da possibilità troppo invasive di abusare di questo potere crescente, le leggi sulla protezione dei dati non possono fermare gli sviluppi di per sé. Restano inoltre **numerose possibilità di manipolazione del comportamento individuale**: i modelli predittivi (con tutta probabilità le persone che fanno A e B fanno/pensano/provano anche X) si basano ad esempio sull'analisi di dati anonimi (o anonimizzati) di altre persone, ma in combinazione con i dati personali di una determinata persona possono essere utilizzati per trarre conclusioni sensibili su di essa.

Nell'ambito del diritto vigente, proprio il **consenso richiesto dal diritto della protezione dei dati** rappresenta una particolare sfida, poiché in situazioni di estremo squilibrio di potere può dare l'illusione di un controllo. Se nel caso del riconoscimento di un volto, dell'interpretazione delle emozioni, dei comportamenti ecc. da parte di una persona è relativamente chiaro come tali elementi siano percepiti da tale persona, le tecnologie di riconoscimento vocale e facciale consentono di ottenere **informazioni molto più dettagliate sui soggetti in questione**. Le persone che non sono in chiaro su questo potenziale potrebbero **acconsentire alla leggera al trattamento dei loro dati**. Nel caso del riconoscimento delle emozioni, persone convinte che la tecnologia non è infallibile potrebbero acconsentire alla sua utilizzazione cercando di superarla in astuzia adattando il loro comportamento a proprio vantaggio, senza pensare che potrebbero essere smascherate e penalizzate. Al tempo stesso, l'esempio del riconoscimento delle emozioni evidenzia un'altra cosa: proprio come una tecnologia non infallibile comporta dei pericoli, si delineano **sfide da non sottovalutare** anche nel caso di una **tecnologia sostanzialmente sicura sul piano tecnico**. Una tecnologia sempre più perfezionata acutizzerà con particolare urgenza pericoli come una sorveglianza di massa da parte dello Stato o la possibilità etichettata come «identificazione a tappeto», che consente di identificare «passanti» sconosciuti.

A medio-lungo termine, i produttori di tecnologie stanno pianificando l'integrazione di tecnologie di riconoscimento vocale e facciale in **sistemi sociotecnologici convergenti**, ad esempio il ricorso al riconoscimento delle emozioni nell'ambito del banking telefonico e dei colloqui di candidatura oppure la loro installazione negli smart speaker. Sulla scia di questa evoluzione, la quantità e soprattutto la qualità dei dati raccolti sul comportamento delle persone segnerebbero un'impennata. Le tendenze alla quantificazione e al controllo del comportamento e della vita degli esseri umani finirebbero con il conquistare gli ultimi settori non ancora digitalizzati. Sistemi così avanzati, soprattutto diagnostici o di assistenza, offrirebbero grandi vantaggi come strumenti di sostegno nella vita di tutti i giorni. In seguito a questi *feedback loop* (cicli di retroazione) tecnologici, la vita umana verrebbe però sempre più controllata dai calcoli e dalle routine della tecnologia. Se fosse garantito un controllo democratico degli algoritmi alla base di questi processi, ciò potrebbe benissimo aumentare l'autodeterminazione sociale e individuale. In caso di controllo algoritmico poco trasparente, l'autodeterminazione potrebbe invece rivelarsi una farsa, dal momento che le decisioni sulle questioni importanti, comprese quelle più private, sarebbero dettate da questi algoritmi.

Un'altra sfida importante è legata ai guadagni in efficienza spesso presentati come un grande vantaggio del riconoscimento vocale e facciale automatizzato. È lecito chiedersi se i guadagni in efficienza resi possibili si tradurranno effettivamente in un miglioramento della qualità del lavoro della polizia, in una diminuzione dei prezzi per accedere agli stadi o in un miglioramento delle cure ai pazienti o se gli effetti di razionalizzazione economica non porteranno invece a **tagli del personale** e a una **distribuzione unilaterale di eventuali utili**.

Le riflessioni sulle sfide etiche evidenziano in particolare la **necessità di un dibattito sociale urgente**, che in ultima analisi dovrà ruotare sul come conciliare i potenziali del progresso tecnologico con un sistema politico democratico, nel rispetto dei diritti fondamentali.

Percezione pubblica

Nell'ambito dell'analisi della percezione pubblica è emerso che i cittadini sanno assolutamente riconoscere e apprezzare le **opportunità offerte dalle tecnologie**. Tra i vantaggi hanno identificato l'atteso **guadagno in termini di sicurezza, salute** e **comodità**. Si aspettano però anche **aumenti dell'efficienza**, che consentano di destinare le risorse umane o finanziarie liberate ad altri impieghi intelligenti. Nella valutazione positiva svolge un ruolo importante anche l'idea, non necessariamente assodata, che l'uso delle tecnologie potrà ridurre le discriminazioni fabbricate dall'uomo e che i responsabili dell'uso delle tecnologie saranno entità degne di fiducia.

Al tempo stesso non manca una certa **diffidenza** nei confronti delle tecnologie e in particolare di chi le gestisce**:** a prevalere è il **timore di un uso non trasparente della tecnologia** e **dell'abuso dei dati personali**, come l'intercettazione dei dati di accesso al conto o la sorveglianza di massa senza motivo. Le persone interpellate temono inoltre che le tecnologie di riconoscimento vocale e facciale possano funzionare in modo non abbastanza affidabile, con il rischio di conseguenze indesiderate, come un controllo o un arresto in seguito a un **riconoscimento erroneo** oppure **errori di diagnosi** fatali nell'ambito delle applicazioni mediche. Molte delle persone interpellate si sono inoltre dette preoccupate per il fatto che le tecnologie potranno essere impiegate anche contro i loro interessi (pubblicità indesiderata, discriminazione di prezzo ecc.).

L'analisi dell'opinione pubblica evidenzia però anche notevoli **lacune di conoscenze** e di conseguenza un **bisogno d'informazione**. È quanto emerge anche dal sostegno o dal rifiuto di diverse applicazioni: l'identificazione a tappeto e l'analisi dell'attenzione nelle scuole sono chiaramente respinte, mentre l'uso del riconoscimento facciale negli stadi è chiaramente sostenuto. Per tutte le altre applicazioni spicca invece una **quota elevata di indecisi**, che sottolinea l'importanza del dibattito pubblico sulle relative tematiche.

I cittadini ritengono che spetti allo Stato e ai gestori dei servizi garantire un uso affidabile delle tecnologie di riconoscimento vocale e facciale. Si fa riferimento alla **necessità di un'ampia regolamentazione della trasparenza, divieti, campagne d'informazione, una supervisione e una valutazione degli usi della tecnologia** da parte di esperti indipendenti, test preliminari dell'efficacia e un obbligo di certificazione nonché al fatto che l'impiego da parte dello Stato deve fondarsi su basi legali il più possibile concrete.

Raccomandazione

I risultati dello studio evidenziano la **necessità di agire**. Come possono far fronte alle molteplici sfide delle tecnologie di riconoscimento vocale e facciale i decisori della politica, dell'economia, delle autorità di protezione dei dati, dei media e della società civile? Formuliamo delle opzioni. In base ai risultati del presente studio sembrano adeguati in particolare un **divieto** in tre settori o una **moratoria** in un settore:

- **divieto della sorveglianza statale automatizzata in tempo reale e del** social scoring **(credito sociale) statale mediante riconoscimento vocale o facciale**;

- **divieto delle decisioni completamente automatizzate basate su tecnologie di riconoscimento vocale e facciale in importanti ambiti della vita** (p. es. nel campo della salute, del perseguimento penale, delle finanze e della concessione di crediti, delle assicurazioni o del mondo del lavoro) fintanto che non saranno garantiti sufficienti livelli di affidabilità ed equità tecnica e organizzativa – i risultati dei sistemi di supporto alle decisioni parzialmente automatizzati dovrebbero essere verificati con spirito critico e approvati da personale qualificato;
- **divieto dell'uso di smart glasses** e di altre tecnologie non riconoscibili direttamente, che si prestano alla sorveglianza **in pubblico**, come piccole videocamere abbinate alla tecnologia di riconoscimento facciale;
- **moratoria per il riconoscimento delle emozioni e delle malattie in importanti ambiti della vita** (p. es. nel campo del perseguimento penale, delle finanze e della concessione di crediti, delle assicurazioni o del mondo scolastico e lavorativo) – in particolare divieto dell'uso del riconoscimento delle emozioni e delle malattie nell'ambito dell'autenticazione tramite la voce e della selezione di candidati da parte di privati nonché divieto del riconoscimento dell'attenzione nelle scuole. Determinate applicazioni indesiderate potrebbero anche essere classificate come applicazioni ad alto rischio e quindi essere autorizzate solo a determinate condizioni.

Attiriamo inoltre l'attenzione sul fatto che le **caratteristiche biometriche e in particolare la voce non dovrebbero essere utilizzate come fattore di autenticazione (unico)**. A differenza delle caratteristiche di autenticazione modificabili (come le password), le caratteristiche fisiche non possono infatti essere modificate una volta che sono state violate.

Le altre raccomandazioni si suddividono in raccomandazioni specifiche per determinati campi d'applicazione e raccomandazioni generali. Queste ultime sono le seguenti.

- **Disciplinamento delle applicazioni ad alto rischio**, in particolare nel campo della salute, del perseguimento penale, della concessione di crediti, delle assicurazioni o del mondo del lavoro, poiché celano un alto rischio per i diritti fondamentali e la convivenza sociale. La regolamentazione dovrebbe comprendere una gestione dei rischi, compresi in particolare una valutazione indipendente dell'affidabilità tecnico-organizzativa, obblighi di documentazione, un obbligo di supervisione umana, prescrizioni sulla sicurezza dei dati nonché meccanismi di controllo indipendenti.
- **Basi legali esplicite per l'uso da parte di servizi pubblici**, visto che sono trattati dati personali degni di particolare protezione o si effettua un profiling. Al di là dall'obbligo di meccanismi volti a garantire lo Stato di diritto occorre prevedere, nell'ambito della procedura legislativa, una verifica della necessità dell'uso della tecnologia.
- **Formazione e perfezionamento adeguati** del personale che utilizza le applicazioni di riconoscimento vocale e facciale, è responsabile del loro impiego o è incaricato di verificare e approvare le decisioni automatizzate, in modo da consentire un uso critico e responsabile delle tecnologie.
- **Guide per i gestori di applicazioni di riconoscimento vocale e facciale** in modo da aiutarli a utilizzare i loro sistemi conformemente al diritto sulla protezione dei dati.

- **Sostegno dei diretti interessati e miglioramento dell'applicazione del diritto nel loro interesse**, per aiutarli a far valere i loro diritti ed eventualmente anche a utilizzare correttamente le applicazioni di riconoscimento vocale e facciale nonché per tutelarli meglio dalle conseguenze negative.
- **Dibattito sociale sui vantaggi e gli svantaggi** delle tecnologie di riconoscimento vocale e facciale e sul loro impiego, compresi il pericolo di una sorveglianza di massa statale, la regolamentazione della diffusione di pubblicità politica, la necessità di negoziare norme sociali nel contesto degli smart speaker o la gestione sociale delle sfide dell'identificazione a tappeto.
- **Messa a disposizione di sufficienti risorse** per terzi di fiducia e rappresentanti dei media, in modo che possano adempiere il loro mandato di sostegno e informazione dei diretti interessati e della società.

Con il presente studio, gli autori si augurano di dare il via al dibattito pubblico e politico urgente, contribuendo così a un uso responsabile delle tecnologie di riconoscimento vocale e facciale. Ai nostri occhi è importante **non sopravvalutare né sottovalutare le opportunità e le sfide di queste tecnologie**. Siccome prime applicazioni sono già in uso e altre seguiranno a ritmo serrato è necessario parlare in particolare dei **limiti da tracciare**, limiti che secondo noi sono necessari ovunque si delineano asimmetrie di potere insormontabili, si verificano ingerenze ingiustificate nei diritti fondamentali o la tecnologia è utilizzata per perseguire scopi condannati dalla società.

1. Einleitung und Kontext

“There are two ways that this technology can hurt people: One way is by not working. The second situation is when it does work – where you have the perfect facial recognition system, but it’s easily weaponized against communities to harass them.”

Deborah Raji

1.1. Hintergrund und Zielsetzung der Studie

Die Digitalisierung von immer mehr Bereichen des alltäglichen Lebens schreitet weiterhin voran. Entsprechend waren die vergangenen Jahre von Themen wie Big Data, Internet der Dinge, k**ünstliche Intelligenz und** digitale Desinformation geprägt. Nachdem – anders als andere biometrische Verfahren – die Stimm- und Spracherkennung sowie die Gesichtserkennung über viele Jahre nur ein Nischenthema waren, erleben diese Technologien seit geraumer Zeit einen Aufschwung, der zu einer verstärkten Debatte in Wissenschaft und (Fach-)Öffentlichkeit geführt hat. So galt etwa die computerisierte Erkennung von Stimme und Sprache trotz intensiver Forschung und Entwicklung über viele Jahre hinweg als nicht ausreichend zuverlässig (siehe z.B. Shneiderman 2000). Mit Fortschritten im Bereich des Maschinenlernens fand die Stimm- und Spracherkennung schliesslich vor einigen Jahren zunächst ihren Weg auf Smartphones in Form von Sprachassistenten – und nur etwas später erfolgte der Durchbruch in Gestalt smarter Lautsprecher. Derzeit dominiert bei Sprachassistenten zwar noch die Nutzung für vergleichsweise rudimentäre Zwecke, doch zeichnet sich angesichts der zunehmenden Konvergenz digitaler Geräte und des allmählichen **Übergangs von visuellen und haptischen Interfaces zu Sprachinterfaces eine weitere Revolution der Techniknutzung ab**. Vielfältige neue Nutzungsmöglichkeiten der zugrunde liegenden Spracheingaben für private wie auch öffentliche Stellen werden voraussichtlich Ergebnis dieses Umbruchs sein. Schon heute kommen Stimm- und Spracherkennung auch in der Schweiz in verschiedenen Bereichen des gesellschaftlichen Lebens zum Einsatz, etwa zur umstrittenen Authentifizierung des Sprechers mittels Stimmerkennung beim Telefonbanking (Emmenegger 2019) oder in Strafverfahren zum Zwecke der Identifizierung von Tatverdächtigen (Urech 2019). Zudem stehen erste Smartphone-Apps zur Erkennung von Krankheiten mittels Stimm- und Sprachanalyse einer internationalen Nutzerschaft unabhängig vom Wohnort zum Download zur Verfügung (Latif et al. 2020, S. 346). Die weltweit rasant zunehmende Verbreitung smarter Lautsprecher blieb hingegen aus, was auf den verspäteten Markteintritt der grossen Anbieter einerseits und auf Schwierigkeiten bei der Erkennung Schweizer Dialekte andererseits zurückgeführt wurde (Städeli 2019). Für die nächsten Jahre wird erwartet, dass Stimm- und Spracherkennung zunehmend miteinander und mit weiteren Systemen zusammengeführt werden und so bspw. in Bewerbungsverfahren zur Emotionserkennung (Linhart 2019) und zur Authentifizierung von Mitarbeitenden sowie die Analyse ihres Gemütszustands eingesetzt werden (doitvoluntarily 2018). In Italien wurde auch die Erkennung von rassistischen Sprechchören mittels Spracherkennung und

die anschliessende Identifikation der Sprecher mittels Gesichtserkennung geplant (Chiusi 2020).

Während die Verbreitung von Anwendungen zur Erkennung von Stimme und Sprache voranschreitet und ihr mit vergleichsweise wenig öffentlicher Aufmerksamkeit und Kritik begegnet wurde, stand die biometrische Gesichtserkennung im Mittelpunkt zahlreicher internationaler Debatten der vergangenen Jahre. Angefangen mit dem erfolgreichen Widerstand gegen die Einführung der biometrischen Gesichtserkennung bei Facebook und der Datenbrille «Google Glass» vor rund zehn Jahren rückte in den letzten Jahren insb. der Widerstand gegen den staatlichen Einsatz von Gesichtserkennungssystemen in den Mittelpunkt der öffentlichen Debatte. Einerseits richtete sich die Kritik dabei auf den diskriminierenden Einsatz der Systeme gegen Minderheiten und insb. gegen Dunkelhäutige, weil diese **überproportional** (anders als Menschen hellerer Hautfarbe) nicht korrekt erkannt bzw. fälschlicherweise als gesuchte Personen erkannt und in der Folge polizeilichen Massnahmen ausgesetzt wurden (Singer und Metz 2019). Wenig später wurde bekannt, dass das erst 2017 gegründete US-amerikanische Unternehmen Clearview AI auf u.a. vorgeblich illegale Weise Milliarden Bilder aus dem weltweiten Internet abgegriffen, in seine Gesichtersuchmaschine integriert und diesen Dienst weltweit an staatliche und private Stellen verkauft bzw. zum Testen angeboten hatte – ohne dass dies zuvor in der Öffentlichkeit der jeweiligen Staaten diskutiert worden war. Der Clearview AI-Fall stellte allerdings nur eine besonders brisante Form des polizeilichen Einsatzes von Gesichtserkennung dar. Auch abseits dessen initiierten bzw. intensivierten verschiedene Staaten den Einsatz entsprechender Systeme, die zum Teil von autoritären Regimen stammen (Big Brother Watch 2018; Feldstein 2019). Wohin der unkontrollierte staatliche Einsatz von Gesichtserkennung führen kann, verdeutlicht deren Einsatz in autoritären Staaten. Während sie in Russland zur Identifikation und Verhaftung von Demonstranten verwendet wird (Rebiger 2017), plant die Volksrepublik China die schrittweise Einführung eines sog. Sozialkreditsystems, das Gesichtserkennungsdaten mit Daten aus möglichst vielen anderen Quellen zusammenführen und so eine möglichst lückenlose staatliche Kontrolle des Verhaltens aller Bürgerinnen und Bürger ermöglichen soll (Hoffman 2017).

In der Schweiz kommt Gesichtserkennung bspw. bereits seit 2017 zur automatisierten Passkontrolle bei der Ausreise (Kanton Zürich 2018) und seitens mehrerer Kantonspolizeien zu Strafverfolgungszwecken (Luchetta 2021a) zur Anwendung. Abseits der verbreiteten Nutzung zur Entsperrung von Smartphones oder in Gesichtsfilter-Apps ist die Gesichtserkennung im privatwirtschaftlichen Bereich hingegen noch kein Thema. Einsätze zu Zwecken der Einlasskontrolle und Identifikation von Gewalttätern in Sportstadien (Knupfer 2020b) oder für Marketingzwecke im Detailhandel wurden zwar diskutiert (Benjamin Weinmann 2019), konnten sich bislang jedoch nicht durchsetzen. International wird Gesichtserkennung im Privatsektor derweil bereits für vielfältige Zwecke eingesetzt, so z.B. zur Unterstützung der Auswahl von Bewerberinnen und Bewerbern auf Grundlage von Emotionserkennung (Linhart 2019) oder zur Zahlung beim Einkaufen mittels Gesichtserkennung (Jacobs und Zheng 2018). Und selbst Privatpersonen können mittlerweile Gesichtserkennung einsetzen: Der deutsche Hersteller Bosch bietet über sein Start-Up Azena sog. «intelligente» Überwachungskameras zum Kauf an, auf deren App-Store «Ethinizitätserkennung, Geschlechtserkennung, Gesichtserkennung, Emotionserkennung und Erkennung verdächtigen Verhaltens» angeboten werden (Campbell und Jones 2022).

Generell ist festzustellen, dass sich die gesellschaftlichen Diskussionen und insb. die Kritik rund um Gesichtserkennung auf ihren unmittelbar staatlichen Einsatz fokussieren. Der fortschreitende Einsatz von Gesichtserkennung seitens staatlicher Stellen führte in westlichen Demokratien zu einem bis heute anhaltenden zivilgesellschaftlichen Widerstand und zu zahlreichen Kampagnen zum Verbot der automatisierten Gesichtserkennung. Im Fokus der Kritik steht dabei die Verletzung von Grundrechten und die Ablehnung gesellschaftlicher Massenüberwachung (EDRi 2020). Ein Ergebnis dieser Debatten war das Verbot des Einsatzes von Gesichtserkennung seitens der Stadt San Francisco in den Vereinigten Staaten (Reuter 2019a). Und auch auf EU-Ebene wurde zeitweise über ein generelles Moratorium debattiert, das letztlich aber zunächst im Kommissionsvorschlag zum sog. AI Act im vorgesehenen Verbot von biometrischer Echtzeitgesichtserkennung seitens staatlicher Stellen mündete (Veale und Zuiderveen Borgesius 2021, S. 4–8). Aus der Kritik am unregulierten Einsatz von Gesichtserkennung seitens Schweizer Polizeien initiierten die drei Non-Profit-Organisationen Amnesty International, AlgorithmWatch CH und Digitale Gesellschaft Ende 2021 eine Petition für ein Verbot des Einsatzes von automatischer Gesichtserkennung im öffentlichen Raum in der Schweiz. Die Petition erreichte zwar das Sammelziel, konkrete politische Konsequenzen aufgrund der Petition oder der vorangegangenen medialen Debatten sind bislang allerdings ausgeblieben (AlgorithmWatch CH et al. 2021).

1.2. Zielsetzung

Die vorliegende Studie verfolgt das Ziel, den Diskurs über die Stimm-, Sprach- und Gesichtserkennung (Funktionsweise der Technologien, Akteursüberblick, zentrale Anwendungsbeispiele sowie die gesellschaftliche Wahrnehmung der Materie) systematisch aufzubereiten, zu bewerten und fundiertes Orientierungswissen für Entscheidungstragende, aber auch für die Bevölkerung sowie Technologiehersteller und Diensteanbieter bereitzustellen. Die mit der Studie angestrebte Verbesserung der Informationslage der (politischen) Entscheidungstragenden soll schliesslich idealerweise in besser reflektierte (politische) Entscheidungen münden.

Die von uns vorgeschlagene Studie möchte zwischen Potenzialen und Herausforderungen vermitteln und auf Basis interdisziplinärer Forschung die verantwortungsvolle Nutzung dieser Technologien voranbringen.

1.2.1. Fragestellung der Studie

Die Fragestellung der Studie ist in drei Fragekomplexe untergliedert: (1) Überblick und Analyse von Stimm-, Sprach- und Gesichtserkennungstechnologien und -anwendungen sowie der Akteure, die sie einsetzen; (2) Bewertung von relevanten Anwendungsfällen und -gebieten; (3) Empfehlungen. Zu den unbeantworteten Fragen zählen aktuell insb. die folgenden:

Ist- und Trendanalyse: Technologie- und Anwendungsüberblick
FF 1.1: Welche Varianten der Stimm-, Sprach- und Gesichtserkennung existieren derzeit? D.h. insb., welche unterschiedlichen technologischen Grundlagen und Methoden existieren aktuell?

FF 1.2: Was können die entsprechenden Technologien grundsätzlich leisten? (Betrifft die Zuverlässigkeit und Aussagekraft der jeweiligen Technologievarianten, bspw. Emotionserkennung)

FF 1.3: Im Rahmen welcher (nationalen wie internationalen) Anwendungsfälle werden welche dieser Technologievarianten eingesetzt?

FF 1.4: Welche dieser Anwendungsfälle und -gebiete könnten zukünftig für die Schweiz von Relevanz sein?

Vertiefte Untersuchung von Anwendungsfeldern

FF 2.1: Wie sind die Anwendungsfälle aus technologischer Perspektive zu bewerten? Was können die jeweiligen Anwendungen wirklich leisten?

FF 2.2: Wie sind die relevanten Anwendungsfälle und -gebiete datenschutzrechtlich zu bewerten? Was ist unter welchen Bedingungen erlaubt?

FF 2.3: Was ist datenschutzrechtlich erlaubt, aber möglicherweise aus anderen Gründen, bspw. aus der Perspektive der Ethik, des Konsumenten- oder Diskriminierungsschutzes, in welcherlei Hinsicht problematisch?

FF 2.4: Welcher Informationsstand bezüglich der Nutzung, der Möglichkeiten und Risiken von Stimm-, Sprach- und Gesichtserkennungstechnologien liegt aufseiten der Bevölkerung vor?

FF 2.5: Wie werden die Nutzung, Möglichkeiten und Grenzen dieser Technologien, insb. bereits existierende bzw. künftig wahrscheinliche Anwendungsfälle und -gebiete, seitens der Bevölkerung bewertet? Was gilt als nützlich und wünschenswert und was als riskant?

Empfehlungen

FF 3.1: Wie sollte ein nachhaltiger und verantwortungsbewusster gesellschaftlicher Umgang mit Stimm-, Sprach- und Gesichtserkennungstechnologien aussehen?

FF 3.2: Wo besteht dringender Handlungsbedarf, wo optionale Handlungsspielräume? (stets mit Fokus auf (politische) Entscheidungstragende, aber auch in Richtung Technologiehersteller und Diensteanbieter sowie der Bevölkerung)

1.3. Wichtige Definitionen

Gesichtserkennung: Bei der Gesichtserkennung handelt es sich um ein biometrisches Verfahren zur Erkennung von Personen. Dieses kann sowohl zur Identifikation als auch zur Verifikation verwendet werden.

Stimmerkennung: Bei der Stimmerkennung handelt es sich um ein biometrisches Verfahren zur Erkennung von Personen. Dieses kann sowohl zur Identifikation als auch zur Verifikation verwendet werden.

Spracherkennung: Von der Stimmerkennung ist die Spracherkennung zu unterscheiden, weil bei Letzterer der Sprachinhalt erkannt wird.

Identifikation: Bei einer Identifikation wird ein Gesichtsbild oder eine Stimmaufzeichnung einer Person mit einem Datenpool aus Gesichtsbildern bzw. Stimmaufnahmen abgeglichen (1:N), um bspw. den Namen einer Person zu bestimmen.

Verifikation: Bei der Verifikation wird ein Gesichtsbild bzw. eine Stimmaufzeichnung mit einem weiteren Bild bzw. einer weiteren Stimmaufzeichnung verglichen. Diese 1:1-Vergleiche werden insb. zur Authentifizierung genutzt.

Emotionserkennung: Emotionserkennung ist ein Prozess, in dem aus audiovisuellen Signalen Aussagen über emotionale Zustände abgeleitet werden.

1.4. Methodologie

Das Projektkonsortium vereinte interdisziplinäre Perspektiven, um die o.g. Fragestellungen zu beantworten. Dabei wurde auf einen Methodenmix zurückgegriffen, der in Abschnitt 1.4.1 kurz vorgestellt wird. Detaillierte Informationen zu der jeweiligen Methodik können den entsprechenden inhaltlichen Kapiteln entnommen werden. Die Fachexpertise des Projektteams wird in 1.4.2 vorgestellt.

1.4.1. Arbeitsschritte und Methoden

1. **Literaturanalyse:** Das zentrale methodische Instrument der Studie ist die Literaturanalyse. Dies schliesst die bibliometrische Analyse in Kapitel 2.3 und die technischen, rechtlichen und ethischen Detailanalysen zu den jeweils untersuchten Anwendungsfeldern in Kapitel 3 ein.
2. **Analyse der Medienberichterstattung:** Zur Identifikation der für die Schweiz relevanten Anwendungsfälle und -gebiete in Kapitel 2.4 wurde auf die quantitative und qualitative Analyse der Medienberichterstattung in der Schweiz zurückgegriffen.
3. **Bevölkerungsumfrage:** Zur quantitativen Untersuchung der Bevölkerungsmeinung wurde eine repräsentative Bevölkerungsumfrage durchgeführt. Dabei wurde auf die Dienste des Meinungsforschungsdienstleisters IPSOS zurückgegriffen.
4. **Fokusgruppen:** Zur qualitativen Untersuchung der Bevölkerungsmeinung führte TA-SWISS gemeinsam mit dem Projektteam Fokusgruppen durch.

Die Studie ist in fünf inhaltliche Kapitel unterteilt. Das erste inhaltliche Kapitel 2 dient im Rahmen einer Ist- und Trendanalyse der Einführung in das Themengebiet. Dazu wird zunächst in die Geschichte und Funktionsweise der Stimm-, Sprach- und Gesichtserkennung (2.1) eingeführt. Daran schliesst die bibliometrische Analyse des Publikationsaufkommens zu Stimm, Sprach- und Gesichtserkennungstechnologien in 2.2 und die Identifikation von Anwendungsfällen und -gebieten in 2.3 an. Kapitel 3 beginnt mit einer allgemeinen Einführung in die juristischen Grundlagen (3.1) sowie die gesellschaftlichen und ethischen Herausforderungen (3.2). Im übrigen Teil von Kapitel 3 werden alle acht Anwendungsfelder der Reihe nach im Hinblick auf die technischen Grundlagen und Möglichkeiten analysiert,

juristisch bewertet sowie die gesellschaftlichen und ethischen Herausforderungen erörtert. In Kapitel 4 werden die Ergebnisse der Fokusgruppen und in Kapitel 5 die Ergebnisse der Bevölkerungsumfrage vorgestellt und diskutiert. Kapitel 6.1 beinhaltet die Empfehlungen und 6.2 die Schlussfolgerungen der Studie.

1.4.2. Projektkonsortium

Die Projektmitarbeiterinnen und -mitarbeiter des Konsortiums waren entsprechend ihrer jeweiligen Expertise für unterschiedliche Teile der Studie federführend zuständig.

Das **Fraunhofer-Institut für System- und Innovationsforschung (ISI)** in Karlsruhe, Deutschland, ist eine von mehr als 80 Forschungseinrichtungen der Fraunhofer-Gesellschaft. Es hat einen Fokus auf sozioökonomische Forschung, Vorausschau, Evaluation, Wirkungsanalysen und Technologiefolgenabschätzung sowie Politikberatung. Das ISI ist Mitglied in der European Technology Assessment Group (ETAG) und im globalTA Network. Für die Durchführung der geplanten Studie waren Wissenschaftler aus dem Geschäftsfeld «Informations- und Kommunikationssysteme» im Competence Center «Neue Technologien» zuständig.

Das Fraunhofer ISI war für die Bearbeitung der Kapitel 1, 2, 4 und 5 und die Analyse der technischen Grundlagen und Möglichkeiten sowie der gesellschaftlichen und ethischen Herausforderungen federführend zuständig. Bei der Formulierung der Empfehlungen unterstützte das Projektteam des Fraunhofer ISI das Projektteam des Instituts für Europarecht.

Das **Institut für Europarecht der Universität Freiburg i.Ue.** gehört zu den führenden Forschungseinrichtungen für Datenschutz in der Schweiz. Seit über zehn Jahren richtet das zweisprachige (d./fr.) Institut, zusammen mit dem Eidgenössischen Datenschutzbeauftragten (EDÖB), den Schweizerischen Datenschutzrechtstag sowie Weiterbildungen für Berufspraktikerinnen und -praktiker im Datenschutzrecht aus.

Das Projektteam des Instituts für Europarecht war federführend für die juristische Bewertung der Anwendungsfelder und die Erarbeitung der Empfehlungen (mit Unterstützung des Fraunhofer ISI) zuständig. Ausserdem wirkte das Projektteam des Instituts für Europarecht unterstützend bei den Arbeitsschritten, für die das Fraunhofer ISI federführend war. Das Projektteam wurde beratend von Prof. Dr. Astrid Epiney unterstützt.

2. AP 2: Ist- und Trendanalyse

2.1. Funktionsweise der Technologien

Sowohl bei der Gesichts- als auch der Stimmerkennung handelt es sich um biometrische Verfahren, da sie Charakteristiken eines Menschen nutzen. Grundsätzlich ist zwischen zwei Funktionsweisen zu unterscheiden: Identifikation und Verifikation. Bei der Identifikation wird ein neues Bild oder Stimmsample mit vielen weiteren aus einer Datenbank verglichen, um bspw. den Namen einer Person zu bestimmen (1:N). Bei der Verifikation wird hingegen ein Bild mit einem weiteren verglichen. Diese 1:1-Vergleiche werden insb. zur Authentifizierung genutzt. Grundsätzlich gilt die Identifikation als die schwierigere Methode, weil es sich bei dem Abgleich eines Bildes bzw. Audiosamples mit einer Vielzahl von Bildern bzw. Audiosamples um einen weitaus komplizierteren Rechenvorgang handelt (Gunther et al. 2017, S. 697). Im Folgenden wird die Funktionsweise der Technologien näher betrachtet (Michael und Roth 2001, S. 10).

2.1.1. Die Anfänge der Gesichtserkennung

Während andere biometrische Methoden, etwa der Vergleich von Fingerabdrücken (Daktyloskopie) oder die Vermessung der Körper(-gliedmassen-)länge, bereits seit mehr als 100 Jahren Anwendung finden, ist die Gesichtserkennung eine vergleichsweise neue Methode. Ab 1964 beschäftigte sich der Informatiker Woodrow Wilson Bledsoe mit der Frage, ob Computer menschliche Gesichter erkennen können (Bledsoe 1966). Hierzu entwickelte er ein einfaches Computerprogramm, welches die Haarlinie, Augen und Nase schematisch kartografierte. Dieses Vorhaben war jedoch nicht erfolgreich, da die Varianz von Kopfbewegungen und -neigungen, die Intensität des Lichts und dessen Einstrahlungswinkel sowie unterschiedliche Gesichtsausdrücke und Alterungsprozesse eine Erkennung erschwerten (Bledsoe 1966). So dauert es bis in die Mitte der 1970er-Jahre, als Computerhardware einen gewissen Leistungsstand erreichte und dadurch weitere Merkmale wie Haarfarbe und Lippendicke mit einbezogen werden konnten (Kirby und Sirovich 1990). Allerdings stand zunächst nicht die Gesichts-, sondern die Objekterkennung im Fokus der frühen Forschung. Bis zur Entwicklung der ersten automatischen Gesichtserkennungssysteme dauerte es weitere 20 Jahre. Der technologische Durchbruch gelang schliesslich 1991, als mit dem «Eigenface»-Algorithmus ein erstes Verfahren zur automatischen Gesichtserkennung entwickelt worden war. An die Stelle der merkmalbasierten Vermessung der Anatomie einzelner Gesichter setzten Turk und Pentland die statistische Auswertung grosser Bilddatenmengen mittels Hauptkomponentenanalyse (Turk und Pentland 1991).

Der mit dem Eigenface-Algorithmus erzielte Erfolg initiierte einen Forschungsboom. Allerdings stellte sich schon bald heraus, dass sich die Erkennungsraten nur schlecht miteinander vergleichen liessen, da jedes Forschungsteam eine eigene Bilddatenbank nutzte (Meyer 2020). Um diesem Problem zu begegnen, erfolgte im Jahr 1993 der Startschuss für

das vom US-amerikanischen Verteidigungsministerium (Department of Defence, DoD) initiierte «Face Recognition Technology» (FERET) Programm (Rauss et al. 1997; Zhao et al. 2003). Das Ziel dieses Programms war u.a. die Erstellung einer Datenbank, die den Vergleich unterschiedlicher Algorithmen ermöglichen sollte (NIST 2017).

Darauf aufbauend entwickelte sich ab 2000 der jährliche «Facial Recognition Vendor Test» (FRVT) des amerikanischen «National Institute of Standards and Technology» (NIST), der bis heute regelmässig durchgeführt wird und weltweit als Quasi-Standard zum Vergleich von Gesichtserkennungsalgorithmen angesehen wird (Chambers 2020). Hierbei handelt es sich um eine grossangelegte unabhängige Evaluierung von verschiedenen Gesichtserkennungsalgorithmen (NIST 2020). In der 2020er-Version des Tests wurden 189 Algorithmen von 99 unterschiedlichen Herstellern untersucht. Anzumerken ist, dass sich einige Firmen wehren, ihre Algorithmen zum Testen freizugeben, wodurch die Aussagekraft des Tests geschmälert wird (Chambers 2020). Das Design der Tests berücksichtigt zudem gesellschaftliche Debatten. Im Ergebnis der Diskussionen rund um ethnische Diskriminierung im Zusammenhang mit Gesichtserkennung wurden die Tests daraufhin überarbeitet, das Hauptaugenmerk auf demografische Effekte, genauer gesagt, auf soziale Verzerrungen zu legen, also Unterschiede bei der Erkennungswahrscheinlichkeit zwischen Geschlechtern, Ethnien und Alter (Grother et al. 2019).

Bis 2006 wurde zudem die «Face Recognition Grand Challenge (FRGC)» durchgeführt, die Programmierer, Forscher, aber auch Firmen aufforderte, Gesichtserkennungsalgorithmen zu verbessern und gegeneinander antreten zu lassen (NIST 2016). Gleichzeitig gab es auch weitere Tests, wie z.B. den «Face-in-Video-Evaluation (FIVE)», welcher Gesichtsmorphe und demografische Effekte von Gesichtserkennung innerhalb von Videosequenzen testete (NIST 2019). Hier werden vier verschiedene Szenarien betrachtet: Erkennung in Menschenmassen, Überwachungsvideos, Videokonferenzen und Fernsehen.

Einen Paradigmenwechsel für die Biometrie markierten schliesslich die Terroranschläge vom 11. September 2001. Dieser äusserte sich sowohl in einem allgemeinen Klima, das neueren Überwachungstechnologien wie der Gesichtserkennung gegenüber positiv gestimmt war, als auch in der biometrischen Aufrüstung von Polizeien, Grossveranstaltungen und Grenzkontrollen. Auf Druck der Vereinigten Staaten mussten die 26 Mitglieder des Visa Waiver Program (VWP), in dessen Rahmen Visa-freie Kurzaufenthalte in den USA möglich waren, um weiterhin am VWP teilnehmen zu können, biometrische Pässe einführen. Sowohl die EU als auch die Schweiz passten in der Folge ihre Gesetze entsprechend an und legten mit der Einführung standardisierter biometrischer Gesichtsbilder ab dem Jahr 2006 die Grundlagen für die staatliche Gesichtserkennung (Schweizerische Eidgenossenschaft 2006, S. 5).

Nachdem der Markt von Gesichtserkennungsanwendungen in den 1990er- und 2000er-Jahren von spezialisierten Unternehmen aus der Sicherheitsbranche dominiert und die relevante Forschung v.a. von der Wissenschaft forciert worden war, drängten ab den 2010er-Jahren in zunehmendem Masse neue Akteure auf den Markt bzw. die Forschungslandschaft. Akteure aus der aufstrebenden Digitalbranche konnten dabei unter Rückgriff auf die durch die Dienste-Nutzerinnen und -Nutzer bereitgestellten Daten schnell grosse Erfolge vorweisen (Meyer 2020). Facebook etwa implementierte im Juni 2011 in Europa die Funktion Deepface, die Gesichter auf jedweden auf die Plattform hochgeladenen Bildern erkennen und dem Profil der entsprechenden Person zuordnen können sollte. Nach der Kritik

von Datenschützern wurde diese Funktion jedoch wieder entfernt (Ihlenfeld 2011). Apple wiederum schaffte es, mit der Einführung von FaceID beim IPhone X aus dem Jahr 2017 zur Entsperrung des Smartphones mittels Gesichtserkennung die Technologie weiten Teilen der Gesellschaft bekannt zu machen und dadurch die Nutzung von Gesichtserkennung im Privatbereich zur Alltagstechnik werden zu lassen (Beuth 2017).

2.1.2. Technischer Hintergrund zur Gesichtserkennung

Aufgrund von Fortschritten bei Gesichtserkennungsalgorithmen und in der Hardwareentwicklung kann Gesichtserkennung heutzutage bereits mit kostengünstiger Hardware und Software durchgeführt werden. Grundsätzlich kann eine einfache, handelsübliche Kamera verwendet werden, um Gesichtserkennung zu ermöglichen. Der dabei entscheidende Faktor ist die Bildqualität der Kamera: Dies bezieht sich sowohl auf die technischen Eigenschaften der Kameras selbst als auch auf externe Faktoren, wie Verschmutzungen an der Linse, unterschiedliche Wetterbedingungen (Schnee, Nebel) oder Unschärfe, die die Bildqualität und somit die Erkennung verschlechtern können. Beeinträchtigt wird die Erkennungsleistung ausserdem von schnellen Kopfbewegungen, die ein verwaschenes Bild zur Folge haben, oder einer ungünstigen Kopfneigung, sodass nicht alle erforderlichen Gesichtselemente aufgezeichnet werden (Gunther et al. 2017, S. 697). Zur Gewährleistung einer ausreichend guten Bildqualität im professionellen Einsatz kommen daher immer häufiger mehrere Kameras zum Einsatz. Die Nutzung mehrerer Kameras ermöglicht einen sog. Stereoblick. Wärmebildkameras oder Time-of-Flight-Kameras, die den Abstand von Objektpunkten zur Kamera berechnen, ermöglichen eine Lebend- und 3D-Gesichtserkennung (Kakkirala et al. 2017; Song und Liu 2018).

Unabhängig von der eingesetzten Hardware und Software besteht der Gesichtserkennungsprozess grundsätzlich aus fünf Schritten (Kaur et al. 2020).

1. **Bilderfassung (Image Capturing)**
   Als Eingangsdaten können sowohl bestehende Bilder und Videos aus Datenbanken, z.B. (biometrische) Passbilder, als auch neu erzeugte Bilder und Videos von Überwachungskameras dienen.

2. **Gesichtsdetektion (Face Detection)**
   Die Gesichtsdetektion ist ein bestimmter Fall der Objekterkennung. Ein Algorithmus sucht in dem Bildmaterial meist zuerst nach einer Augenpartie (Sheu et al. 2014). Sofern ein mögliches Auge gefunden ist, versuchen weitere evolutionäre (also selbst lernende und selbst optimierende) Optimierungsverfahren bzw. Eigenface-Algorithmen weitere Gesichtsbestandteile wie Nasenlöcher oder Mundwinkel zu finden (Setiowati et al. 2017). An dieser Stelle können den Algorithmen bereits sog. Schwellenwerte, also bspw. eine minimale Wahrscheinlichkeit, ab wann ein Gesicht als Gesicht erkannt werden soll, voreingestellt werden. Diese Werte haben grossen Einfluss auf die Fehlertoleranz oder Falscherkennung von Gesichtern.

3. **Merkmalsextraktion (Feature Extraction)**
   In diesem Schritt werden aus dem gefundenen Gesicht bestimmte und einzigartige Bereiche extrahiert, um aus einer grossen Datenmenge kleinere, untersuchbare Einheiten

zu bilden. Diese sind in der Regel Koordinaten von markanten Punkten oder Linien. Oftmals werden hierfür v.a. die Augenpartien genutzt, da sich diese selten verändern oder bedeckt werden (Chen et al. 2018). Aus den Daten wird anschliessend ein Vektor, eine Art Gesichtsvorlage, generiert und abgespeichert.

4. **Datenbankabgleich (Database Matching)**
   Die in Schritt 3 erstellte Gesichtsvorlage wird nun mit anderen Gesichtsvorlagen aus der jeweils genutzten Datenbank verglichen. Sofern es sich bei dem Verfahren um einen 1:1-Vergleich handelt, dient als Gesichtsvorlage das jeweils zu verifizierende Gesicht. Im Falle eines angestrebten 1:N-Vergleichs verfügen die Datenbanken über beliebig viele Einträge, die mit der zu identifizierenden Gesichtsvorlage abgeglichen werden.

5. **Personenfeststellung (Person Identification)**
   Kommt es in Schritt 4 zu einem Treffer, dann ist die Person bei einem 1:1-Vergleich als berechtigt ausgewiesen oder bei einem 1:N-Vergleich deren Identität erkannt.

2.1.2.1. Herausforderungen und Trefferraten bei der Gesichtserkennung

Die Erkennungsraten von System und Algorithmen werden seit Jahren in verschiedenen Tests und sog. Challenges verglichen. Die weltweit federführende Institution auf diesem Gebiet ist das National Institute of Standards and Technology (NIST), welches eine Vielzahl von Test veröffentlicht, z.B. den Vergleich von verschiedenen Herstellern (NIST 2020) oder die Gesichtserkennung in Videomaterial (NIST 2019). Diese bieten eine gewisse Vergleichbarkeit untereinander, da dieselben Testbilder als auch dasselbe Betriebssystem zur Durchführung der Tests genutzt werden. In der Tabelle auf NIST (2020) sind die Ergebnisse von mehreren Hundert verschiedenen Algorithmen dargestellt. Sowohl in den 1:1- als auch in den 1:N-Tests weisen die bestbewerteten Algorithmen enorm hohe Erkennungsraten rund um 99 % auf.

Allerdings zeigen die NIST-Tests, dass die getesteten Algorithmen nur für einen bestimmten Anwendungsfall trainiert wurden, weshalb die Genauigkeit eines Gesichtserkennungsalgorithmus sehr stark vom konkreten Nutzungskontext und der dort verwendeten Hard- und Software abhängt (Ho et al. 2020, S. 7). Zudem ist die generelle Aussagekraft der Genauigkeit sehr gering. Laut Ho et al. (2020, S. 9) liegt dies an den sog. «domain shift» und «institutional shift». Erstere sagt aus, dass unterschiedliche Typen von Bildern von den Herstellern und den Auditoren genutzt werden, um das Modell zu trainieren und zu evaluieren (so auch bei dem NIST-Report). Die «institutional shift» beschreibt die unterschiedlichen Ausprägungen der Genauigkeit beim Einsatz in verschiedenen Anwendungsfeldern. Green und Chen (2019) konkretisieren, dass die statistische Genauigkeit nicht die tatsächliche praktische Genauigkeit widerspiegelt. Dies kann auch dann passieren, wenn dieselben Bilder beim Trainieren und Evaluieren verwendet werden (Ho et al. 2020, S. 11). Folglich stellen die NIST-Tests eine hilfreiche erste Vergleichsbasis dar, erlauben es aber nicht, generelle Aussagen über die Genauigkeit eines Algorithmus zu treffen. Um Angaben über Trefferraten im Praxiseinsatz zu erzielen, bleibt es unabdingbar, einen Algorithmus im jeweiligen Verwendungskontext zu evaluieren.

Konkret untersuchten die NIST-Forschenden im Face Recognition Vendor Test (FRVT) u.a. die kulturellen Verzerrungen der 189 getesteten Algorithmen – also wie Algorithmen mit un-

terschiedlichen demografischen Eigenschaften umgehen. So waren die Erkennungsraten bei Dunkelhäutigen, Asiaten, amerikanischen Ureinwohnern, Frauen, Kindern und Älteren erheblich schlechter als bei weissen Männern mittleren Alters (Grother et al. 2019, S. 7). Dies war v.a. bei False-Positive-Fehlern (also zwei verschiedene Gesichter wurden als die gleiche Person erkannt) auffällig. So war die Wahrscheinlichkeit, dass das Gesicht einer dunkelhäutigen Frau falsch erkannt wird, bei 30 Algorithmen bis zu zehnmal grösser und in drei Fällen sogar 100-mal grösser (Chambers 2020). Abbildung 1 zeigt die Erkennungsrate der getesteten Systeme im Vergleich zur Erkennung von weissen Männern. Es zeigen sich sowohl für weisse Frauen, schwarze Männer und schwarze Frauen erhebliche schlechtere Erkennungsraten, wobei die Erkennung dunkelhäutiger Frauen am schlechtesten funktionierte. Dass derartige fehlerhafte Gesichtserkennungsalgorithmen auch zu realer Diskriminierung durch Strafverfolger führen können, zeigen mehrere Beispiele in der westlichen Welt (Bacchini Fabio und Lorusso Ludovica 2019).

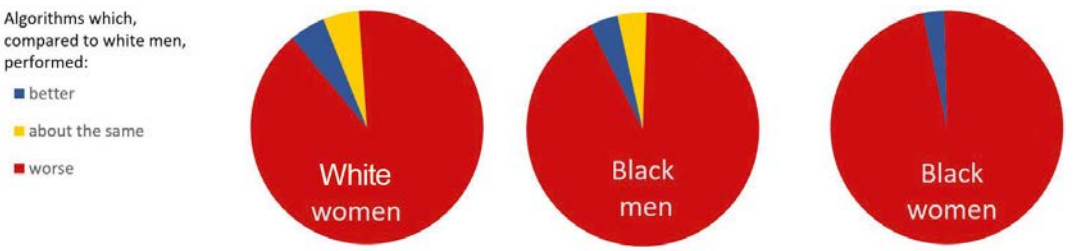

Abbildung 1: Algorithmenerkennung von versch. Gruppen im Vergleich zu weissen Männern (eigene Darstellung basierend auf Chambers 2020)

Cavazos et al. (2020) erforschen die technischen Gründe für diese Diskriminierung und zeigen, dass Diskriminierung bei Bildern zunimmt, auf denen Gesichter schwer zu erkennen sind. Zudem seien Besonderheiten des Quellenmaterials, insb. die mangelnde Vielfalt in den Trainingsdaten (wie z.B. eine geringe Anzahl an Gesichtern von dunkelhäutigen Frauen), ein Grund. Doch auch technische Besonderheiten, die weit vor der Gesichtserkennung entwickelt wurden, hätten Einfluss auf die Erkennungsalgorithmen. So ist die Farbfotografie grundsätzlich auf hellere Hauttöne optimiert (Kaltheuner und Obermüller 2018), sodass dunkle Bereiche und damit auch Gesichter bereits von technischer Seite her weniger gut aufgenommen werden, sodass die Gesichtserkennung erschwert wird. Die Behebung der vielfältigen möglichen Gründe für die Diskriminierung sei allerdings ein komplexes und schwieriges Verfahren (Kaltheuner und Obermüller 2018). So konnte selbst Google die Fehlzuordnung von dunkelhäutigen Menschen als «Gorillas» nur dadurch beheben, indem das Schlagwort Gorilla entfernt wurde (Simonite 2018).

Des Weiteren hat der vorgegebene Schwellenwert zur Identifizierung Einfluss auf die Erkennungsleistung. So zeigte sich bspw., dass für Ostasiaten ein höherer Schwellenwert nötig war als für kaukasische Gesichter, um eine gleiche Falsch-Akzeptanzrate zu erreichen. Gleichzeitig sei es erforderlich, dass der Test für jede Anwendung individuell durchgeführt wird. Oft würden Schwellenwerte bei Tests jedoch willkürlich gesetzt, sodass ein Vergleich kaum aussagekräftig sei (Chambers 2020). Lediglich der bereits erwähnte NIST FRVT nutzt relative, statt absolute Schwellenwerte für das Benchmarking (Grother et al. 2019, S. 20). Oft wird auch auf die Festlegung von Schwellenwerten verzichtet. Als konkretes Beispiel

lässt sich hier der Fall der Gesichtserkennung mittels Amazons «Rekognition» durch die Polizei in Oregon aufführen. Amazon selbst gibt an, dass der Schwellenwert für die korrekte Erkennung bei Polizeibehörden mindestens auf 99 % gesetzt werden solle. Die Polizei in Oregon setzt jedoch keinerlei Schwellenwerte, wodurch viel mehr Falsch-Positive Ergebnisse entstehen (Menegus 2019).

Die Erkennungsraten von Algorithmen werden zudem oft mit denen von Menschen verglichen. So zeigten O'Toole et al. (2007), dass bereits vor mehr als zehn Jahren insgesamt neun Algorithmen gleiche Gesichter unter verschiedenen Lichtbedingungen besser zu erkennen in der Lage waren als Menschen, sofern die Gesichter den Menschen unbekannt waren. Menschen waren hingegen dann besser in der Gesichtserkennung, wenn die ihnen gezeigten Gesichter ihnen bereits bekannt waren (O'Toole et al. 2007, S. 1645). In einem neueren Experiment (Phillips et al. 2018) verglich das gleiche Forscherteam die Erkennungsrate von «Experten» (Fingerabdruckexperten, forensische Gesichtserkennungsexperten und sog. Super-Recognizer, welche sich überdurchschnittlich gut Gesichter einprägen können) mit vier sog. «deep convolutional neural networks» (DCNN), die zwischen 2015 und 2017 entwickelt wurden. Die Ergebnisse zeigten, dass sich die Erkennungsrate der Algorithmen in den Jahren monoton erhöhte. Tabelle 1 fasst die Ergebnisse der Erkennungsrate (in Area Under The Curve AUC) zwischen den Algorithmen und den Experten zusammen. Es zeigt sich, dass sich die Treffergenauigkeit der Algorithmen stetig verbessert hat und mittlerweile auf demselben Niveau liegt wie bei Gesichtserkennungsexperten und teilweise sogar besser ist (Nguyen et al. 2020). Der Algorithmus mit der höchsten Trefferrate musste sich dem forensischen Gesichtserkennungsexperten bspw. lediglich in einem einzigen Fall geschlagen geben (A2017a). Phillips et al. (2018, S. 6171) weisen in diesem Zusammenhang darauf hin, dass eine Zusammenarbeit von Algorithmus und Mensch die besten Ergebnisse liefen könne.

Tabelle 1: Erkennungsrate von DCNN-Algorithmen im Vergleich zu menschlichen Experten

Algorithmus (Jahr)	**Leistung (AUC)**	**Menschlicher Experte**	**Leistung (AUC)**
A2015	0.68	Studenten	0.68
A2016	0.76	Fingerabdruckexperten	0.76
A2017a	0.85	Super-Recognizer	0.83
A2017a	0.85	forensischer Gesichtsprüfer (forensic facial reviewers)	0.87
A2017b	0.96	forensischer Gesichtsprüfer (forensic facial examiners)	0.93

Schliesslich ist die Erkennung von bewegten Bildern bzw. teilverdeckten Gesichtern eine technische Herausforderung: In einem Feldexperiment mit Fussgängern an einer US-Universität untersuchten Gunther et al. (2017, S. 704) drei verschiedene Algorithmen zur Gesichtserkennung bei bewegten Bildern und Gesichtern, die in der Datenbank noch nicht vorhanden waren («unbekannt»). In diesem Fall zeigte sich, dass der Algorithmus entweder eine sehr hohe Falscherkennungsrate hatte oder nur 60 % der Gesichter identifizierte (Gun-

ther et al. 2017, S. 704). Allerdings gab es in den vergangenen Jahren auch auf diesem Gebiet grosse Fortschritte: Nach Beginn der Coronapandemie wurde die Leistung von Algorithmen (1:1-Vergleich) zur Erkennung von mit einem Mund-Nasen-Schutz teilverdeckten Gesichtern bis um den Faktor 10 verbessert (Krempl 2020).

2.1.3. Die Anfänge der Stimm- und Spracherkennung

Ähnlich wie bei der Gesichtserkennung verliefen die ersten Forschungsvorhaben bei der Stimm- und Spracherkennung in den 1960er-Jahren zunächst wenig erfolgreich. Aufgrund der nicht ausreichenden Computerleistung war die Erkennung auf einzelne wenige Wörter begrenzt. Einen Durchbruch bildete die von IBM entwickelte Software «IBM Shoebox», die im Jahr 1962 als erste Software in der Lage war, 16 gesprochene englische Wörter zu erkennen (IBM Corporation 2011). In den folgenden Jahrzehnten wurde das System weiter verbessert, sodass es 1984 bereits 5000 englische Wörter automatisiert erkennen konnte (IBM Corporation 2011). Diese Verbesserung war möglich, da die Bedeutung gleichklingender Wörter (Homophone) mittels Häufigkeitsanalysen unterschieden werden konnte (Merialdo 1988). Dafür waren jedoch noch ein Grossrechner und mehrere Minuten Rechenzeit erforderlich.

Eine leistungsfähigere Alternative folgte mit der Spracherkennung der Firma Dragon Systems, die auf einem tragbaren Computer genutzt werden konnte (Baker und Baker 1989). Das EU-Projekt SUNDAIL zeigte zwischen 1988 und 1993, dass Spracherkennung von 1000–2000 Wörtern in deutscher, französischer und italienischer Sprache in «Telefon-Qualität» möglich ist (European Commission 1994). Etwa zeitgleich stellte IBM mit «TANGORA» ein System vor, welches 20.000 bis 30.000 deutsche Wörter erkennen konnte (Das und Picheny 1996). 1993 veröffentlichte IBM schliesslich mit dem «VoiceType» ein kommerzielles Diktiersystem (IBM Corporation 2011), welches grossen Anklang fand. Es folgten viele weitere Lösungen von Firmen wie Dragon NaturallySpeaking (Nuance 2021) und Philips, welche sich hauptsächlich als Bürostandard etablierten. In dieser Zeit schritt auch die Stimmerkennung voran, mit der die Erkennung unterschiedlicher Sprecher und ihre Authentifizierung bzw. Identifikation möglich wurde (Myers 2022, S. 16).

In den 2000er-Jahren stellte auch Microsoft erste Spracherkennungssysteme für sein Betriebssystem Windows, bestimmte Programme und die Office-Pakete zur Verfügung. So konnte z.B. Windows Vista teilweise mittels Sprachbefehlen bedient werden (Odell und Mukerjee 2007). Für das Apple-Betriebssystem Mac OS wurde mit den Namen «MacSpeech» seit 2006 eine Spracherkennungssoftware vertrieben. Diese wurde jedoch 2010 in die Dragon NaturallySpeaking-Produktreihe integriert. Als plattformunabhängiges System hat sich seither Dragon NaturallySpeaking durchgesetzt. Zunächst kam das seit 2007 entwickelte, 2010 von Apple aufgekaufte und seit 2011 beginnend mit dem iPhone 4s kommerziell vertriebene System «Siri» auf den Markt (Bosker 2013). Ende 2014 folgte schliesslich das in kleine, preiswerte und selbstständige Lautsprecher integrierte System Amazon Alexa, das weltweit zu einer erhöhten Nutzung von Stimm- und Spracherkennung führte (Yoffie et al. 2018, S. 25).

2.1.4. Technischer Hintergrund zur Stimm- und Spracherkennung

Grundsätzlich muss zwischen Spracherkennung einerseits und Stimm- bzw. Sprechererkennung andererseits unterschieden werden. Bei Ersterer handelt es sich um die Extraktion von Inhalten und Bedeutungen aus einer Aussage, um bspw. die Steuerung von Informationssystemen mittels Spracheingaben zu ermöglichen (Munteanu und Penn 2018). Letztere bezeichnet hingegen ein biometrisches Verfahren zur Erkennung von Personen, z.B. für die Authentifizierung an einem Computersystem (Dellwo et al. 2018, S. 777). Weiter lassen sich die Systeme in sprecherabhängige, sprecherunabhängige und sprecheradaptive Spracherkennung unterteilen (Pfister und Kaufmann 2017, S. 334). Bei einer sprecherunabhängigen Erkennung, welche hauptsächlich in modernen Spracherkennungssystemen, wie Smart Speakern und Smartphones, eingesetzt wird, kann die Erkennung ohne eine vorhergehende Trainingsphase beginnen.

Technisch gesehen unterscheiden sich Stimm- und Spracherkennung kaum. Die Erkennung von Stimme und Sprache kann in zwei grobe Schritte unterteilt werden:

1. **Vorarbeit und Texterkennung:**
   Die gesprochene Sprache wird als Tondatei aufgezeichnet. Dann folgt eine Reihe von Vorarbeiten. Hierbei wird das analoge Signal zuerst digitalisiert. Danach werden Stör- und Hintergrundgeräusche herausgefiltert (Pfister und Kaufmann 2017, S. 337) und aus dem Signal das Frequenzspektrum ermittelt (Pfister und Kaufmann 2017, S. 60). Als letzte Vorarbeit wird ein Merkmalsvektor aus dem digitalen Sprachsignal erstellt, welcher bspw. Periodizitäten im Spektrum beinhaltet.

   Zur eigentlichen Texterkennung werden die Sprachsignale anschliessend in kleine sprachliche Einheiten (Phoneme) zerlegt und mittels Hidden-Markov-Modelle oder neuronaler Netze mit vorhandenen Sprachdaten verglichen, um die höchste Übereinstimmung zu finden (Mustafa et al. 2019; Wang und Chen 2018).

2. **Sinnerkennung und Sprachmodell:**
   Im zweiten Schritt wird versucht, passende Wortkombinationen zu finden und so eine sinnhafte Spracherkennung zu ermöglichen. Dies geschieht mithilfe von Wahrscheinlichkeiten einzelner Wortkombinationen. Die Auftrittswahrscheinlichkeit von zwei oder drei Wörtern (Bi- oder Trigrammstatistik) kommt aus Grammatikmodellen (Damavandi et al. 2016). Hiermit konnte auch das Problem gleichklingender Wörter (Homophone) gelöst werden, da bspw. «Vielen Dank» viel wahrscheinlicher ist als «Fielen Dank», obwohl beides gleich ausgesprochen wird.

2.1.4.1. Herausforderungen und Trefferraten bei der Stimm- und Spracherkennung

Die Spracherkennung ist oft schwierig, da die oben genannten Prozesse problemanfällig sind. Dies hängt zuerst massgeblich von der Qualität des Tonsignals ab. Auf die Tonqualität haben wiederum insb. die verwendete Hardware (Mikrofonqualität) und verschiedene Sprechlautstärken (Flüstern, Schreien) oder Sprechweisen (z.B. Nuscheln) grossen Einfluss. Zudem haben verschiedene Laute unterschiedlichen Einfluss auf die Erkennungsrate von Systemen. So führen Vokale und Nasallaute zu einer besseren Spracherkennung (Dell-

wo et al. 2019, S. 62). Zudem zeigt sich, dass Füllwörter wie «uh» und «ehm» Spracherkennungssysteme vor Herausforderungen stellen (Stolcke und Droppo 2017, S. 137).

Unabhängig von der Art der Laute muss die Datenbasis, also das Wörterbuch, mit der die Phoneme verglichen werden, eine Vielzahl von Inhalten haben. So basiert bspw. der deutsche Wortschatz auf einer Vielzahl von Endungen oder Wortformen wie «gehen, ging, gegangen». Zudem hat in dieser Sprache die Gross- und Kleinschreibung oft eine wichtige Bedeutung. Das Gelingen der Spracherkennung erfordert die Berücksichtigung all dieser Faktoren in der Datenbasis.

Auch Dialekte können die Erkennung negativ beeinflussen (Yoo et al. 2019, S. 5716). Um diese Probleme zu umgehen, ist es nötig, die maschinellen Lernsysteme zu trainieren. So setzten bspw. Beck et al. (2018) Crowdsourcing ein, um Schweizerdeutsch zu transkribieren. Ebenso zeigen die Erkennungsraten aktueller Spracherkennungssysteme Ungleichheiten zwischen unterschiedlichen Ethnien, die auf eine fehlerhafte Akzenterkennung zurückzuführen ist. So zeigte sich bei der Untersuchung von Koenecke et al. (2020, S. 7684) bei weissen Sprechern eine Fehlerrate von 0,19, während diese bei schwarzen Sprechern mit 0,35 fast doppelt so hoch lag.

Für die Interaktion mit digitalen Assistenten ist die Sinnerkennung einer Eingabe wichtig. Die Spracheingabe des Nutzers hat hier jedoch oft viele Füll- oder Hilfswörter, welche erkannt und entfernt werden müssen. Aus einer einfachen Abfrage wie «Wie wird das Wetter am Montag in Bern» sind für das System eigentlich nur drei Wörter inhaltlich wichtig: «Wetter», «Montag» und «Bern». Schliesslich kommt es nicht allein auf die Wortbedeutung, sondern auch auf die Betonung an. «Ein Schild umfahren» etwa kann je nach Betonung bedeuten, dem Schild auszuweichen oder es zu treffen (Schulz 2012).

Aus diesen Gründen ist es wenig verwunderlich, dass sich die Fehlerraten bei Spracherkennungsalgorithmen stark unterscheiden. In der Wissenschaft ist die «word error rate» (WER) eine weitverbreitete Messzahl für die Korrektheit von Spracherkennungsalgorithmen und ermöglicht somit eine Vergleichbarkeit (Ali und Renals 2018). Die WER schliesst u.a. falsch eingefügte Wörter ein, z.B. «Spracherkennung funktioniert ganz gut» anstelle von «Spracherkennung funktioniert gut». Oder das falsche Löschen von Wörtern, z.B. «TA-SWISS hat … Sitz in der Schweiz» anstelle von «TA-SWISS hat ihren Sitz in der Schweiz». Jedoch liefert die WER keine Informationen über genaue Fehler oder Gründe für Falscherkennungen. Die «word accuracy» (WAcc), die auch oft als Messzahl genannt wird, spiegelt den Kehrwert der WER dar (Ogawa et al. 2012). Des Weiteren gibt es noch die Kennzahl «equal error rate» (EER), die die Zahl repräsentiert, bei der ein Typ-1-Fehler auch ein Typ-2-Fehler ist. Wenn also bspw. ein Schwellenwert auf 5 % festgelegt ist, also 5 % der Korrekten werden nicht erkannt und gleichzeitig 5 % der Falschen akzeptiert, resultiert daraus eine EER von 5 %.

Latif et al. (2020, S. 347) vergleichen sieben verschiedene Algorithmen anhand des WER-Werts und zeigen, dass sich diese zwischen 5,78 % und 1,90 % WER befinden. In einem Test, basierend auf standardisierten Testdaten von NIST 2000, welche u.a. 150 Stunden Telefongespräche beinhalten, zeigten Stolcke und Droppo (2017, S. 137) eine WER von zwischen 5 % und 11 %, je nach Anwendungsfall und Ort. Im Test mit Menschen zeigten sich ähnliche Werte. So findet sich bei Untersuchungen von Gesprächen in Telefonzentralen ein Wert von 5,8 % für Algorithmen und 5,9 % für Menschen (Stolcke und Droppo

2017). Die nah beieinanderliegenden Werte weisen aber auch daraufhin, dass es Laute gibt, die weder eine Maschine noch ein Mensch korrekt verstehen kann. Dabei wird auch darauf hingewiesen, dass die Erkennungsraten von Menschen ebenfalls variieren können. Gründe hierfür seien v.a. unterschiedliche Situationen, z.B. Telefon, oder persönlich, Gespräche zwischen Bekannten oder Fremden. Dementsprechend zeigen sich in unterschiedliche Arbeiten menschliche WER von 4,1 % bei sehr sorgfältiger Transkription, aber auch 9,6 % bei schneller Transkription (Glenn et al. 2010). Bei der Aussprache einzelner Buchstaben zeigte sich eine menschliche WER von nur 1,6 % (Lippmann 1997, S. 6). Neuere Forschungsergebnisse zeigen, dass Algorithmen Sprache zunehmend besser erkennen als Menschen. Diese werden als «super-human performance» bezeichnet wird. Mit einem WER von exakt 5,0 % zeigten Nguyen et al. (2020) ein solches praxistaugliches onlinebasiertes System, welches ohne grosse Bearbeitungsverzögerungen arbeitet. Doch es gibt aufgrund von unterschiedlichen Kennzahlen auch sehr unterschiedliche Ergebnisse. So untersuchten Chadha et al. (2015, S. 30) Algorithmen von zehn verschiedenen Autoren und zeigten, dass die Fehlerraten zwischen 2 % und 48,3 % liegen.

Viele der in diesem Abschnitt besprochenen Herausforderungen lassen sich mit gut trainierten Algorithmen lösen. Daher setzen viele Hersteller von Spracherkennungssystemen auf menschliche Transkription (Kremp 2019). Somit ist zu erwarten, dass die Erkennungsraten sukzessive besser werden. Allerdings können sowohl externe Faktoren, wie eine schlechte Netzverbindung oder Mikrofonqualität, als auch algorithmische Faktoren, wie eine unerwartete Sprechweise, die beim Training des Algorithmus nicht berücksichtigt wurde, auch weiterhin zu Fehlern bei den Erkennungen führen.

2.1.5. Zwischenfazit

Der Kurzüberblick zur Funktionsweise, den Herausforderungen und Trefferraten von Stimm-, Sprach- und Gesichtserkennung zeigt, dass einerseits grössere technische Fortschritte zu erwarten sind, die voraussichtlich zu einem deutlichen Anstieg der Trefferraten führen werden. Hierzu zählen die Verbesserung der Kamera- und Mikrofonqualität, der Datenbasis und Algorithmen.

Andererseits gibt es eine Reihe von Gründen, weshalb zu erwarten ist, dass keine dieser Technologien bzw. darauf basierende Anwendungen auf absehbare Zeit eine 100-prozentige Treffergenauigkeit im Praxiseinsatz erzielen wird. Hierzu zählen technische Herausforderungen, wie Videoaufnahmen unter schlechten Licht- und Wetterbedingungen oder Veränderungen an Gesichtern, aber auch die anhaltende Schwierigkeit, Bias in Datenbanken auszuräumen (Matzner 2016). Besonders relevant im Praxiseinsatz sind ausserdem soziale Faktoren, ob bspw. die von einer Kamera oder einem Mikrofon aufgezeichneten Personen den Erfordernissen der Technik entsprechen und ihr Gesicht in einer günstigen Position halten oder in ausreichender Reichweite zum Mikrofon mit normaler Stimme sprechen. Die Betreiber von Stimm-, Sprach- und Gesichtserkennungssystemen sind zudem damit konfrontiert, durch organisatorische Massnahmen sicherzustellen, dass die eingesetzte Technik ordnungsgemäss funktioniert. In diesem Zusammenhang könnte es, etwa im Falle wetterbedingt verschmutzter Linsen, selbst erforderlich sein, die Kameralinsen zu reinigen.

2.2. Bibliometrische Auswertung wissenschaftlicher Publikationen

Zur Erfassung des wissenschaftlichen Publikationsaufkommens und der Positionierung der Schweiz im internationalen Vergleich wurde eine bibliometrische Auswertung durchgeführt.

2.2.1. Methodik

Zur Auswertung wurde auf die Datenbank Scopus mittels Structured Query Language (SQL) zugegriffen. Zu Auswahl stand neben der Scopus- auch die Web-of-Science(WoS-)Datenbank. Da die WoS-Datenbank jedoch erheblich weniger Ergebnisse für die Suchbegriffe lieferte, wurde ausschliesslich Scopus genutzt. Von Interesse waren die Publikationen der zehn publikationsstärksten Länder zu den Themen Stimm-, Sprach- und Gesichtserkennung und jeweils zusätzlich Schweiz (CH), Deutschland (DE) und EU (gesamt). Daraus ergeben sich die Länder: Kanada (CA), Schweiz (CH), China (CN), Deutschland (DE), Frankreich (FR), Grossbritannien (GB), Indien (IN), Japan (JP), Südkorea (KR), Taiwan (TW) und USA (US). Es wurden ausschliesslich deutsch- und englischsprachige Publikationen untersucht.

Die Suche nach relevanten Publikationen wurde in einem iterativen Prozess durchgeführt. Mittels SQL-Zugang zur Scopus-Datenbank konnten sehr präzise Suchanfragen erstellt werden. Die Suche fand im November 2020 statt, wobei die Datenbank alle Publikationen bis Kalenderwoche 17, 2020 enthielt.

Die Suche wurde auf die Schlagworte (Author_keywords) und den Titel der Publikationen angewandt. Für die gefundenen Publikationen wurden alle Schlagworte extrahiert und gruppiert ausgegeben. Als Ausgangssuchworte für die erste Suchiteration dienten die Schlagworte in *kursiv* in Gross- und Kleinschreibung:[1]

1. *%Gesichtserkennung%* | *%Gesichtsbiometrie%* | *Facial Recog%* | *Face Recog%* **| Face Detection% | Facial Detection% | Biometric% AND Facial | Biometric% AND Face | Feature Extraction AND Face | Feature Extraction AND Facial**
2. *Spracherkennung%* | *Sprachenerkennung%* | *Sprach-Erkennung%* | *Sprachen-Erkennung%* | *Stimm-Erkennung%* | *Stimmen-Erkennung%* | *Stimmerkennung%* | *Stimmenerkennung%* | *Speech Recogn%* | *Voice Recogn%* | *Vocal Recogn%* **| Feature extraction AND voice | Feature extraction AND speech | Feature extraction AND vocal | ASR AND voice | ASR AND speech | ASR AND vocal | Language recogn%**

Aus den Ergebnissen der ersten Suchinteraktion wurden die meistverwendeten Schlagworte extrahiert. Diese dienten anschliessend zur Identifikation weiterer relevanter Schlagworte **(fett)**.

[1] Die Prozentzeichen stellen sog. Wildcards dar, womit weiterführende Wörter, wie z.B. «Gesichtsbiometrie*erkennung*» als Suchbegriff gelten. Der senkrechte Strich stellt ein OR (Oder) dar, sodass nur eines der Suchworte im Ergebnis vorkommen muss.

2.2.2. Ergebnisse zur Gesichtserkennung

Zum Thema Gesichtserkennung wurden insgesamt 16.309 relevante Publikationen gefunden. Wie in Abbildung 2 zu sehen, wurde ein Grossteil durch Autorinnen und Autoren aus China, den USA und Indien veröffentlicht, wobei China mit 4554 Publikationen mehr als doppelt so viel veröffentlicht als die USA oder Indien. Europäische Länder, wie die Schweiz, Deutschland, Frankreich, aber auch Grossbritannien, kommen auf 89 bis 745 Publikationen im kompletten Zeitverlauf. Alle EU-Länder und die Schweiz sowie Grossbritannien (EU_gesamt)[2] kommen auf insgesamt 1660 Publikationen. Abbildung 2 zeigt zudem die Anzahl der Publikationen zum Thema Gesichtserkennung im Vergleich zu der Gesamtanzahl aller Publikationen des Landes. Hier zeigt sich ebenfalls ein heterogenes Bild. Während unter den europäischen Ländern der prozentuale Anteil zwischen 0,011 % und 0,020 % schwankt, zeigt sich in Hongkong (HK) (0,083 %) und Indien (IN) (0,081 %) ein mehr als viermal so hohes relatives Publikationsaufkommen. In China ist diese etwa dreimal so hoch (0,067 %).

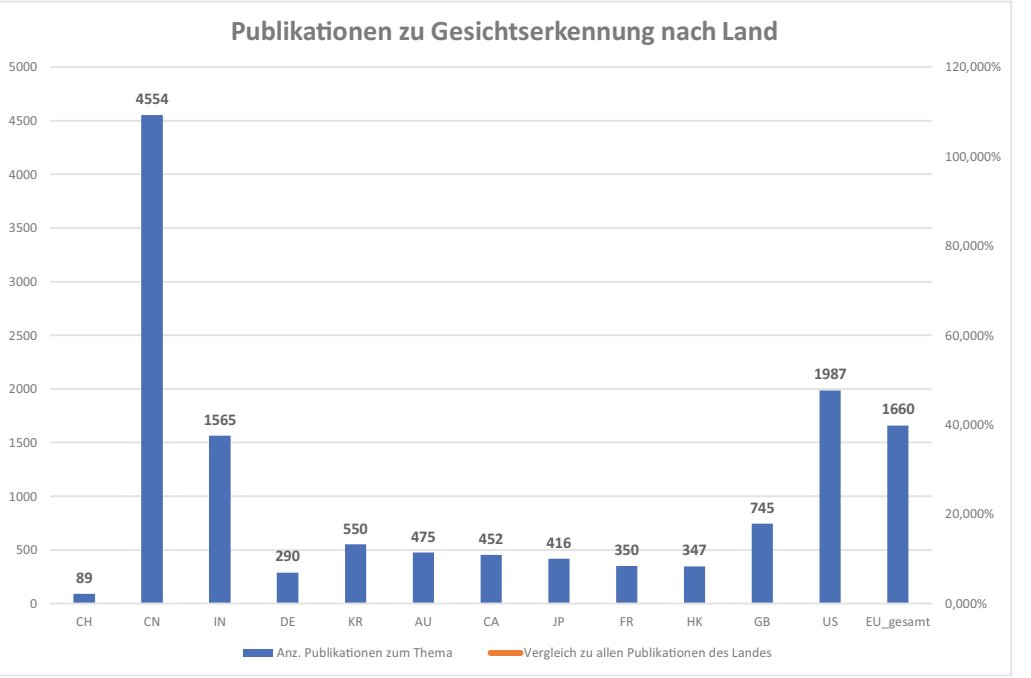

Abbildung 2: Top-10-Anzahl der Veröffentlichungen pro Land sowie die Anzahl der Publikationen zum Thema Gesichtserkennung im Vergleich zur gesamten Publikationstätigkeit des Landes (EU_gesamt umfasst alle Eurostaaten, GB und die Schweiz)

Ein Vergleich der Publikationstätigkeiten gruppiert nach Ländern im zeitlichen Verlauf ist im Onlineappendix[3] verfügbar. Hierbei wird deutlich, dass die chinesische Dominanz auf dem Gebiet der Publikationen zum Thema Gesichtserkennung nicht erst seit wenigen Jahren besteht, son-

[2] Hierzu zählt in allen weiteren Auswertungen auch die Schweiz.

[3] 10.5281/zenodo.6838749.

dern China die Vereinigten Staaten bereits 2004 auf diesem Gebiet überholte. Eine Analyse der zeitlichen Verteilung aller Publikationen weltweit von 1996 bis 2020 zeigt ein ähnliches Bild. So gab es 1996 nur 34 Publikationen. Im Jahr 2011 wurde dann zum ersten Mal die Hundertermarke überschritten (112 Publikationen). Im Jahr 2018 schliesslich waren es 1129 Publikationen.

2.2.3. Ergebnisse zur Stimm- und Spracherkennung

Zum Thema Stimm- und Spracherkennung wurden insgesamt 17.210 relevante Publikationen gefunden. Die Hauptakteure sind die USA und China, gefolgt von Japan und Indien (siehe Abbildung 3). In diesem Themenfeld hat die USA jedoch eine Vorreiterrolle (3815 Publikationen). Bei einer Betrachtung der europäischen Länder fällt Deutschland als Vorreiter auf (991 Publikationen), gefolgt von Grossbritannien (839 Publikationen). In Relation der Anzahl der Publikationen zur Gesamtanzahl aller Publikationen eines Landes (vgl. Abbildung 3) ergibt sich ein ziemlich homogenes Bild über alle untersuchten Länder (0,022 % bis 0,030 %).

Ein Vergleich der Publikationstätigkeiten gruppiert nach Ländern im zeitlichen Verlauf ist im Onlineappendix[4] verfügbar. Hier zeigt sich eine kontinuierliche Dominanz der US-amerikanischen Publikationen. Zugleich ist eine enorme Zunahme der chinesischen Publikationen seit 2018 zu erkennen. Eine Analyse der zeitlichen Verteilung aller Publikationen weltweit von 1996 bis 2020 zeigt ein ähnliches Bild. So gab es 1996 296 Publikationen. Deren Anzahl stieg in den darauffolgenden Jahren fast linear an, bis 2018 zum ersten Mal über eintausend Publikationen erreicht wurden. Insgesamt ist aber eine starke Zunahme ab dem Jahr 2017 zu verzeichnen.

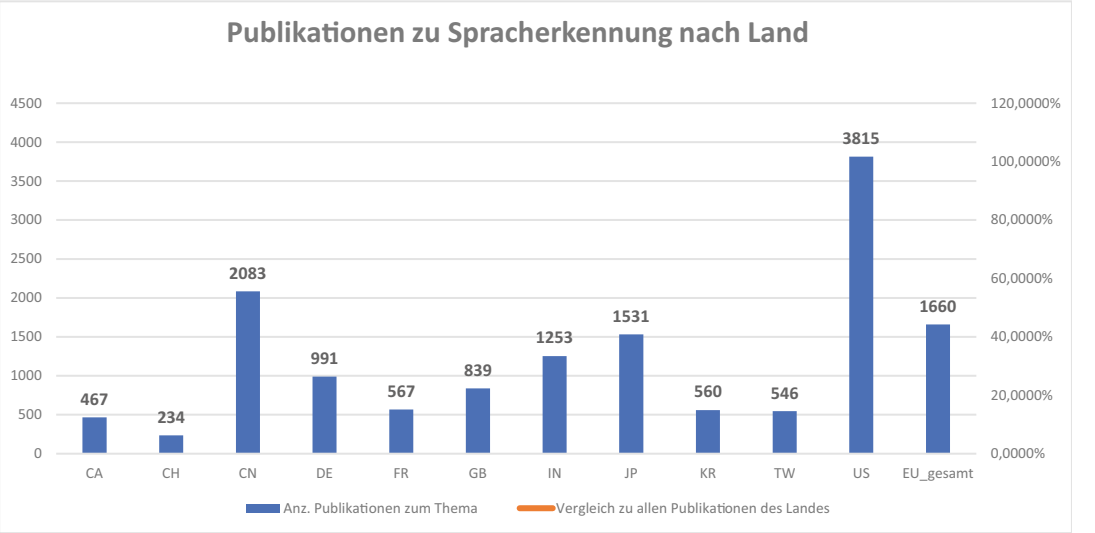

Abbildung 3: Top-10-Anzahl der Veröffentlichungen pro Land sowie die Anzahl der Publikationen zum Thema Stimm- und Spracherkennung im Vergleich zur gesamten Publikationstätigkeit des Landes (EU_gesamt umfasst alle Eurostaaten, GB und die Schweiz)

[4] 10.5281/zenodo.6838763.

2.2.4. Zwischenfazit

Der vergleichende Blick auf das wissenschaftliche Publikationsaufkommen bestätigt die führende Rolle der Vereinigten Staaten und der Volksrepublik China im Bereich der Forschung zur Stimm-, Sprach- und Gesichtserkennung. Auffallend ist die relative Häufigkeit der entsprechenden Forschungen in nicht westlichen Staaten. So beträgt der Anteil der Forschung zur Stimm-, Sprach- und Gesichtserkennung in China, Indien, Hongkong, Korea (teils auch Japan und Taiwan) am Gesamtpublikationsaufkommen in diesen Ländern ein Mehrfaches des Aufkommens in den untersuchten westlichen Ländern.

Das Publikationsniveau in der Schweiz bewegt sich grundsätzlich auf einem vergleichbaren Niveau wie in den übrigen westlichen Staaten. Lediglich auf dem Gebiet der Forschung zur Gesichtserkennung ist das relative Publikationsaufkommen in der Schweiz etwas niedriger als in anderen westlichen Staaten.

2.3. Anwendungsübersicht und Identifikation der zu untersuchenden Anwendungsfälle und -gebiete

Zur Identifizierung einschlägiger Anwendungsfelder der Stimm-, Sprach- und Gesichtserkennung wurde in einem ersten Schritt mittels Nexis (LexisNexis 2021) das gesamte Publikationsaufkommen in der Schweiz zu den Themen der Stimm-, Sprach- und Gesichtserkennung erfasst (Abbildung 4). Dieser Arbeitsschritt basierte auf demselben Such-String, der auch in der bibliometrischen Analyse verwendet wurde:

- *%Gesichtserkennung%* | *%Gesichtsbiometrie%* | *Facial Recog%* | *Face Recog%* | Face Detection% | Facial Detection% | Biometric% AND Facial | Biometric% AND Face | Feature Extraction AND Face | Feature Extraction AND Facial
- *Spracherkennung%* | *Sprachenerkennung%* | *Sprach-Erkennung%* | *Sprachen-Erkennung%* | *Stimm-Erkennung%* | *Stimmen-Erkennung%* | *Stimmerkennung%* | *Stimmenerkennung%* | *Speech Recogn%* | *Voice Recogn%* | *Vocal Recogn%* | Feature extraction AND voice | Feature extraction AND speech | Feature extraction AND vocal | ASR AND voice | ASR AND speech | ASR AND vocal | Language recogn%

Diese Suche lieferte 4437 einzelne Nachrichteneinträge. Der früheste Eintrag war vom 3. September 1993, der jüngste Beitrag vom 26. November 2020.

Auswertung

Zur weiteren Bearbeitung wurden die Daten zunächst in Microsoft Excel überführt und folgende Schritte durchgeführt:

1. Extraktion des Jahres für jeden Eintrag.
2. Entfernung von Duplikaten basierend auf den Inhalten aller Felder (38 Duplikate entfernt).
3. Entfernung von Duplikaten basierend auf den Inhalten der Felder Überschrift und Abstract (223 Duplikate entfernt).

Die manuelle Durchsicht von 4176 (=4437-38-223) Einträgen wurde auf alle Beiträge seit 2018 beschränkt. Dies ist auch der Zeitpunkt, ab dem ein signifikanter Anstieg der Berichterstattung über moderne Gesichtserkennungsanwendungen zu beobachten war. Zusätzliche drei Kriterien waren im Hinblick auf die Einschränkung der letztlich verwendeten Medien entscheidend. In das finale Sample wurden nur die Beiträge jener Medien aufgenommen, die 1. sehr oft über das Thema berichtet haben, die 2. über eine vergleichsweise hohe Auflage verfügen. Zudem wurde bei der Auswahl 3. darauf geachtet, dass die letztlich verwendeten Medien in der Summe ein möglichst breites Ressortspektrum abdeckten. Die ausgewählten Medien, samt Beitragszahl und Auflage, sind in Tabelle 2 zu finden.

Tabelle 2: Ausgewählte Medien

Medium	Gesamtzahl der Beiträge	Ausgewählte Beiträge	Auflage[5]	Ressort
Neue Zürcher Zeitung	913	199	85.261 (2016)	Allgemein
Tages-Anzeiger	290	77	147.146 (2016)	Allgemein
Cash Online	212	173	61.000	Finanzen
Handelszeitung	194	105	35.711 (2018)	Wirtschaft, Politik, Finanzen
St. Galler Tagesblatt	180	54	122.958 (2016)	Allgemein, regional
SonntagsZeitung	152	20	147.566 (2018)	Allgemein, eher boulevardesk
Elektronik-Praxis	144	60	–	Technik
	2085	**688**		

Durch diese Kriterien blieben 688 Beiträge übrig, die manuell von zwei Codern zu je gleichen Teilen gesichtet wurden. Diese bewerteten zuerst die Relevanz der Nachricht und des Anwendungsfalls für diese Studie. Zu jedem passenden Anwendungsfall wurde eine eindeutige ID vergeben und kategorisiert, ob es sich um Stimm,- Sprach- oder Gesichtserkennung sowie Identifikation oder Verifikation handelt. In einem iterativen Prozess wurden Codes für den Anwendungsfall (Was?) sowie den Anwendungsbereich (Wo?) festgelegt. Zusätzlich wurde erfasst, ob der Anwendungsfall im öffentlichen oder privaten Raum sowie durch eine private Firma oder öffentliche Institution eingesetzt wird. Zusätzlich zu den 688 Beiträgen wurden auch weitere Beiträge in die Liste aufgenommen, auf die das Projektteam im Laufe der Literaturrecherche gestossen war. Im Ergebnis wurde eine Liste bestehend aus rund 90 Anwendungsmöglichkeiten der Stimm-, Sprach- und Gesichtserkennung erstellt (vgl. die zugehörige Tabelle im Onlineappendix[6]).

[5] Auflage laut WEMF AG für Werbemedienforschung, erreichbar via Webarchiv unter: https://web.archive.org/web/20190116102939/https://wemf.ch/de/downloads/audit-statistcs/auflagebeglaubigung/wemf-auflagebulletin-2018.pdf.

[6] 10.5281/zenodo.6838769.

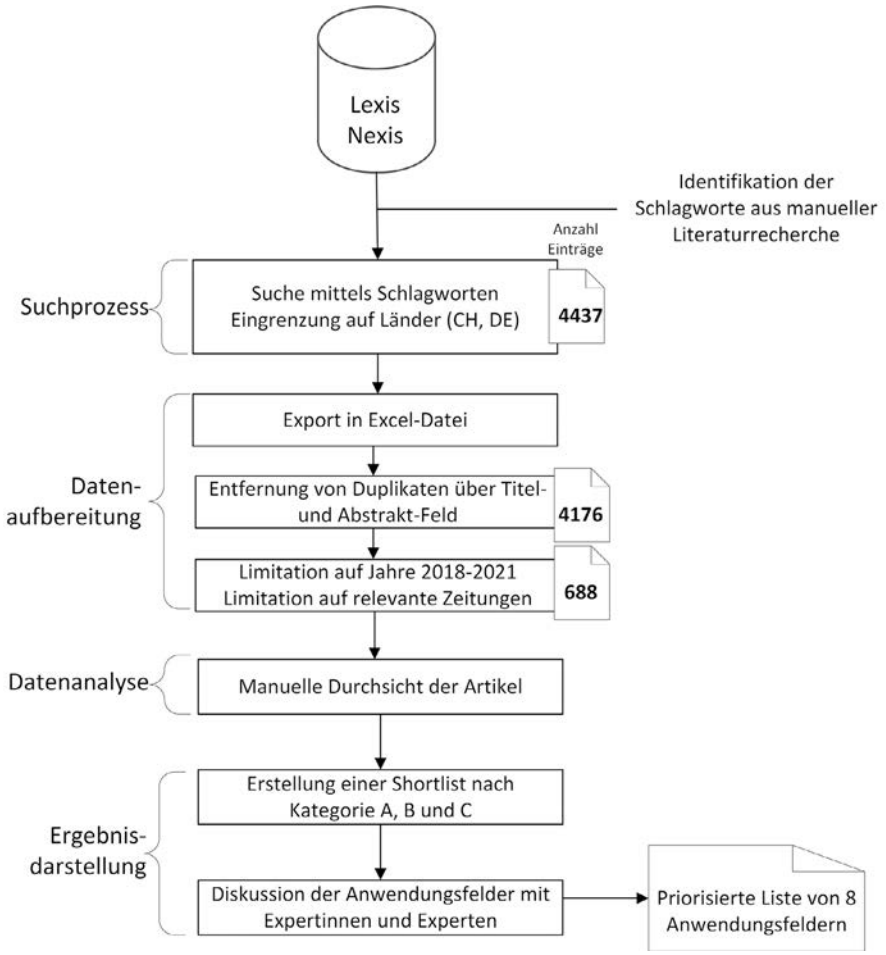

Abbildung 4: Identifikation der zu untersuchenden Anwendungsfälle und -gebiete

Im nächsten Schritt wurde ein dreistufiges Ranking dieser Anwendungsmöglichkeiten durch vier Coder durchgeführt. Das Ziel des Rankings bestand in der Selektion der zu untersuchenden Anwendungsfälle und -gebiete. Auswahlkriterien waren:

- Die Erreichung einer Mischung aus aktuell bereits praktizierten Anwendungen sowie von mittel- bis langfristig relevanten Anwendungen.
- Spezifische Relevanz für die Schweiz.
- Inklusion sowohl von Stimm- als auch Sprach- und Gesichtserkennung.
- Die Systeme sollten sowohl seitens öffentlicher und privatwirtschaftlicher Stellen als auch seitens Privatpersonen eingesetzt werden.
- Schliesslich sollten die Anwendungen in Privathaushalten, privatwirtschaftlichen und öffentlichen Bereichen anzutreffen sein.

Auf Grundlage dieser Kriterien wurden der Begleitgruppe 17 Anwendungsmöglichkeiten vorgestellt. Aus der gemeinsamen Diskussion fiel die Entscheidung für die folgenden acht Anwendungsgebiete.

Nr.	Anwendungsgebiet	Technologie	Einsatzort	Betrieb seitens
1	Smarte Lautsprecher	Stimme und Sprache	Privater Lebens-bereich	Gewerblich
2	Gesichts- und Spracherken-nung durch polizeiliche Stellen	Gesicht	Öffentlicher Raum, privater (gewerb-licher) Raum	Öffentlich
3	Authentifizierung via Stimme beim Telefonbanking	Stimme	Übergreifend	Gewerblich
4	Gewaltprävention und -aufklärung in Sportstadien	Gesicht	Privater (gewerb-licher) Raum	Gewerblich
5	Erkennung physischer und psychischer Krankheiten	Stimme und Sprache und Gesicht	Öffentlicher Raum, privater Raum	Öffentlich, gewerblich, individuell
6	Emotionserkennung	Stimme, Sprache und Gesicht	Übergreifend	Gewerblich und öffentlich
7	Aufmerksamkeitsanalyse in Schulen	Gesicht	Semi-Öffentlicher Raum	Gewerblich und öffentlich
8	Jedermann-Identifikation	Gesicht	Übergreifend	Übergreifend

3. Analyse der Anwendungsgebiete

Die folgende Untersuchung der Anwendungsgebiete umfasst die drei Aspekte:

- **Analyse der technischen Grundlagen und Möglichkeiten:** Bewertung der technischen Leistungsfähigkeit von Stimm-, Sprach- und Gesichtserkennungsdiensten (vgl. auch FF 2.1).[7]
- **Juristische Bewertung insb. der datenschutzrechtlichen Rahmenbedingungen:** Untersuchung der Rechtskonformität der Anwendungsfelder, mit einem Fokus auf die datenschutzrechtlichen Regelungen sowie, je nach Anwendungsfeld, unter Einbezug weiterer einschlägiger Rechtsvorschriften (vgl. auch FF 2.2).[8]
- **Erörterung gesellschaftlicher und ethischer Herausforderungen:** Diskussion der möglichen gesellschaftlichen und ethischen Herausforderungen, wenn Stimm-, Sprach- und Gesichtserkennungstechnologien in den diskutierten Anwendungsfeldern zum Einsatz kommen (vgl. FF 2.3).[9]

Einführend werden zunächst (3.1) die grundlegenden juristischen Vorgaben dargelegt. Daran schliesst sich (3.2) eine Einführung in die Grundlagen zur Diskussion der gesellschaftlichen und ethischen Herausforderungen an. Schliesslich werden alle Anwendungsgebiete (Unterkapitel 3.3–3.10) im Hinblick auf die oben genannten drei Aspekte untersucht.

3.1. Juristische Grundlagen

Einschlägige rechtliche Vorgaben für die Stimm-, Sprach- und Gesichtserkennung ergeben sich aus Völker-, Europa-, Verfassungs- und einfachgesetzlichen Grundlagen. Im Folgenden wird der sich daraus ergebende Rahmen für die Stimm-, Sprach- und Gesichtserkennung übersichtsweise dargestellt. Eine eingehende Anwendung dieses Rahmens erfolgt sodann spezifisch für die einzelnen Anwendungsgebiete in den folgenden Unterkapiteln. Zu unterscheiden ist für diesen allgemeinen Überblick zwischen Handlungen des Staates (3.1.1) sowie Handlungen Privater (3.1.2), wobei gewisse Vorgaben, namentlich die datenschutzrechtlichen Grundsätze, auch für öffentliche und private Akteure gleichermassen anwendbar sind (3.1.3).

---

[7] Die entsprechenden Kapitel wurden von Frank Ebbers, mit Unterstützung von Murat Karaboga und Michael Friedewald verfasst.

[8] Die entsprechenden Kapitel wurden von Nula Frei und Sophia Rovelli verfasst.

[9] Die entsprechenden Kapitel wurden von Murat Karaboga, mit Unterstützung von Frank Ebbers und Michael Friedewald verfasst.

3.1.1. Verwendung von Stimm-, Sprach- und Gesichtserkennungstechnologien durch den Staat

Sind es staatliche Behörden (auf Bundes-, Kantons- oder Gemeindeebene sowie Private, die öffentliche Aufgaben wahrnehmen), welche diese Technologie zu verwenden beabsichtigen, müssen sie sowohl die allgemeinen verfassungsrechtlichen Vorgaben für das Staatshandeln beachten wie auch die spezialgesetzlichen, namentlich datenschutzrechtlichen, Vorschriften einhalten. Insb. muss jegliches staatliches Handeln auf einer gesetzlichen Grundlage beruhen (Legalitätsprinzip, Art. 5 Abs. 1 BV) und verhältnismässig sein (Art. 5 Abs. 2 BV).

Tangiert der Einsatz der Technologie eines oder mehrere Grundrechte, so sind die Anforderungen an Grundrechtseinschränkungen von Art. 36 BV zu beachten. Betroffene Grundrechte sind bei der Verwendung von Stimm-, Sprach- und Gesichtserkennungstechnologien durch den Staat in erster Linie das Recht auf Privatsphäre (Art. 13 Abs. 1 BV) sowie der Anspruch auf Schutz vor Missbrauch der persönlichen Daten (Art. 13 Abs. 2 BV). Jedoch können auch weitere Grundrechte tangiert sein, bspw. die Kommunikationsgrundrechte (Meinungsäusserungsfreiheit, Versammlungs- und Vereinigungsfreiheit, Art. 16, 22, 23 BV), auf welche die staatliche oder private Verwendung von Stimm-, Sprach- oder Gesichtserkennungstechnologie einen *chilling effect* ausüben kann (Schreiber und Joss 2020), die Verfahrensgrundrechte (Art. 29, 29a, 30 BV) sowie das Verbot der Diskriminierung (Art. 8 Abs. 2 BV), welches etwa dann relevant werden kann, wenn eine bestimmte Technologie für bestimmte Personen (z.B. Frauen, persons of colour) eine höhere Falschidentifizierungsrate zeigt (Buolamwini und Gebru 2018; FRA 2019).

Grundrechtseingriffe benötigen eine gesetzliche Grundlage, welche den Anforderungen an die Bestimmtheit genügt (Art. 36 Abs. 1 BV). Schwerwiegende Eingriffe müssen in einem Gesetz im formellen Sinn vorgesehen sein; ob es sich um einen schwerwiegenden Einzelfall handelt, muss im Einzelfall – hier also in Bezug auf die konkrete Technologie und ihre konkrete Verwendung – beurteilt werden. Grundrechtseingriffe müssen durch ein zulässiges öffentliches Interesse gerechtfertigt werden, wobei hierbei dem Gesetzgeber ein Gestaltungsspielraum zukommt, solange er sich nicht in Widerspruch zu verfassungsrechtlichen Wertentscheidungen begibt. Schliesslich müssen Grundrechtseingriffe im Lichte ihrer Eingriffsintensität und des verfolgten öffentlichen Interesses verhältnismässig sein, d.h. geeignet, erforderlich und zumutbar. Auch die Verhältnismässigkeit richtet sich nach der konkreten Ausgestaltung im Einzelfall. Zum Schluss darf ein Grundrechtseingriff nicht den Kerngehalt der betroffenen Grundrechte aushöhlen. Im Kontext der Verwendung von Überwachungstechnologie wäre der Kerngehalt des Rechts auf Privatleben namentlich dann tangiert, wenn die Überwachung einen nicht näher eingegrenzten Personenkreis trifft oder wenn durch eine permanente Überwachung ein diffuses Gefühl konstanter Überwachung, mithin ein «gläserner Bürger», entsteht.

Darüber hinaus ist der Staat nicht nur bei eigener Verwendung dieser Technologien durch die Grundrechte verpflichtet. Gemäss Art. 35 Abs. 3 BV sorgen die Behörden dafür, dass die Grundrechte, soweit sie sich dazu eignen, auch unter Privaten wirksam werden. In Bezug auf die vorliegende Thematik heisst dies, dass bei festgestellten Gefährdungen von Grundrechten durch die private Verwendung von Stimm-, Sprach- oder Gesichtserkennungstech-

nologie – wobei es sich hier um die gleichen Grundrechte wie oben ausgeführt handeln dürfte – auch ein Auftrag an den Staat besteht, bspw. durch den Erlass von Vorschriften dafür zu sorgen, dass solche Verletzungen unterbleiben bzw. geahndet werden können.

Über die verfassungsrechtlichen Vorgaben hinaus müssen staatliche Stellen, die Stimm-, Sprach- oder Gesichtserkennungstechnologie einsetzen wollen, selbstredend auch sämtliche bestehenden einfachgesetzlichen Regelungen beachten, namentlich das Datenschutzrecht von Bund respektive Kantonen sowie die datenschutzrechtlichen und sonstigen Vorgaben in den jeweiligen Fachgesetzen und -verordnungen. Auf die allgemeinen datenschutzrechtlichen Vorgaben des neuen Datenschutzgesetzes (nDSG)[10], das voraussichtlich am 1. September 2023 in Kraft tritt, wird noch eingegangen (sogleich sowie unten, 3.1.3), während die spezialgesetzlichen Regelungen sowie, falls einschlägig, die kantonalen Datenschutzgesetze bei den jeweiligen Anwendungsfällen untersucht werden.

Eine Bundesbehörde darf Daten nur dann bearbeiten, wenn sie dafür über eine gesetzliche Grundlage verfügt. Hierfür taugt grundsätzlich auch eine Verordnungsbestimmung, ausser es handelt sich um besonders schützenswerte Personendaten oder Persönlichkeitsprofile. Für Letztere bedarf es einer Grundlage in einem formellen Gesetz,[11] nach dem nDSG zudem ebenfalls für Datenbearbeitungen, die nach ihrem Bearbeitungszweck oder der Art und Weise der Bearbeitung zu einem schwerwiegenden Eingriff in die Grundrechte der betroffenen Person führen können.[12] Eine Bearbeitung ohne eine hinreichende Rechtsgrundlage ist aber in Ausnahmefällen möglich, namentlich wenn die betroffene Person eingewilligt oder ihre Daten allgemein zugänglich gemacht hat.[13] Die gleichen Anforderungen an eine gesetzliche Grundlage gelten für die Datenbekanntgabe durch Bundesorgane.[14]

3.1.2. Spezifische Bearbeitungsvorschriften für Private

Für die Datenbearbeitung durch Private gilt, dass Datenbearbeitungen erlaubt sind, sofern sie keine widerrechtliche Persönlichkeitsverletzung darstellen.[15] Eine Persönlichkeitsverletzung liegt insb. vor, wenn Personendaten entgegen den Datenschutzgrundsätzen oder entgegen der ausdrücklichen Willenserklärung der betroffenen Person bearbeitet werden sowie bei Bekanntgabe von besonders schützenswerten Personendaten an Dritte (Rosenthal 2020). Die Persönlichkeitsverletzung ist widerrechtlich, wenn sie nicht durch Einwilligung des Betroffenen oder durch ein überwiegendes privates oder öffentliches Interesse oder durch Gesetz gerechtfertigt ist (Rampini 2014).[16] Im Zusammenhang der Datenbearbeitung durch Stimm-, Sprach- und Gesichtserkennungstechnologien durch Private kommt, neben

---

[10] In der vorliegenden Studie wird aufgrund des baldigen Inkrafttretens konsequent auf die Bestimmungen des neuen Datenschutzgesetzes (nDSG) verwiesen.

[11] Art. 34 nDSG.

[12] Art. 34 Abs. 2 lit. c nDSG.

[13] Art. 34 Abs. 4 lit. b nDSG.

[14] Art. 36 Abs. 1 nDSG.

[15] Art. 30 Abs. 1 nDSG.

[16] Art. 28 Abs. 2 ZGB.

der Einwilligung (dazu unten noch näher), insb. der Rechtfertigungsgrund der überwiegenden Interessen infrage.[17] Es muss stets im Einzelfall geprüft werden, ob das Interesse des Bearbeiters dasjenige des Betroffenen überwiegt.

Der Rechtsschutz gegen Datenbearbeitungen durch Private ist auf dem Klageweg im Rahmen des Zivilrechts wahrzunehmen. Die betroffene Person kann insb. das Verbot einer bestimmten Datenbearbeitung oder einer Datenbekanntgabe sowie die Löschung oder Vernichtung von Personendaten verlangen.[18] Kann weder die Richtigkeit noch die Unrichtigkeit der betreffenden Personendaten festgestellt werden, so kann ein Bestreitungsvermerk angebracht werden.[19]

3.1.3. Vorgaben für Behörden und Private: Die datenschutzrechtlichen Grundsätze

Über diese spezifischen Regeln für öffentliche und für private Datenbearbeiter hinaus stellt das Datenschutzgesetz eine Vielzahl von Vorschriften auf, die für alle Kategorien von Datenbearbeitern gelten.

Anknüpfungspunkt für die Anwendbarkeit des Datenschutzgesetzes ist das Vorliegen von Personendaten. Personendaten sind alle Angaben, die sich auf eine bestimmte oder bestimmbare natürliche Person beziehen, wobei der Begriff weit zu fassen ist.[20] Sofern mit der Stimm-, Sprach- und Gesichtserkennung Daten bearbeitet werden, die sich auf eine natürliche Person beziehen, ist die Anwendbarkeit des Datenschutzgesetzes gegeben. Anders wäre es z.B., wenn die Daten anonymisiert oder pseudonymisiert worden sind und es keine technische Möglichkeit mehr gibt, daraus auf eine natürliche Person zu schliessen. Dies dürfte bei den hier untersuchten Anwendungsfällen nie der Fall sein.

Werden Personendaten bearbeitet, sind die allgemeinen Bearbeitungsgrundsätze zu beachten (Art. 4, 5, 7 DSG resp. Art. 6 und 8 nDSG). Die Datenbearbeitung muss *rechtmässig* sein, darf also nicht gegen eine Norm des objektiven Rechts verstossen.[21] Sie muss nach *Treu und Glauben* erfolgen, darf mithin also nicht treuwidrig sein.[22] Der Grundsatz von Treu und Glauben verlangt loyales und vertrauenswürdiges Verhalten im Rechtsverkehr. Der Grundsatz kann als Generalklausel verstanden werden, die immer dann zur Anwendung gelangt, wenn andere Bearbeitungsgrundsätze nicht anwendbar sind (Epiney und Nüesch 2015, S. 69). Aus dem Grundsatz kann auch der Grundsatz der Transparenz jeglicher Datenbearbeitungsschritte abgleitet werden (Rosenthal 2020).[23] Gemäss dem Grundsatz der *Transparenz* muss sowohl die Datenbeschaffung als auch deren Zweck für die Betroffenen

---

[17] Art. 32 Abs. 2 nDSG.

[18] Art. 32 Abs. 2 nDSG.

[19] Art. 32 Abs. 3 nDSG.

[20] Art. 5 Bst. a nDSG; siehe BGE 147 I 346, Urteile 1B_510/2017 vom 11. Juli 2018 E. 3.3; 1C_509/2016 vom 9. Februar 2017 E. 3.1; 1C_74/2015 vom 2. Dezember 2015 E. 3.2.

[21] Art. 6 Abs. 1 nDSG.

[22] Art. 6 Abs. 2 nDSG.

[23] Der Art. 6 Abs. 3 nDSG bezieht sich nur auf den Datenbearbeitungsschritt des Erhebens, welcher erkennbar sein muss. Art. 6 Abs. 2 nDSG ist für die anderen Datenbearbeitungsschritte ausschlaggebend.

erkennbar sein.[24] Die Bearbeitung durch private wie auch öffentliche Stellen muss *verhältnismässig* sein.[25] Ein besonderer datenschutzrechtlicher Grundsatz ist derjenige der Zweckbindung, demnach muss die Datenbearbeitung mit dem bestimmten Zweck, zu dem sie erfolgt, vereinbar sein und dieser Zweck muss für die betroffene Person im Zeitpunkt der Beschaffung erkennbar sein.[26] Sobald die Daten für diesen Zweck nicht mehr erforderlich sind, müssen sie vernichtet oder anonymisiert werden.[27] Der Grundsatz der *Datenrichtigkeit* verlangt, dass der Bearbeiter sich über die Richtigkeit der bearbeiteten Personendaten in Hinblick auf den Zweck der Beschaffung oder der Bearbeitung vergewissern und alle angemessenen Massnahmen treffen muss, damit in diesem Sinne unrichtige Daten berichtigt, gelöscht oder vernichtet werden.[28] Der Grundsatz der *Datensicherheit* schliesslich verpflichtet den Bearbeiter (Verantwortliche wie Auftragsbearbeiter) dazu, durch technische und organisatorische Massnahmen Verletzungen der Datensicherheit zu vermeiden; diese Massnahmen müssen dem Risiko der Datenbearbeitung angemessen sein.[29] Der Grundsatz der Datensicherheit verlangt, dass die Vertraulichkeit, Integrität und Verfügbarkeit von Personendaten sichergestellt wird (Rosenthal 2020). Wird die Datensicherheit nicht gewährleistet, kann es zu schwerwiegenden Konsequenzen für die Betroffenen kommen, etwa durch eine unautorisierte Bekanntgabe von besonders schützenswerten Personendaten an Dritte (CoE 2021).

Die Anforderungen an die Einwilligung sind mit dem neuen Datenschutzgesetz präzisiert worden.[30] Die Einwilligung stellt in der Praxis einen der wichtigsten Rechtfertigungsgründe für Datenbearbeitungen durch Private dar und kommt auch bei Bearbeitungen durch Bundesorgane zum Zug, namentlich wenn eine gesetzliche Grundlage fehlt. Eine Einwilligung ist nur gültig, wenn sie für eine oder mehrere Bearbeitungen nach angemessener Information freiwillig erteilt wird.[31] Für die Bearbeitung besonders schützenswerter Personendaten, ein Profiling durch ein Bundesorgan oder ein Profiling mit hohem Risiko durch eine private Person muss die Einwilligung ausdrücklich erfolgen.[32]

Besonders schützenswerte Personendaten sowie das Profiling sind in Art. 3 DSG resp. Art. 5 nDSG definiert. Hier stellt sich insb. die Frage, ob Daten, die im Kontext der Stimm-, Sprach- und Gesichtserkennung beschafft und bearbeitet werden, besonders schützenswerte Personendaten darstellen, was insb. für die Kategorie der biometrischen Daten, die eine natürliche Person eindeutig identifizieren,[33] der Fall sein könnte. Fraglich ist deshalb,

---

[24] Art. 6 Abs. 3 nDSG.

[25] Art. 6 Abs. 2 nDSG.

[26] Art. 6 Abs. 3 nDSG.

[27] Art. 6 Abs. 3 nDSG.

[28] Art. 6 Abs. 5 nDSG.

[29] Art. 8 Abs. 1 und 2 nDSG sowie Verordnung vom 14. Juni 1993 zum Bundesgesetz über den Datenschutz (VDSG), SR 235.11.

[30] Der ursprüngliche Vorschlag des Bundesrates, dass die Einwilligung auch *eindeutig* erfolgen muss, wurde in den parlamentarischen Beratungen allerdings wieder gestrichen.

[31] Art. 6 Abs. 6 nDSG.

[32] Art. 6 Abs. 7 nDSG.

[33] Art. 5 Bst. c Ziff. 4 nDSG.

ob Daten im Kontext der Stimm-, Sprach- und Gesichtserkennung *biometrische Daten darstellen, die eine Person eindeutig identifizieren*. Darunter fallen Daten, die durch ein spezifisches technisches Verfahren zu den physischen, physiologischen oder verhaltenstypischen Merkmalen eines Individuums gewonnen werden und die eine eindeutige Identifizierung der betreffenden Person ermöglichen oder bestätigen. Es handelt sich dabei bspw. um einen digitalen Fingerabdruck, Gesichtsbilder, Bilder der Iris oder Aufnahmen der Stimme. Diese Daten müssen zwingend auf einem spezifischen technischen Verfahren beruhen, das die eindeutige Identifizierung oder Authentifizierung einer Person erlaubt. Dies ist bspw. grundsätzlich nicht der Fall bei gewöhnlichen Fotografien (Bundeskanzlei 2017). Mit dieser Definition folgt das revidierte DSG dem Ansatz der DSGVO[34]. Es ist somit zu differenzieren: Sofern bei den Stimm-, Sprach- und Gesichtserkennungstechnologien lediglich eine Aufnahme (bspw. eine Fotografie, eine Video- oder eine Tonbandaufnahme) verwendet wird, ohne dass diese mittels technischer (automatisierter) Verfahren zur eindeutigen Identifikation einer Person führt, liegt keine Bearbeitung besonders schützenswerter Personendaten vor. Werden diese Technologien allerdings gerade zwecks Identifikation, Verifikation oder Authentifizierung der Identität von Personen verwendet, so ist zweifellos von besonders schützenswerten Personendaten auszugehen. Dabei reicht es aus, dass die Personendaten mit speziellen technischen Mitteln bearbeitet werden, die die eindeutige Identifizierung oder Authentifizierung einer natürlichen Person ermöglichen, es kommt also nicht auf die tatsächliche Wahrnehmung dieser Möglichkeit an (Council of Europe 1981, S. 11). Dies steht im Einklang mit der bundesgerichtlichen Rechtsprechung.[35]

Des Weiteren ist zu beachten, dass die Bearbeitungsvorgänge im Kontext der Stimm-, Sprach- oder Gesichtserkennungstechnologien auch aus anderen Gründen besonders schützenswerte Personendaten darstellen können, selbst wenn dadurch eine Person nicht eindeutig identifiziert wird. Wird die Technologie zur Erkennung physischer oder psychischer Krankheiten eingesetzt, werden Daten über die Gesundheit[36] bearbeitet; bei Überwachung etwa von Kundgebungen oder Streiks im öffentlichen Raum oder des Eingangs zu religiösen Einrichtungen könnten Erkenntnisse über religiöse, politische oder gewerkschaftliche Tätigkeiten[37] gewonnen werden (Nguyen und Alexander 2020).

Letztlich bleibt also festzuhalten, dass die Verwendung von Stimm-, Sprach- und Gesichtserkennungstechnologien bei einer grossen Anzahl der Fälle eine Bearbeitung von besonders schützenswerten Personendaten darstellt, wenngleich im jeweiligen Einzelfall zu differenzieren ist.

Ebenfalls stellt sich die Frage, ob mit bestimmten Anwendungen der Stimm-, Sprach- und Gesichtserkennung ein *Profiling* respektive ein *Profiling mit hohem Risiko* einhergeht. Ein Profiling liegt vor bei jeder Art von automatisierter Datenbearbeitung, die Daten verwendet, um bestimmte persönliche Aspekte einer natürlichen Person zu analysieren oder vorherzusagen, worunter insb. die Arbeitsleistung, wirtschaftliche Lage, Gesundheit, persönliche

---

[34] Art. 4 Ziff. 14 und Art. 9 Abs. 1 DSGVO, vgl. auch Art. 3 Ziff. 13 Richtlinie 2016/680.

[35] Vgl. BGE 136 II 508 (*Logistep*).

[36] Art. 5 Bst. c Ziff. 2 nDSG.

[37] Art. 5 Bst. c Ziff. 1 nDSG.

Vorlieben, Interessen, Zuverlässigkeit, Verhalten, Aufenthaltsort oder Ortswechsel dieser Person gehören.[38] Ein Profiling mit hohem Risiko liegt bei einem Profiling vor, das durch Verknüpfung von Daten eine Beurteilung wesentlicher Aspekte der Persönlichkeit einer natürlichen Person erlaubt und damit ein hohes Risiko für die Persönlichkeit oder die Grundrechte dieser Person mit sich bringt.[39] Die Qualifikation der Verwendung von Stimm-, Sprach- oder Gesichtserkennungstechnologie als eine Form des Profiling (mit oder ohne hohem Risiko) kann nur im konkreten Anwendungsfall unter Berücksichtigung aller Umstände beantwortet werden und wird deshalb weiter hinten bei der Beurteilung der einzelnen Anwendungsfälle wieder aufgegriffen.

Die Qualifizierung als besonders schützenswerte Personendaten oder als Profiling hat folgende datenschutzrechtliche Konsequenzen:

- Bei Datenbearbeitung durch Private muss die *Einwilligung* ausdrücklich erfolgen (bei Profiling allerdings nur, wenn es sich um ein Profiling mit hohem Risiko handelt).[40]
- Ebenso dürfen Private keine besonders schützenswerte Personendaten *weitergeben.*[41]
- Eine Datenbearbeitung durch ein Bundesorgan muss auf einer *formellen gesetzlichen Grundlage* beruhen.[42]
- Ist eine *Verhältnismässigkeitsabwägung* zu treffen, etwa beim Rechtfertigungsgrund des überwiegenden Interesses sowie allgemein im Rahmen des datenschutzrechtlichen Verhältnismässigkeitsgrundsatzes, so dürfte diese bei besonders schützenswerten Personendaten sowie bei Profiling eher zugunsten der Interessen der betroffenen Person und gegen die Interessen des Bearbeiters ausfallen.

Weitere datenschutzrechtliche Vorschriften, die allgemein bei der Verwendung von Stimm-, Sprach- und Gesichtserkennungstechnologien beachtet werden müssen, sind die Rechte der Betroffenen, namentlich deren Recht auf Information (das sich als Informationspflicht des Bearbeiters bei der Beschaffung von Personendaten äussert[43]), das Auskunftsrecht[44], das Recht auf Datenherausgabe oder -übertragung[45] sowie das Berichtigungsrecht.[46]

Werden Personendaten ins Ausland übermittelt, sind die Vorschriften von Art. 16 ff. nDSG zu beachten. Sofern die betroffenen Personen nicht im Einzelfall ausdrücklich in die Bekanntgabe eingewilligt haben[47] oder ein anderer Ausnahmetatbestand[48] vorliegt, dürfen Per-

---

38 Art. 5 Bst. f nDSG.

39 Art. 5 Bst. g nDSG.

40 Art. 5 Bst. c Ziff. 4 nDSG.

41 Art. 30 Abs. 2 lit. c nDSG.

42 Art. 34 Abs. 2 nDSG; siehe auch Bundesrat, Botschaft DSG, BBl 2017, 82.

43 Art. 19–21 nDSG.

44 Art. 25, 26 nDSG.

45 Art. 28, 29 nDSG.

46 Art. 32 Abs. 1 und 41 Abs. 2 nDSG.

47 Art. 17 Abs. 1 lit. a nDSG.

48 Namentlich Vertragsabwicklung, überwiegende öffentliche Interessen oder Rechtsansprüche, Notfälle, allgemeine Zugänglichmachung oder gesetzlich vorgesehenes Register, Art. 17 Abs. 1 lit. b–f nDSG.

sonendaten nur dann ins Ausland bekannt gegeben werden, wenn ein Angemessenheitsentscheid des Bundesrates vorliegt oder der Datenschutz durch völkerrechtlichen Vertrag, vertragliche Datenschutzklauseln, Standarddatenschutzklauseln oder unternehmensinterne Datenschutzvorschriften gewährleistet wird.[49]

Weitere Pflichten der Bearbeiter umfassen namentlich die datenschutzfreundliche Ausgestaltung der Datenbearbeitung («privacy by design and by default»),[50] das Führen eines Verzeichnisses der Bearbeitungstätigkeiten,[51] die Durchführung einer Datenschutzfolgenabschätzung[52] sowie die Meldung von Verletzungen der Datensicherheit[53].

3.2. Grundlagen zur Diskussion der gesellschaftlichen und ethischen Herausforderungen

Über Möglichkeiten der missbräuchlichen Verwendung der Gesichtserkennungstechnologie und wie diesen Einhalt geboten werden könnte, wird seit über zwei Jahrzehnten kritisch diskutiert (Steve Wright 1998). Vergleichbare lange zurückreichende Diskussionen zu Stimm- und Spracherkennungstechnologien hat es keine gegeben. Diese haben allerdings in den letzten Jahren im Hinblick auf konkrete stimm- und spracherkennungsbasierte Anwendungen zugenommen, beispielweise Sprachassistenten (Suarez 2021) oder dem Einsatz im medizinischen Bereich (Villongco und Khan 2020).

Vor dem Hintergrund des kontinuierlichen Ausbaus der Videoüberwachungsinfrastruktur in Europa, einer unzureichenden regulatorischen Einbettung der Technologie und der Erwartung, dass der Einsatz von Gesichtserkennung mit Beginn des neuen Jahrtausends ebenso zunehmen würde, wurden ab der zweiten Hälfte der 1990er-Jahre zunächst Verhaltenskodexe und Best Practices ausgearbeitet. Mit diesen sollte der Einsatz der Technologie entlang bestimmter Prinzipien geregelt werden. Zu den Autoren dieser Dokumente zählten zunächst Interessengruppen aus Wirtschaft und Zivilgesellschaft sowie lokale Behörden. Auffällig an dieser frühen Debatte ist, dass in den entsprechenden Dokumenten neben menschenrechtlichen Gesichtspunkten insb. auf Datenschutzbedenken eingegangen wurde (Webster 2004).

Aus dieser ersten Phase der Erarbeitung von Verhaltenskodizes sind insb. jene von Steve Wright aus dem Jahr 1998 (Steve Wright 1998) und von Chris Pounder aus dem Jahr 2008 (Pounder 2008) hervorzuheben. Beide Kodizes stellen hilfreiche Annäherungen an die Bewertung von video- bzw. gesichtserkennungsbasierten Überwachungspraktiken dar und weisen Überlappungen mit der zeitgenössischen wissenschaftlichen Debatte zur Einhegung von KI-Technologien auf. Fragen hins. einer möglichen Implementation bzw. Operationalisierung der Kodizes blieben allerdings unbeantwortet. Erweiterte Hilfestellung hinsicht-

[49] Art. 16 Abs. 1 und 2 nDSG.

[50] Art. 7 nDSG.

[51] Art. 12 nDSG.

[52] Art. 22 nDSG.

[53] Art. 24 nDSG.

lich der Implementation bot insb. das von der britischen Datenschutzbehörde im Jahr 2000 veröffentlichte «CCTV Code of Practice» (Information Commissioner 2000). Diese Kodizes beziehen sich zwar unmittelbar auf Videoüberwachungssysteme, doch waren sie v.a. an die Betreiber der Systeme gerichtet, um den ordnungsgemässen Betrieb der Systeme sicherzustellen. In Bezug auf die Bewertung von Datenbearbeitungen hins. ihrer Datenschutzkonformität wurden parallel zu den Kodizes Vorschläge für sog. *Privacy Impact Assessments* entwickelt. Diese hatten teils die Begleitung der Entwicklung von Datenbearbeitungsprozessen zur Gewährleistung der Datenschutzkonformität von der Planung bis zur Umsetzung einer Datenbearbeitung zum Ziel (was später als Privacy by Design bekannt werden sollte) und teilweise die Evaluation einer Datenbearbeitung vor deren Inbetriebnahme (Friedewald et al. 2017).

Derartige Versuche der gesellschaftsvertraglichen Einhegung des zunehmenden Ausbaus von Überwachungstechnologien ernteten jedoch im Laufe der 2000er-Jahre zunehmende Kritik aus dem Lager der Surveillance-Forschenden. Im Fokus der Kritik stand zum einen die philosophische Basis des Datenschutzes bzw. der Privatheit. So wurde argumentiert, dass Datenschutz bzw. Privatheit, verstanden als individuelle Praxis, die ihren Ursprung im liberalen Individualismus hat, die dem Einzelnen den Rückzug aus der Gesellschaft und damit auch die Freiheit von Überwachung ermöglichen soll, nicht mehr der Realität einer zunehmend vernetzten Gesellschaft entspreche. Mehr noch, könne die in diesem Verständnis etablierte Idee des individuellen Rückzugs aus der Gesellschaft nicht dazu taugen, den Ausbau von Überwachungstechnologien abzuwehren, da invasive Überwachungspraktiken immer dann den Vorzug gegenüber Datenschutz bzw. Privatheit erhalten würden, sofern sie nur ausreichend gut klarmachten, dass sie wichtigen gesamtgesellschaftlichen Zielen (öffentliche Sicherheit, Betrugsprävention, u.v.m.) und nicht bloss dem individuellen Streben nach Privatheit dienten (Stalder 2002; Gilliom 2011). Zusammenhängend mit dieser Kritik wurde zum anderen bemängelt, dass Datenschutz bzw. Privatheit nur einen Aspekt schützenswerter ethischer Gesichtspunkte darstelle. Demnach seien trotz der Allgegenwart von Datenschutzgesetzen und daraus abgeleiteten Verhaltenskodizes sehr weitreichende Datenbearbeitungen mit weitreichenden Folgen für Individuen und Gesellschaft möglich, unabhängig davon, ob diese auf Basis der intendierten Einhaltung aller Datenschutzpflichten erfolge oder durch die Ausnutzung etwaiger gesetzlicher Spielräume (Pohle 2016). In anderen Worten: Die aus datenschutzrechtlich unbedenklichen Datenbearbeitungen resultierenden und sich verschärfenden Machtungleichgewichte zwischen Individuen gegenüber Staat und Unternehmen blieben unberücksichtigt und würden ihren Schatten nicht zuletzt auf die Funktionsfähigkeit der Demokratie selbst werfen (Seubert 2017). Christen et al. (2020, S. 293) weisen auf diese Herausforderung in der TA SWISS-Studie «Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz» ebenfalls hin: «Der heutige Ansatz des Datenschutzrechts ist auf den Vorgang der Bearbeitung von Personendaten ausgerichtet und weitestgehend blind für die Folgen dieser Datenbearbeitungen.» Das heisst, solange das Datenschutzrecht eingehalten wird, bleiben andere ethisch fragwürdige Praktiken der Datenbearbeiter unberücksichtigt und trotz der Einhaltung des Datenschutzrechts können sonstige unerwünschte Auswirkungen Resultat dieser Bearbeitungen sein.

Unter anderem vor dem Hintergrund dieser Kritiken entstanden seit Mitte der 2000er-Jahre Vorschläge für sog. Surveillance Impact Assessments. Im Unterschied zu Privacy Impact Assessments haben Surveillance Impact Assessments (SIA) die Bewertung von

Überwachungssystemen und -technologien nicht nur im Hinblick auf deren Datenschutz- bzw. Privatheitsfolgen zum Ziel, sondern auch im Hinblick auf deren gesellschaftliche, wirtschaftliche, politische, rechtliche und ethische Folgen. Nach Wright et al. ist ein SIA als ein Risikoabschätzungsprozess zu verstehen, der zunächst die Wahrscheinlichkeit des Eintritts von Risiken identifiziert, die damit verbundenen Konsequenzen aufzeigt und schliesslich zur Eindämmung der identifizierten Risiken dient. Ein SIA soll durch die für die jeweilige Datenbearbeitung Verantwortlichen unter Einbindung von weiteren Interessengruppen in den Bewertungsprozess durchgeführt und im Falle von relevanten Änderungen der bewerteten Überwachungspraktik wiederholt werden müssen (Wright und Raab 2012; Wright et al. 2015). Dieser Trend, dass bei den Vorschlägen für Folgenabschätzungen Datenschutz als ein Aspekt und nicht als Kernaspekt betrachtet werden sollte, setzte sich auch in den Folgejahren fort (Raab 2020), so insb. in Form von Ethical Impact Assessments (Wright und Friedewald 2013), Human Rights Impact Assessments (Mantelero 2018) und schliesslich Algorithmic Impact Assessments (Reisman et al. 2018). Weiterhin ist das Ziel dieser Impact Assessment-Frameworks die Etablierung einer Verpflichtung für die Betreiber zur Bestimmung der Kritikalität einer Datenbearbeitung und im Falle des Vorhandenseins von Risiken zur Einführung von Schutzmassnahmen.

In der Debatte drückt sich der Wandel aus, dass Datenschutz zwar weiterhin klar als ein wichtiges schützenswertes Gut angesehen wird, das einen zentralen Baustein zur Einhegung der negativen Folgen von modernen Datenbearbeitungen und Überwachungstechnologien darstellt. Zugleich wird aber deutlich, dass diese Datenbearbeitungen und Überwachungstechnologien Folgen mit sich bringen, die über das Datenschutzrecht hinausgehen und dass zu deren Einhegung weitere Massnahmen erforderlich sind, die über das Datenschutzrecht hinausweisen. Diese Einsichten spiegeln sich auch in den aktuellen Debatten zur Einhegung der negativen Folgen von künstlicher Intelligenz wider, in deren Rahmen hilfreiche Vorschläge erarbeitet wurden, die einen Beitrag zur Diskussion der gesellschaftlichen und ethischen Herausforderungen von Stimm-, Sprach- und Gesichtserkennungsanwendungen leisten können und im folgenden Unterkapitel vorgestellt werden.

3.2.1. Adressierung ethischer Gesichtspunkte in Datenbearbeitungen und daten- und KI-basierten Anwendungen: Die ethische Wende

Vor dem Hintergrund der ungebremsten Digitalisierung der Gesellschaft, die in zunehmendem Masse auf der Bearbeitung grosser Datensätze mittels fortschrittlicher Algorithmen basiert, und angesichts der Erkenntnis, dass die Gewährleistung des Datenschutzes zwar einen elementaren Bestandteil gesellschaftsvertraglicher Digitalisierung darstellt, dies allein jedoch aller Voraussicht nach alleine nicht zur Erreichung dieses Ziels genügt, waren die letzten Jahre von der Suche nach weiteren Lösungsmöglichkeiten geprägt. Das Ergebnis dieser Suche war der sog. «Ethical Turn». Forschende, Behörden und Unternehmen aus unterschiedlichen Bereichen forcierten die ethische Wende: Philosophie (Floridi und Taddeo 2016), Soziologie, Psychologie, Marktforschung (Raab 2020), Medizin (Mittelstadt und Floridi 2016) sowie Datenschutz (EDPS 2015) und Surveillance Studies (Lyon 2014). Dabei sind derartige Ansätze nicht an einzelne Technologien gebunden, sondern beziehen sich auf Digitaltechnologien und KI-basierte Anwendungen im Allgemeinen und sind deshalb sowohl auf Stimm- und Sprach- als auch auf Gesichtserkennungstechnologien anwendbar.

Die Frage nach der ethischen Reflexion hat in den letzten Jahren auch Eingang in die Wissenschaftscommunity der Gesichtserkennungsforschung gefunden. So stiess bspw. die Anwendung von Gesichtserkennung auf Uiguren und andere chinesische Minderheiten auf eine breite Front der Kritik in der Forschungscommunity. Eine wachsende Zahl an Wissenschaftlern hat zudem zwischenzeitlich damit begonnen, Aufrufe bezüglich der kritischen Evaluation der Eingebundenheit der eigenen Zunft in möglicherweise unethische Projekte oder bezüglich der Gefahr der unethischen Weiternutzung von Forschungsergebnissen zu initiieren. Journals und Konferenzen haben damit begonnen, Ethiküberprüfungen der eingereichten Beiträge durchzuführen. Die Ergebnisse einer Umfrage unter 480 Forschenden zu Gesichtserkennung veranschaulichen, dass sich die Mehrzahl der Befragten mit den meisten modernen gesichtserkennungsbasierten Überwachungspraktiken (Echtzeitüberwachung des öffentlichen Raums durch öffentlichen Stellen oder Private, Emotionserkennung in der Schule, Jedermann-Identifikation, Emotionserkennung bei Bewerbungsgesprächen) unwohl fühlt (van Noorden 2020).

Der Ethical Turn spiegelt sich auch in der Entstehung und weltweiten Verbreitung einer Vielzahl an Ethikräten, Gremien, Ausschüssen, Prinzipien, Richtlinien, Projekten und Ähnlichem wider, die sich der gesellschaftsvertraglichen Einhegung von zunächst Big Data, später Algorithmen, Machine Learning und künstlicher Intelligenz verschrieben haben. Zu den Initiatoren zählen sowohl privatwirtschaftliche Unternehmen, Forschende als auch behördliche und politische Institutionen (Fjeld et al. 2020).

Seinen deutlichsten Ausdruck im politischen Raum in der Schweiz fand dieser Prozess in der Einsetzung der verwaltungsinternen «Interdepartementalen Arbeitsgruppe künstliche Intelligenz», die im Rahmen des Aktionsplans «Digital Schweiz» im September 2018 ins Leben gerufen wurde (Bundesrat 2018).[54] Diese war zwischen dem dritten Quartal 2019 und November 2020 für die Erarbeitung von Empfehlungen zum Umgang mit künstlicher Intelligenz verantwortlich. In dem Bericht der Arbeitsgruppe vom Dezember 2019 zu den «Herausforderungen der künstlichen Intelligenz» (WBF 2019) werden vier Cluster als Herausforderungen identifiziert (WBF 2019, S. 36–37):

1. «Autonomie, Verantwortlichkeit und Haftung»,
2. «Nachvollziehbarkeit und Transparenz»,
3. «Bias und Diskriminierung» sowie
4. «Datenzugang und Datenschutz».[55]

Auf Grundlage des Berichts der Arbeitsgruppe beauftragte der Bundesrat das WBF damit, in Zusammenarbeit mit UVEK und der Arbeitsgruppe strategische Leitlinien für den Umgang mit den Herausforderungen der künstlichen Intelligenz auf Ebene des Bundes bzw. in der Bundesverwaltung auszuarbeiten. In dem im November 2020 veröffentlichten Ergebnis-

[54] Bestehend aus Vertretern von armasuisse, GS-UVEK, SBFI, EDA, BFS, SIF, GS-EFD, SECO, BAKOM, BJ, EZV und WBF (GDS 2020).

[55] Etwas weniger umfassend, aber im Grundsatz affirmierend, sind auch die Empfehlungen der Schweizerischen Akademie der Technischen Wissenschaften (SATW 2019).

bericht finden sich die sieben Leitlinien (WBF, UVEK und Interdepartementale Arbeitsgruppe künstliche Intelligenz 2020):

1. Den Menschen in den Mittelpunkt stellen,
2. Rahmenbedingungen für Entwicklung und Anwendung von KI,
3. Transparenz, Nachvollziehbarkeit und Erklärbarkeit,
4. Verantwortlichkeit,
5. Sicherheit,
6. Aktive Mitgestaltung der Governance von KI,
7. Einbezug aller relevanten nationalen und internationalen Akteure.

In der «Strategie Digitale Schweiz» vom September 2020 finden sich schliesslich folgende fünf Ziele (GDS 2020, S. 5):

1. Chancengleiche Teilhabe aller ermöglichen und Solidarität stärken,
2. Sicherheit, Vertrauen und Transparenz gewährleisten,
3. Digitale Befähigung und Selbstbestimmung der Menschen weiter stärken,
4. Wertschöpfung, Wachstum und Wohlstand sicherstellen,
5. Ökologischen Fussabdruck und Energieverbrauch verringern.

Während der erste Bericht der Arbeitsgruppe Herausforderungen benennt, werden in den anderen Dokumenten Leitlinien bzw. Ziele formuliert. Dadurch unterscheidet sich zwar die Stossrichtung der drei Dokumente etwas, ein Vergleich der dort benannten Punkte zeigt allerdings, dass sich in fünf Feldern Übereinstimmungen finden lassen (vgl. Tabelle 19). Dies betrifft die Punkte bzw. Prinzipien:

- Nachvollziehbarkeit und Transparenz
- Fairness, Bias und Diskriminierung
- Verantwortlichkeit, Rechenschaftspflicht
- Datenschutz, Privatheit
- Sicherheit

Im Bereich der Schweizer Wirtschaft wird momentan versucht, mittels der Setzung von ethischen Standards eine Vorreiterrolle zu übernehmen. Dies hat sich die «Swiss Digital Initiative» (SDI) zum Ziel gesetzt, die auf dem Weltwirtschaftsforum 2020 lanciert wurde und die unter ihrem Dach sowohl Schweizer Unternehmen und Universitäten als auch Partner aus anderen Staaten versammelt. Mittels Einführung eines sog. Digital Trust Labels soll Nutzerinnen und Nutzern Aufschluss über vier Dimensionen gegeben werden: (1) Sicherheit, (2) Datenschutz, (3) Zuverlässigkeit des Systems,[56] und (4) Faires User Management (Transparenter Umgang mit automatisierten Entscheidungsfindungen) (Swiss Digital Initia-

[56] Der Inhalt dieser Dimension entspricht dem Prinzip der Gerechtigkeit bzw. Fairness.

tive 2021). Die genannten vier Dimensionen entsprechen den im politischen Raum verhandelten Prinzipien, lediglich das Prinzip der Verantwortlichkeit bzw. Rechenschaftspflicht ist nicht enthalten.

Schliesslich finden sich die im politischen Raum diskutierten fünf Ethikprinzipien sowie die vier Prinzipien der SDI auch in zahlreichen weiteren weltweit entstandenen Dokumenten, die sich dem Thema der Gewährleistung ethischer KI widmen, wie die Ergebnisse dreier einschlägiger Metastudien, von Jobin et al. (2019), Fjeld et al. (2020) sowie Hagendorff (2020), demonstrieren. Ein Vergleich der drei Metastudien (vgl. Tabelle 3) zeigt, dass in allen Publikationen in wesentlichen inhaltlichen Punkten übereinstimmend fünf Prinzipienbündel hervorstechen. Diese werden im Folgenden kurz vorgestellt.

Privatheit, Datenschutz. Privatheit und Datenschutz werden laut Jobin et al. (2019, S. 395) sowohl als ein aufrechtzuerhaltender Wert, der eng mit Autonomie verbunden ist, als auch als ein schutzbedürftiges Rechtsgut aufgefasst. In das Spektrum Privatheit und Datenschutz fallen nach Fjeld et al. (2020, S. 21) acht Unterpunkte, darunter die Einwilligung und Betroffenenrechte wie das Bestehen von Lösch- und Widerspruchsmöglichkeiten.

Verantwortlichkeit und Rechenschaftspflicht. Angesichts der erwarteten Zunahme KI-basierter automatisierter Entscheidungen wird darauf verwiesen, dass auch solche Entscheidungen letztlich auf einen Verantwortlichen (ob als natürliche oder juristische Person) zurückgeführt werden können müssen, um etwa Haftungsansprüche geltend machen oder wirksam Beschwerde einlegen zu können. Verantwortlichkeit und Rechenschaftspflicht heben aber auch die Bedeutung der Durchführung von Impact Assessments in der Designphase von KI-basierten Systemen und die Durchführung von Evaluationen und Audits hervor.

Sicherheit und Schutz vor Schäden. Dieses Prinzipienbündel baut insb. auf der Erwartung, dass KI-Systeme voraussichtlich sehr weitreichende gesellschaftliche und individuelle Auswirkungen haben werden und dass Angriffe auf die Vertraulichkeit und Integrität der KI-Systeme oder ihr Missbrauch unvorhersehbare Konsequenzen nach sich ziehen könnten. Folglich geht es zum einen um Gesichtspunkte wie die Gewährleistung von IT-Sicherheit und zum anderen um den Schutz der Betroffenen vor möglichen durch KI-Einsatz entstehenden Schäden.

Denn während im Falle einer *Kompromittierung eines Pins oder Passworts dieses sich schnell und einfach ändern lässt, ist dies bei biometrischen Daten wie der* Stimme oder dem Gesicht eines Menschen nicht mehr möglich. Einmal kompromittierte Daten bleiben für immer (Sheldon 2020).

Transparenz und Erklärbarkeit. Die Forderung nach der Gewährleistung von Transparenz basiert auf der Feststellung, dass KI-Systeme trotz ihrer teils schon heute weitreichenden Folgen oftmals nicht als solche gekennzeichnet werden. Folglich können Betroffene teilweise nicht erkennen, dass sie Gegenstand einer KI-basierten Entscheidung sind. Zudem beraubt die praktizierte Intransparenz über den Einsatz von KI-basierten Systemen die Öffentlichkeit der Möglichkeit der Deliberation über entsprechende Einsätze. Die Forderung nach der Gewährleistung von Erklärbarkeit wiederum gründet auf dem Umstand, dass die Gründe für KI-basierte Entscheidungen, insb. wenn Entscheidungen auf neuronalen Netzen basieren, häufig selbst seitens der Programmierenden nicht mehr erklärt werden können

(Blackbox-Problem). Daher wird diesbezüglich gefordert, dass möglichst relevante Auskunft bezüglich der verwendeten Trainingsdaten, zugrunde liegenden Algorithmen und, wenn möglich, über die Entscheidungen selbst gegeben wird.

Gerechtigkeit, Fairness und Nicht-Diskriminierung. Dieses Prinzipienbündel basiert auf der Einsicht, dass KI-basierte Systeme sich gegenüber der Welt und den Menschen nicht neutral verhalten, sondern stets von ihren sozialen Entstehungskontexten geprägt sind und somit bestehende Ungerechtigkeiten replizieren.

Die Arbeiten von Fjeld et al. (2020) und Hagendorff (2020) identifizieren ausserdem übereinstimmend eine sechste und siebte Dimension. «**Menschliche Kontrolle der Technik**» verweist auf die Forderung, dass KI-basierte Analyseergebnisse vor ihrer Anwendung von einem Menschen kontrolliert und interpretiert werden sollten. Auf die Relevanz dieses Punktes weisen auch Medienberichte hin, wonach Strafverfolgungsbehörden die falsch-positiven Treffer der Gesichtserkennungssoftware ohne ausreichende Überprüfung der Ergebnisse zur Verhaftung unschuldiger Menschen herangezogen hatten (Springer et al. 2018). Allerdings ist dieser Punkt insofern kontrovers, weil auch Menschen zu Fehlern neigen. Dies liegt zum einen daran, dass Nutzer entweder den Algorithmen nicht trauen, sich selbst für besser qualifiziert halten oder das Ergebnis schlicht nicht einordnen können (Stevenson 2018). Andere akzeptieren hingegen nur Ergebnisse, welche deren eigene Voreingenommenheit unterstützen (Skeem et al. 2020).

«**Förderung menschlicher Werte**» verweist auf die normative Forderung, dass KI-basierte Anwendungen stets auf die Förderung des Gemeinwohls im weitesten Sinne ausgerichtet sein sollten.

Hierzu zählen in zunehmendem Masse auch die **Umweltwirkungen digitaler Infrastrukturen**. Nach Jahrzehnten des ungebremsten Wachstums digitaler Infrastrukturen, wird derzeit – noch mit starkem Fokus auf Blockchain-Währungen – vermehrt über die Vermeidung unnötigen Energie- und Ressourcenverbrauchs diskutiert. Doch auch der Anstieg des Datenvolumens, v.a. durch Gesichtserkennungsanwendungen, kann zu einem erhöhten Energie- und Ressourcenverbrauch führen: Schon heute machen rund 80 % des Datenverkehrs in Telekommunikationsnetzen Videoinhalte aus (UBA 2020, 11 f.).[57] Die prognostizierte Zunahme der Gesichtserkennung wird zu einer weiteren Zunahme insb. des Energieverbrauchs führen.

Tabelle 3: Überblick über die in den einschlägigen Metastudien identifizierten Kern-Prinzipien für ethische KI (eigene Zusammenstellung)

	Fjeld et al. (2020)	**Jobin et al. (2019)**	**Hagendorff (2020)**	**(Bertelsmann Stiftung und VDE 2020)**
Privacy	X	X	X	X
Accountability, Responsibility	X	X	X	X

[57] Datendurchsatz Video: Ultra-HD-Video = 7 GB/h; Datendurchsatz Audio: 320 kbit/MP3 = rund 100 MB/h.

	Fjeld et al. (2020)	Jobin et al. (2019)	Hagendorff (2020)	(Bertelsmann Stiftung und VDE 2020)
Safety and Security	X	X	X	X (Reliability)
Transparency and Explainability	X	X	X	X
Justice, Fairness and Non-discrimination	X	X	X	X
Human Control of Technology	X		X	
Professional Responsibility	X			
Promotion of Human Values	X		X (Do good)	
Environmental Sustainability				X

3.2.1.1. Methodisches Vorgehen

Grundsätzlich stützt sich die Diskussion der ethischen Herausforderungen auf die Analyse der einschlägigen Literatur zu den jeweiligen Anwendungsgebieten. Zusätzlich wurden in den Anwendungsgebieten der polizeilichen Gesichtserkennung, der Authentifizierung mittels Stimme sowie der Stadionüberwachung Dienstebetreiber kontaktiert und um die Beantwortung eines Ethik-Kriterien-Fragekatalogs gebeten. Die oben diskutierten fünf zentralen Ethikprinzipien dienen bei der Diskussion sowohl der Literatur als auch der empirischen Datenerhebung als grundlegende Orientierung.

Der Ethik-Kriterien-Fragekatalog (siehe Tabelle 20 im Anhang) basiert auf den in Abschnitt 3.2.1 diskutierten fünf zentralen Ethik-Prinzipienbündeln Privatheit/Datenschutz, Verantwortlichkeit/Rechenschaftspflicht, Sicherheit, Transparenz/Erklärbarkeit sowie Gerechtigkeit/Fairness/Nicht-Diskriminierung. Hilfreiche Antworten auf die Frage, wie sich die Prinzipien operationalisieren lassen, bieten insb. eine Studie der Bertelsmann-Stiftung, in der die beteiligten Forschenden Vorschläge zur Operationalisierung von KI-Ethik formulieren, sowie ein Aufsatz von Castelluccia und Le Métayer (2020), die eine Risiko-Analyse-Methodolo-

gie zur Impact-Analyse von Gesichtserkennungsanwendungen vorschlagen.[58] In der Bertelsmann-Studie wird ein mehrschichtiges Evaluationssystem vorgeschlagen, das auf der obersten Ebene den entsprechenden Wert (Privatheit, Gerechtigkeit, Transparenz usw.) beinhaltet und auf den niedrigeren Stufen Kriterien, Indikatoren und Beobachtungen vorsieht. Beispielsweise werden dem Wert Transparenz die Kriterien «Offenlegung der Herkunft von Datensätzen», «Offenlegung der Eigenschaften des verwendeten Algorithmus/Modells» und «Zugänglichkeit» zugeordnet. Um dann das Kriterium der Offenlegung der Herkunft von Datensätzen zu untersuchen, werden wiederum vier Indikatoren-Fragen vorgeschlagen «Ist die Herkunft der Daten dokumentiert?», «Ist für jeden Zweck plausibel, welche Daten genutzt wurden/werden?» und «Sind die Eigenschaften des Trainingsdatensatzes dokumentiert und offengelegt? Sind die entsprechenden Datenblätter umfassend?». Aus der Studie von Castelluccia und Le Métayer (2020) stammen bspw. das Kriterium «Was ist der erwartete Nutzen des Systems?» und die zugehörigen Indikatoren-Fragen: «Welchen Interessen ist dieser Nutzen zuträglich (privat/öffentlich, Staat, Zivilgesellschaft, einzelne Bürger)?» sowie «Welche Auswirkungen hat das System auf die übrigen Interessengruppen?». Wir verfolgen mit dem Rückgriff auf den Ethik-Kriterien-Fragekatalog im Rahmen dieser Studie das Ziel eines tiefer gehenden Verständnisses der untersuchten Anwendungen und nicht einer mehr oder weniger quantifizierbaren Bewertung dieser.

In den nun folgenden Kapiteln werden die Anwendungsgebiete untersucht.

3.3. Smarte Lautsprecher

Viele Jahrzehnte lang galt die Idee, Computer per Sprachbefehl zu steuern, als Science-Fiction. Wissenschaftler als auch Marktforscher gehen allerdings zunehmend davon aus, dass visuelle und haptische Interfaces allmählich von Sprachinterfaces abgelöst werden (Hoy 2018; Frost & Sullivan 2019, S. 221). Auf Grundlage von Fortschritten im Bereich der künstlichen Intelligenz und des maschinellen Lernens haben smarte Lautsprecher im Laufe des vergangenen Jahrzehnts zunächst Eingang in den Markt und allmählich auch in den Alltagsgebrauch gefunden. Apple veröffentlichte Siri erstmals im Jahr 2010, Microsoft folgte mit Cortana in 2013. Amazon veröffentlichte Alexa 2014 und Google Assistant 2016 (Yoffie et al. 2018, S. 25). Die wesentlichen Funktionen von smarten Lautsprechern umfassen (Hoy 2018, S. 83):

- Versenden und Vorlesen von Textnachrichten, Starten von Telefonanrufen sowie Versenden und Vorlesen von E-Mails.
- Beantworten einfacher Fragen («Wie spät ist es?», «Wie ist die Wettervorhersage?» usw.).
- Einstellen von Timern, Weckern, Kalendereinträgen usw.

[58] In unseren Recherchen sind wir auf keine Studien zur systematischen Untersuchung der spezifischen Folgen von Stimm- und Spracherkennungstechnologien gestossen. Wir sind allerdings der Ansicht, dass sich die Vorschläge aus dem KI-Ethik-Bereich sowie jene von Castelluccia und Le Métayer (2020) sowohl auf Gesichts- als auch auf Stimm- und Spracherkennungstechnologien anwenden lassen, da ihre technologische Funktionsweise sowie ihre Anwendungsweise in wesentlichen Aspekten übereinstimmen.

- Einstellen von Erinnerungen, Erstellung von Listen und Durchführung von mathematischen Berechnungen usw.
- Kontrolle der Mediennutzung über verbundene Dienste wie Amazon, Google Play, iTunes, Netflix, Spotify usw.
- Kontrolle von Internet-der-Dinge-Geräten wie Thermostaten, Lichtquellen, Tür-/Fensterschlössern usw.
- Einfache Unterhaltungsfunktionen, etwa das Erzählen von Witzen oder Geschichten.

Zudem beherrschen die Geräte der verschiedenen Hersteller auch herstellerspezifische Funktionen. Amazon Alexa ermöglicht Amazon-Käufe per Sprachbefehl. Der Google Assistant hingegen merkt sich, wo das Auto geparkt wurde und erinnert den Nutzer auf Nachfrage an den Stellplatz. Nutzerbefragungen zeigen jedoch, dass die Geräte hauptsächlich zum Abspielen von Musik, Radiohören, Internetsuchanfragen und die Steuerung von Haushaltsgeräten genutzt werden (Klö ss 2020; Ammari et al. 2019).

Smarte Lautsprecher basieren auf sog. Sprachassistenten wie Apples Siri oder der Google Assistent und stehen auf einer Reihe unterschiedlicher Geräte zur Verfügung. Siri wurde bspw. zunächst als App auf dem iPhone und später auch auf iPads, Apple Macs und der Apple Watch angeboten. Microsofts Cortana wurde zunächst auf Windows Phones und später auch auf Computern mit dem Betriebssystem Windows 10 angeboten. Google Assistant erschien ebenfalls zunächst auf Smartphones (Yoffie et al. 2018). 2014 bot Amazon den Assistenten Alexa für festinstallierte Wohnzimmergeräte an. Erst auf diese Weise fand die Technologie stärkeren Eingang in die Alltagsnutzung. 2016 veröffentlichte auch Google eine intelligente Lautsprecherreihe mit dem Namen Google Home. Apple betrat den Markt mit dem HomePod im Jahr 2018. Parallel arbeiten die Hersteller an der Integration ihrer Software in eine Reihe weiterer Geräte, etwa Smart Cars und Smart TVs (Hoy 2018). Zugleich lassen sich Sprachassistenten in zunehmendem Masse in herkömmliche Lautsprechersysteme integrieren. Ein Beispiel dafür ist die SYMFONISK-Reihe von Ikea, die in Kooperation mit Sonos entwickelt wurde und eine Steuerung mittels Amazon Alexa und Google Nest ermöglicht (Ikea 2021).

Amazons frühe Expansion auf dem Markt intelligenter Lautsprecher[59] führte dazu, dass bis Anfang 2017 etwa 80 % aller weltweit verkauften Geräte von Amazon stammten. Allerdings hat sich der Markt durch den Markteintritt der Konkurrenz deutlich diversifiziert. Amazon setzt (Stand 2020) weltweit zwar noch immer die meisten Geräte ab und hält einen Marktanteil von etwa 22 %. Google hat jedoch zwischenzeitlich deutlich aufgeholt und folgt auf dem zweiten Platz mit einem weltweiten Marktanteil von ca. 17 % (Strategy Analytics 2020).

Etwas anders stellt sich die Situation mit Blick auf die Schweiz dar: Zum einen ist generell ein geringerer Nutzungsgrad festzustellen. Eine repräsentative Bevölkerungsumfrage der Universität Luzern ermittelte, dass nur 1 % der Schweizer Bevölkerung im Jahr 2018 und nur 3 % im Jahr 2019 intelligente Lautsprecher nutzten (Kunath et al. 2020). Demgegenüber nutzten in Deutschland 2018 bereits 10 % der Bevölkerung intelligente Lautsprecher (Brien 2018). Zum anderen haben intelligente Lautsprecher den Schweizer Markt erst spät

[59] Innerhalb von wenigen Jahren hat sich die Zahl weltweit verkaufter Geräte von etwa einer Million Geräten auf mehr als 30 Millionen Geräte vervielfacht (Strategy Analytics 2020).

und teils noch gar nicht betreten. Das wurde v.a. auf das Sprachverständnis der zahlreichen Schweizer Dialekte zurückgeführt. So gab die Mehrheit der Schweizer Nutzer von Sprachassistenten 2019 an, von der Software eher nicht verstanden zu werden (Vuichard 2020). Weder Apple noch Amazon betraten mit der HomePod- bzw. Echo-Reihe zunächst den Schweizer Markt. Seit dem offiziellen Marktaustritt Amazons aus der Schweiz (Melchior 2018) wurde allerdings darüber spekuliert, ob Amazon Alexa losgelöst vom Versandhandel in der Schweiz einführen wird (Rivas 2021). Auf dem Schweizer Markt für smarte Lausprecher präsent ist hingegen Google mit der Nest-Reihe. Allerdings erfolgte auch die Einführung der Google-Produkte mit Verspätung Ende 2019. Die Studienlage deutet darauf hin, dass Google in der Schweiz Marktführer bei smarten Lautsprechern ist (Mcschindler 2019).

Weil die Google Nest-Reihe bereits auf dem Schweizer Markt präsent ist und die Amazon Alexa-Reihe möglicherweise eingeführt werden könnte, stehen beide Produkte im Zentrum der folgenden Analysen. Zudem ermöglicht die Analyse beider Produkte auch einen Vergleich zwischen zwei unterschiedlichen smarten Lautsprechern.

3.3.1. Technische Grundlagen und Möglichkeiten

Der grundlegende technische Aufbau von smarten Lautsprechern ist – unabhängig vom Hersteller – ähnlich (Park et al. 2019, S. 1076). So verfügen alle smarten Lautsprecher über mindestens ein Mikrofon, einen Lautsprecher und sind über WLAN/Internet mit den Servern des jeweiligen Anbieters verbunden (Dempsey 2017, S. 80; Lau et al. 2018, S. 1). Einige Modelle verfügen zudem über einen Bluetooth-Chip, über den Geräte in der Nähe kontaktiert werden können. Zudem besitzen die Geräte noch einige hardwarebasierte Knöpfe.

Das Mikrofon dient dazu, Sprachbefehle entgegenzunehmen. Die Geräte hören praktisch immer jedes gesprochene Wort und Geräusch mit («always on») und warten auf ein sog. Aktivierungswort. Diese Erkennung bis zum Aktivierungswort wird meist unmittelbar auf den Geräten selbst bearbeitet (Lau et al. 2018, S. 3). Jedoch haben mehrere Studien gezeigt, dass diese Erkennung oft nicht korrekt verläuft und die Spracherkennungsfunktion immer wieder auch ungewollt aktiviert wird (St. John 2020; Dubois et al. 2020). In der englischen Sprache wurden folgende Begriffe als Aktivierungswort am meisten falsch verstanden:

- «okay to go» statt «Hey, Google» (Google Home Mini)
- «Hey, Missy» statt «Hey, Siri» (Apple HomePod)
- «quartet» statt «Cortana» (Microsoft, Harman-Kardon)

Erst nach der Erkennung des Aktivierungswortes beginnt die Aufzeichnung der Stimme. In einigen Geräten gibt es jedoch auch einen physischen Knopf, um die Spracherkennung zu aktivieren oder das Mikrofon ein- bzw. auszuschalten (Lau et al. 2018, S. 3). Die Sprachaufnahme wird danach an die Server der Hersteller gesendet, wo mittels Spracherkennung, dem sog. Natural Language Processing (NLP), eine Auswertung des Befehls erfolgt. Dabei versuchen verschiedene Algorithmen die Sprache zu verstehen, um daraus die Intention des Nutzers abzuleiten. Im Erfolgsfall wird der Befehl an weitere Internet- oder externe Dienste (oft bezeichnet als Skills) delegiert (Edu et al. 2021, S. 4). Dort wird dieser bearbeitet und schliesslich das Ergebnis über eine automatisierte Sprachausgabe (text to speech)

ausgegeben. Die Lautsprecher dienen dann als Feedback zum Nutzer, indem eine Antwort des Systems an den Nutzer ausgegeben wird, z.B. wie das Wetter wird. In anderen Anwendungsfällen können sie anderen Zwecken dienen, etwa dem Abspielen von Musik (Park et al. 2019, S. 1076). Das «Ökosystem» der smarten Lautsprecher besteht also aus drei Komponenten: dem Lautsprecher selbst, einem cloudbasierten Service und einem Set von definierten nutzbaren Funktionen (Skills). Der Ablauf ist Abbildung 5 zu entnehmen.

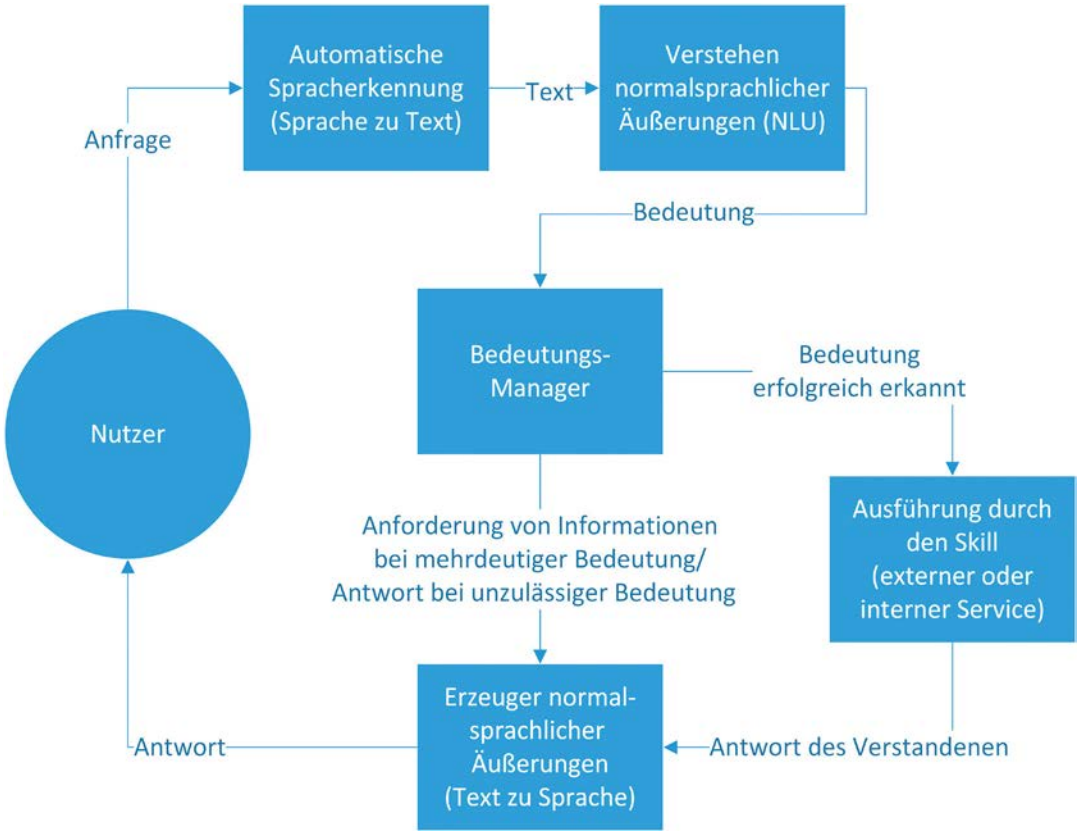

Abbildung 5: Ablauf der Sprach- und Sinnerkennung in Smart Speaker in Anlehnung an Edu et al. (2021, S. 6)

Zur Beantwortung der Anfragen der Nutzenden werden die Sprachaufzeichnungen über das verbaute WLAN-Modul zur Analyse an die Server der Hersteller versendet und Statusinformationen ausgetauscht. Dabei werden umfangreiche (personenbezogene) Daten bearbeitet, wie eine direkte Sprachaufnahme des Aktivierungswortes oder die E-Mail-Adresse des Nutzenden (Akinbi und Berry 2020, S. 271). Aufgezeichnete Tonaufnahmen lassen sich bei vielen Geräten über eine Webseite oder eine verbundene App löschen (Lau et al. 2018, S. 3). Diese Funktionen sind allerdings vielen Nutzenden nicht bekannt bzw. werden von diesen kaum verwendet. Zudem ist unklar, wie effektiv eine Löschung ist. Sie umfasst nämlich in der Regel nur die Rohdaten, nicht aber die Informationen, die die Hersteller daraus abgeleitet haben (Kelly 2019).

Für unterschiedliche Anwendungen nutzen smarte Lautsprecher verschiedene personenbezogene Daten. Vielen Nutzenden dürfte klar sein, dass Audioaufzeichnungen der Stimme für eine Sprachsteuerung nötig sind. Jedoch werden auch Daten erhoben, die nicht direkt ersichtlich sind und zu Profilen zusammengestellt werden, z.B. das Datum inkl. der Uhrzeit einer Anfrage, um daraus eventuell Rückschlüsse auf den Tagesablauf der Person zu ziehen. Weiter werden Kontodaten wie Name, Wohnort und Zahlungsmodalitäten mit Interaktionen des Lautsprechers verknüpft. Als eine Art «Beifang» werden auch die Stimmhöhe der sprechenden Person sowie die Stimmen von anderen Menschen, die zur Aufnahmezeit ggf. im Hintergrund zu hören sind, mit aufgezeichnet (Edu et al. 2021, S. 7; Natatsuka et al. 2019). Durch die Interaktion mit Skills können weitere Daten des Nutzers abgefragt werden, etwa das Geburtsdatum für Erinnerungen oder die Blutgruppe für Gesundheitsanwendungen (Edu et al. 2021, S. 7). Des Weiteren können verschiedene (personenbezogene) Daten übertragen werden, wenn der smarte Lautsprecher mit anderen smarten Geräten, wie Videotürklingeln, interagiert. Aus all diesen Daten bzw. der Analyse dieser Daten lässt sich zudem auf Verhaltensweisen der Nutzer schliessen (Chung et al. 2017). Dadurch werden bspw. Rückschlüsse auf den physischen Gesundheitszustand der Nutzenden oder die Beziehung zwischen mehreren Personen, etwa den in einem Haushalt lebenden Menschen, die mit dem Lautsprecher interagieren, möglich (Edu et al. 2021, S. 8).

Die steigende Nutzung und starke Integration smarter Lautsprecher in alltägliche Abläufe («Domestikation») lockt zudem Kriminelle an (Hoppenstedt 2019). So zeigte sich, dass mittels einer Aufnahme der Stimme des Nutzenden die Authentifizierungsmechanismen der Lautsprecher getäuscht werden können und somit Fremde Zugriff auf sensible Daten erhalten können (Replay-Attacken) (Malik et al. 2019). Auch Erweiterungen für die smarten Lautsprecher stellen eine potenzielle Gefahr für Nutzer da. Diese können, wie ein Computer-Virus (z.B. ein Trojaner), Daten des Nutzers (z.B. Tonaufnahmen oder Passwörter) abfangen und an mögliche Angreifer senden (Zhang et al. 2019).

Forschende haben sich dieser Datenschutz- und Sicherheitsprobleme angenommen (Edu et al. 2021, S. 24–25), beispielsweise durch die Entfernung von Sprachcharakteristiken, die nicht für die Interaktion mit einem smarten Lautsprecher nötig sind (Vaidya und Sherr 2019). Mitev et al. (2020) entwickelten eine Netzwerk-Analysesoftware, mit der die Übertragung von Audiodateien über das Internet kenntlich gemacht werden kann. So können Nutzende unabhängig vom Gerät gewarnt werden, dass das Aktivierungswort ausgelöst wurde. Geforscht wird auch am Schutz vor Replay-Attacken (Malik et al. 2019, S. 523). Zudem gibt es Forschung im Bereich der Aussortierung von Geräuschen, etwa der Gespräche anderer Menschen, die nicht für die Nutzung eines Audio-Interfaces nötig sind (Xia und Jiang 2020).

Tabelle 4: Technische Daten zu Google Home/Nest

Hersteller:	Google
Sitz des Herstellers:	USA
Einführung:	2016 (USA), 2016 (Schweiz)
Anzahl Mikrofone:	2 (Google Home Mini) (Dempsey 2017, S. 81)
Besonderheiten:	Schiebeknopf zum Ausschalten des Mikrofones

Google vertreibt unter unterschiedlichen Namen eine Vielzahl von smarten Geräten. Neben Geräten zur Heimautomatisierung wie smarten Thermostaten oder Überwachungskameras unter dem Firmennamen Nest auch smarte Lautsprecher, wie z.B. den Google Home Mini (Tabelle 4). Damit können umfassende Daten aus einem privaten Haushalt zusammengeführt werden und z.B. durch den Nest Hub, ein Gerät, das einem Tablet ähnelt, gesteuert werden (Zuboff 2018, S. 231–269). Auf diesen Geräten läuft der Google Assistant, der mittels Sprachbefehlen gesteuert werden kann. Dazu werden weitere Cloud-Services, wie z.B. Musikabspielen, ermöglicht. Zusätzlich können auch weitere Geräte wie smarte Glühbirnen mit dem Google Assistant per Sprache gesteuert werden (Gupta 2018, S. 2074). Abbildung 6 zeigt den schematischen Aufbau des Nest-Ökosystems.

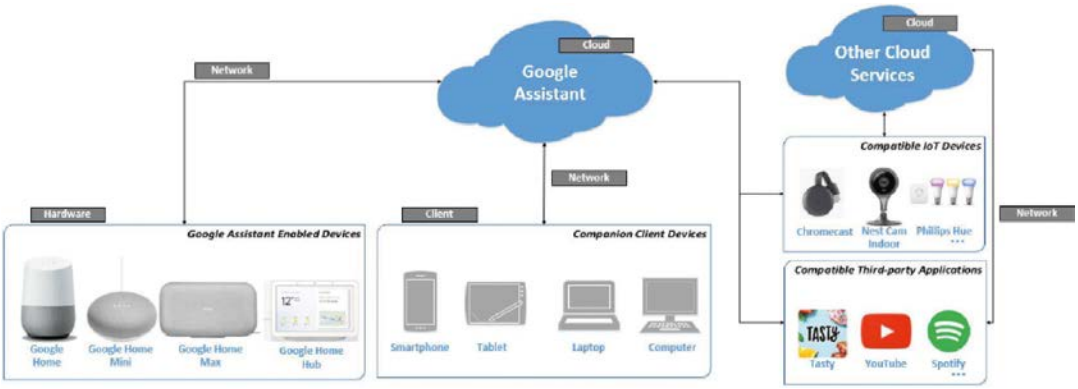

Abbildung 6: Verbindungen zwischen Google Nest und Assistant (Akinbi und Berry 2020, S. 272)

Wird ein Google Nest-Gerät mittels Aktivierungswort aktiviert, erhält der Nutzende ein audiovisuelles Signal, dargestellt durch einen Ton und das Aufleuchten von LEDs auf der Oberseite des Geräts. Auf Nest-Geräten, die über ein Display verfügen (z.B. Google Nest Hub), wird zudem ein Icon auf dem Display angezeigt. Aber auch bei Geräten anderer Hersteller, auf denen der Google Assistant installiert ist (z.B. Sonos One, der auch mit Amazon Alexa kompatibel ist), wird ein audio-visuelles Feedback ausgegeben (Langley und Pattison Tuohy 2019). Somit ist stets ersichtlich, ob das Gerät aktiviert ist. Obwohl es Forschenden gelungen ist, mittels bösartigen Programmcodes, der über «Skills» eingespielt wird, auch nach einem Stopp-Wort weiter ein Mithören der Lautsprecher zu ermöglichen (Beuth 2019), ist unbekannt, ob es auch möglich ist, das visuelle Feedback auszuschalten.

Standardmässig ist der Google Nest im Stand-by-Modus, wartet also auf Spracheingaben, die mit dem Aktivierungswort «Ok Google» beginnen. In der Datenschutzerklärung schreibt Google: «In diesem Modus werden kurze Audio-Snippets bearbeitet (einige Sekunden lang), um eine Aktivierung (wie ‹Ok Google›) zu erkennen. Wird nichts erkannt, werden diese Audio-Snippets weder an Google gesendet noch gespeichert. […] Im Stand-by-Modus sendet Assistant das, was Sie sagen, weder an Google noch an andere Anbieter oder Personen.» (Google Inc. 2021e).

Wird das Aktivierungswort erkannt, wird das Audio-Snippet an Google weitergeleitet. Die Snippets bestehen aus der aufgezeichneten Spracheingabe sowie allen Nebengeräuschen sowohl im Sprechzeitraum als auch jeglichen Geräuschen ein paar Sekunden vor der Ak-

tivierung (Google Inc. 2021f). Zudem wird ein Zeitstempel gespeichert. Akinbi und Berry (2020) unterzogen die Daten, die der Google Assistant bei Sprachsuchen erstellt, einer forensischen Analyse. Diese zeigte, dass folgende Daten zusätzlich sowohl auf den Geräten als auch im Onlineprofil des Nutzenden gespeichert werden:

- E-Mail-Adresse (Google Adresse) des Nutzenden
- Tonaufnahme im mp3 Format mit Zeitstempel im Dateinamen
- Transkription der Tonaufnahme als BLOB-Datei (Binary Large Objects)

Intern nutzt Google die Tonaufnahmen auch zur Verbesserung seiner Dienste. Hierfür werden diese auch an menschliche Transkribierer gesendet (Russell 2020). Bis 2019 wurden die aufgezeichneten Spracheingaben der Nutzer per Standardeinstellung auf Google-Servern oder bei machen Geräten sowie bei fehlender Internetverbindung auch lokal auf dem Gerät gespeichert (Johnson 2019; Google Inc. 2021e). Es bleibt jedoch unklar, auf welchen Servern in welchem Land solche Tonaufnahmen gespeichert werden. Eine Wahlmöglichkeit scheint es für den Nutzer dabei nicht zu geben. Auf der eigenen Webseite gibt Google die Standorte seiner Datencenter an. In Europa gibt es demnach fünf Datencenter: in Irland, Belgien, Niederlande, Dänemark und Finnland (Google Inc. 2021c). Zugriff auf die Daten habe primär Google. Einen Verkauf der Daten an Externe schliesst Google zwar aus, allerdings erfolgen Datenübertragungen an Auftragnehmer, mit denen Google zum Zwecke der Bereitstellung seiner Dienste zusammenarbeitet. Google schreibt auf seiner Webseite, dass u.U. auch Strafverfolgungsbehörden und Geheimdienste Zugriff auf die Daten erhielten (Google Inc. 2021b).[60]

Seit 2019 ist die automatische Speicherung zur Verbesserung der Dienste abgeschafft, kann aber per Opt-In durch den Nutzenden aktiviert werden. Frühere Aufnahmen sind davon jedoch nicht betroffen und werden nicht gelöscht (Google Inc. 2021f). Zudem gibt Google an, dass die Spracheingaben «aus Interaktionen mit der Google-Suche, Google Assistant […] nicht in Ihrem Google-Konto gespeichert» werden und somit nicht zur Profilgenerierung genutzt werden. Da die Bearbeitung aller Sprachbefehle auf Google-Servern stattfindet, ist davon auszugehen, dass die Sprachaufzeichnungen zumindest zur Bearbeitung auf die Server von Google geladen werden (Google Inc. 2021f). Eine Nutzung ganz ohne Google Account scheint nicht möglich zu sein. Zwar ist kein Account bei Nest nötig, jedoch muss die Google Home App installiert sein, die einen Account bei Google voraussetzt (Akinbi und Berry 2020, S. 2; Google Inc. 2021d, 2021f). Löschen und anhören lassen sich die Daten durch den Nutzer jederzeit in den Kontoeinstellungen. Zudem teilte Google mit, dass Aufnahmen, welche nicht mehr zur Produktverbesserung benötigt werden, automatisch gelöscht werden (Russell 2020). Bemerkt ein Nutzer, dass versehentlich eine Sprachaufzeichnung erfasst wurde, kann diese mittels des Befehls «Hey Google, das war nicht für dich» gelöscht werden (Google Inc. 2021e). Zudem können Nutzer die Sensitivität für die Erkennung des Aktivierungsworts einstellen (Google Inc. 2021a). Zwar lassen sich die Sprachaufzeichnungen löschen, die BLOB-Dateien mit den Bearbeitungsergebnissen allerdings nicht (Akinbi und Berry 2020, S. 9). Zudem fanden die Forscher gespeicherte Tonaufnahmen des Aktivierungsworts der Nutzer (Akinbi und Berry 2020, S. 4).

[60] US-amerikanische Behörden können Zugriff auf jedwede Daten US-amerikanischer Unternehmen erhalten, unabhängig davon, wo sich deren Server befinden. Zudem können die betroffenen Unternehmen dazu verpflichtet werden, Stillschweigen über die Datenweitergabe zu bewahren (Karaboga et al. 2014, S. 7).

Tabelle 5: Technische Daten zu Amazon-Echo-Produkten (basierend auf dem Alexa Sprachassistent)

Hersteller:	Amazon
Sitz des Herstellers:	USA
Einführung:	2014 (USA), 2016 (weltweit)
Anzahl Mikrofone:	Echo: 7 (Dempsey 2017, S. 81) Echo Dot v3: 4 (Pawlaszczyk et al. 2019, S. 22)
Konnektivität:	WLAN, Bluetooth

Amazon war der Pionier und brachte den ersten smarten Lautsprecher bereits 2014 in den USA auf den Markt (Tabelle 5). Unter dem Namen Amazon Echo gibt es unterschiedliche Modelle, z.B. den kleinen displaylosen Echo Dot, den Echo Show mit Display oder die eher unauffällige Echo Wall Clock, die einer analogen Wanduhr nachempfunden ist (Albanesius 2021). Abbildung 7 zeigt eine Übersicht einiger Modelle. Auf allen Geräten läuft der Sprachassistent Alexa. Dieser wird u.a. auch in Lautsprechern der Marke Sonos verwendet (Sonos Inc 2021).

Abbildung 7: Amazon-Echo-Modelle (Helpful Home 2020)

Die Funktionsweise der Echo-Geräte ist vergleichbar mit denen von Google. Auch hier wird die Spracherkennung entweder per Aktivierungswort oder Hardwareknopf gestartet. Die Aufnahme und Bearbeitung ist bei den Echo-Geräten an einem weiss-blau leuchtenden Lichtring an der Oberseite des Geräts zu erkennen. Bei ausgeschaltetem Mikrofon wird dieser in Rot angezeigt (AWS 2018, S. 2).

Amazon treibt das Konzept der Erweiterungen (Skills) stark voran. Externe Entwickler können Skills für unterschiedlichste Anwendungen programmieren. In deutscher Sprache gibt es bereits mehr als zehntausend Skills (Kinsella 2020). Auch einige grosse Hersteller aus anderen Bereichen entwickeln Skills. So lassen sich etwa Fahrzeuge des Herstellers BMW mittels Alexa öffnen und verschliessen (BMW AG 2021) oder Reiseinformationen der Schweizerischen Bundesbahnen abrufen (Schweiz Tipps 2021).

Hardwaretechnisch unterscheiden sich die Echo-Geräte, abgesehen von den Geräten mit Bildschirm, wenig von anderen smarten Lautsprechern. Ein grösserer Unterschied zeigt sich allerdings bei der Anzahl der Mikrofone. So waren in der ersten Version des Echo sieben Mikrofone verbaut – im Gegensatz zu den zwei Mikrofonen des Google Home Mini (Dempsey 2017, S. 81; ifixit 2019). Dadurch können auch aus weiten Entfernungen und bei hohem Hintergrundgeräuschpegel brauchbare Audioaufnahmen gemacht werden. Bei mehreren Echo-Geräten in einem Haushalt kann zudem mittels «Echo Spatial Perception» die Spracherkennung automatisch auf dem Gerät ausgeführt werden, das den Nutzer am besten «hört» (Pawlaszczyk et al. 2019, S. 21; Li et al. 2019, S. 6491).

Auch Echo-Geräte sind nicht leistungsfähig genug, Bearbeitung der Audiodaten auf dem Gerät selbst (sog. Edge Computing) durchzuführen (Pawlaszczyk et al. 2019). Für die eigentliche Spracherkennung kommunizieren die Geräte mit den «Alexa Voice Services» (AVS) in einer Cloud-Umgebung. Hierzu werden die Tonaufzeichnungen nach der Erkennung des Aktivierungsworts verschlüsselt an Server von Amazon Web Services (AWS) übertragen (Pawlaszczyk et al. 2019, S. 21).

Erhobene Daten werden an drei Orten gespeichert: auf dem Gerät selbst, auf dem verbundenen Smartphone und im Nutzerkonto bei Amazon. Diese Daten können verschiedenen Kategorien zugeordnet werden: Kontodaten, Einstellungen, weitere verbundene Geräte, Skills, Verhalten der Nutzer und Nutzeraktivitäten (Li et al. 2019, S. 6492). Alle Daten sind zudem mit einem Zeitstempel versehen, sodass sich eine sehr genaue Zeitleiste der Interaktion des Nutzers mit dem Gerät ableiten lässt (Chung et al. 2017). Konkret ermöglichen die bearbeiteten Daten eine eindeutige Identifikation, weil sich darunter das Land, in dem der Nutzer gemeldet ist, Telefonnummer, Vor- und Nachname, die E-Mail-Adresse und Bezahlinformationen befinden (Pawlaszczyk et al. 2019, S. 25–26; Edu et al. 2021, S. 7).

Auf Echo-Geräten und damit verbundenen Smartphones konnten Forscher die transkribierten Aufnahmen finden. Diese werden im JSON-Format vorgehalten. Sie beinhalten neben dem Transkript der verstandenen Sprachbefehle des Nutzers auch einen Zeitstempel der Aufnahme und die URL der zugehörigen Tonaufnahme auf den Amazon-Servern sowie eine eindeutige Identifikationsnummer des verwendeten Geräts, sodass Rückschlüsse auf den Nutzer möglich sind (Pawlaszczyk et al. 2019, S. 26). Gespeicherte Tonaufnahmen haben üblicherweise viele Nebengeräusche, sodass vermutlich auch Sprachaufnahmen von Personen gespeichert werden, die sich zum Zeitpunkt der Aufnahme in der Nähe befanden (Pawlaszczyk et al. 2019, S. 27).

Im Amazon-Account werden zusätzlich noch folgende Daten gespeichert: To-do-Listen, Einkaufslisten, Alarme, der Kundenname (anders als in Amazons eigener Datenbank hier allerdings ohne weitere Daten wie Adresse oder Telefonnummer) sowie Informationen über alle weiteren verbundenen Echo-Geräte (Pawlaszczyk et al. 2019, S. 27). Die Analyse zeigte, dass sich selbst ohne Zugriff auf den verbundenen Amazon-Account Informationen über die letzten Sprachbefehle ermitteln lassen (Pawlaszczyk et al. 2019, S. 28). Über amazon.de/alexaprivacy können Nutzende die von ihnen gespeicherten Tonaufnahmen anhören und löschen. Pawlaszczyk et al. (2019, S. 28) konnten jedoch nicht herausfinden, ob bei dieser Form der Löschung der Tonaufnahmen auch sonstige Transkripte und Meta-Informationen gelöscht werden (Pawlaszczyk et al. 2019, S. 28). Jedoch gab Amazon zu, dass bei einer

Löschung nur die Rohdaten, jedoch nicht abgeleitete Nutzerdaten und Präferenzen des Nutzerprofils gelöscht werden (Kelly 2019).

3.3.2. Juristische Bewertung

Da es sich bei den Anbietern der hier untersuchten smarten Lautsprecher um private Firmen handelt, sind die Datenschutzvorschriften für Private anwendbar. Neben den Anbietern sind u.U. auch weitere Akteure als Verantwortliche beteiligt, namentlich die Anbieter von Erweiterungen oder sog. Skills.

Durch smarte Lautsprecher werden häufig personenbezogene Daten von mehreren Personen bearbeitet. In der Regel gibt es einen Hauptnutzer, welcher Inhaber des zur Nutzung erforderlichen Benutzerkontos ist. Hinzu kommen weitere Personen, die mit dem smarten Lautsprecher interagieren, d.h. das Aktivierungswort aussprechen und ein Kommando eingeben, sowie Personen, deren Stimmen mit aufgenommen werden, selbst wenn sie keinen Befehl eingeben, dies weil sie sich gerade im selben Raum befinden oder weil sie unwissentlich das Aktivierungswort ausgesprochen haben (EDPB 2021a).

3.3.2.1. Die Datenschutzgrundsätze bei smarten Lautsprechern

Aus dem Grundsatz von Treu und Glauben wird der Grundsatz der *Transparenz* jeglicher Datenbearbeitungsschritte abgleitet (Rosenthal 2020). Die Anforderungen an die Transparenz resp. Erkennbarkeit sind für Betreiber von smarten Lautsprechern aus mehreren Gründen besonders schwer zu erfüllen. Ein Grund ist die Vielzahl von Benutzern des Lautsprechers. Während der Inhaber eines Benutzerkontos im Zuge des Abschlusses dieses Benutzerkontos über die Datenbearbeitung informiert werden kann, ist dies für die zusätzlichen Nutzer, deren Daten bearbeitet werden, nur schwer zu erfüllen. Auch die potenzielle Mehrzahl an Verantwortlichen macht diese Aufgabe noch komplexer.

Schliesslich stellen die Besonderheiten von Sprachinterfaces ebenfalls Herausforderungen an die Transparenz (EDPB 2021a). Dabei könnte dies auch Potenzial bieten, denn das Datenschutzgesetz erfordert keine Schriftlichkeit der Information, diese könnte also auch über eine Sprachausgabe erfolgen, was bspw. für Menschen mit Sehbehinderungen Vorteile hätte.

Weitere Schwierigkeiten für die Transparenz birgt die Tatsache, dass ein smarter Lautsprecher verschiedene Funktionsstadien kennt. Er kann lokal auf die Erkennung des Aktivierungswortes warten oder mit einem entfernten Server interagieren, um einen Befehl auszuführen, er kann aber auch viele weitere Stadien einnehmen, abhängig vom Kontext (z.B. wenn Umgebungsgeräusche vorhanden sind) oder vom Benutzer, der mit ihnen spricht (z.B. je nachdem, ob es sich um einen identifizierten oder unbekannten Benutzer handelt). Diese Stadien sind für den Benutzer jedoch kaum erkennbar, weshalb der Europäische Datenschutzausschuss (EDSA) empfiehlt, dass Benutzer stets über den Status informiert werden sollten, in dem sich das Gerät gerade befindet (EDPB 2021a). Ebenfalls sollten sie informiert werden über die genauen Erhebungs- und Bearbeitungsvorgänge, namentlich welche Daten bearbeitet werden und ob diese für die Extraktion von Meta-Informationen weiter-

verwendet werden (z.B. um aus einer Sprachaufnahme Emotionen zu erkennen) (EDPB 2021a). Die Verknüpfung eines smarten Lautsprechers mit weiteren Diensten, z.B. E-Mail, Videostreaming, Musik-Playlisten oder Einkäufen, erfordert in der Regel ziemlich lange und komplexe Datenschutzerklärungen. Die Länge und Komplexität dieser Erklärungen sind der Transparenz ebenfalls nicht dienlich (EDPB 2021a).

Gemäss dem Grundsatz der *Zweckbindung* dürfen Personendaten nur für einen bestimmten und für die Betroffenen erkennbaren Zweck erhoben werden. Der Zweck muss ausdrücklich, spezifisch und rechtmässig sein (CoE 2021). Die Daten dürfen nur in einer Weise bearbeitet werden, die mit dem Zweck vereinbar ist (Rosenthal 2020). Smarte Lautsprecher bearbeiten Daten üblicherweise zu mehreren Zwecken. Der vordergründige Zweck ist die Ausführung der Benutzerbefehle. Dieser Zweck wird in der Regel auch von der Einwilligung abgedeckt sein (siehe dazu noch unten). Des Weiteren werden Daten zum Zweck des Trainings der Algorithmen, häufig auch mittels Machine Learning, verwendet. Dies geht über die blosse Ausführung von Benutzerbefehlen hinaus und muss deshalb ausdrücklich kommuniziert werden. Ein weiterer Bearbeitungszweck kann die Identifizierung des Nutzenden sein. Werden Sprachaufnahmen zur Identifizierung verwendet, liegt eine Bearbeitung biometrischer Daten vor, für welche eine ausdrückliche Einwilligung notwendig ist (siehe näher unten). Schliesslich können die erhobenen Daten auch zum Zweck der Profilbildung der Nutzenden verwendet werden, sei es, um ihnen personalisierte Inhalte auf ihre Abfragen anzubieten, oder für personalisierte Werbung (EDPB 2021a). Während personalisierte Inhalte zumeist ein intrinsisches und von den Nutzenden erwartetes Element eines smarten Lautsprechers darstellen, und somit in der Regel als von der Einwilligung erfasst angesehen werden können (sofern die Bearbeitung verhältnismässig ist), so kann die Profilbildung zu Werbezwecken nicht als Teil der Dienstleistung angesehen werden. Eine gesonderte Einwilligung ist hierfür erforderlich, welche zudem u.U. ausdrücklich erfolgen muss.

Allgemein kann also festgestellt werden, dass smarte Lautsprecher eine Vielzahl von personenbezogenen und nicht personenbezogenen Daten bearbeiten, was eine Bearbeitung für eine Vielzahl von Zwecken ermöglicht, die über die blosse Ausführung von Benutzerbefehlen hinausgehen und die auch völlig unbemerkt bleiben könnten. Durch die Analyse der gesammelten Daten ist es bspw. möglich, Interessen, Tagesabläufe oder Gewohnheiten der Nutzenden abzuleiten, was die Bearbeitung zu unvorhergesehenen Zwecken ermöglicht (z.B. Stimmungsanalyse oder Beurteilung des Gesundheitszustands), die weit über die vernünftigen Erwartungen der Benutzer hinausgehen.

Der Grundsatz der *Datenminimierung*, als Element der Verhältnismässigkeit, schreibt vor, dass Daten zu vernichten oder zu anonymisieren sind, sobald sie zum Zweck der Bearbeitung nicht mehr benötigt werden.[61] Nur die erforderlichen Informationen sollen bearbeitet werden. Smarte Lautsprecher bearbeiten und generieren eine Vielzahl personenbezogener Daten wie Sprachaufnahmen, deren Transkriptionen sowie zahlreiche Metadaten, und diese Daten können zu einer Vielzahl von Zwecken (weiter-)bearbeitet werden. Gemäss dem Grundsatz der Datenminimierung sollten smarte Lautsprecher diese Daten nicht länger speichern, als es für die jeweils zulässigen Zwecke erforderlich ist; dies bedingt u.U. unterschiedliche Datenaufbewahrungsfristen für die unterschiedlichen Bearbeitungszwecke.

[61] Art. 6 Abs. 4 DSG.

Der Grundsatz der *Datensicherheit* verlangt, dass die Vertraulichkeit, Integrität und Verfügbarkeit von Personendaten sichergestellt wird (Rosenthal 2020). Es sind strenge Sicherheitsmassnahmen sowohl auf technischem als auch auf organisatorischem Level vorzusehen, um einen Verlust oder unerlaubten Zugriff auf personenbezogene Daten zu verhindern. Diese Massnahmen müssen für alle Bearbeitungsschritte (bspw. Erfassung, Übermittlung oder Speicherung) vorgesehen werden (CoE 2021). Eine Herausforderung für die Datensicherheit ergibt sich bei smarten Lautsprechern durch die potenzielle Vielzahl von Benutzenden. In der Tat kann jede Person, die in der Umgebung steht und das Aktivierungswort ausspricht, Befehle erteilen und die Dienste des Lautsprechers nutzen. Dies birgt die Gefahr, dass durch unberechtigte Personen Sprachbefehle erteilt werden, welche auf die persönlichen Daten der anderen Benutzer zugreifen, sie verändern oder löschen (z.B. Adressen, E-Mails, Einkaufslisten, Playlisten, Kalendereinträge usw.), wobei hierfür nicht einmal eine bösartige Absicht notwendig ist, da durch Suchanfragen von Drittpersonen sich das Nutzerprofil des Hauptnutzers verändern kann.

Eine weitere Herausforderung für die Datensicherheit ergibt sich aus der Tatsache, dass die aus einem smarten Lautsprecher gewonnenen Daten die Erstellung eines ziemlich genauen Profils ihrer Nutzenden erlauben und weitgehende Einblicke in ihr Privatleben, namentlich auch in ihre Wohnung, ableiten lassen. Dies erhöht das Risiko von Überwachung, sei es zwischen Privaten (namentlich als Form häuslicher Gewalt; Stelkens 2021) oder auch durch Strafverfolgungsbehörden. Verschiedene Strafverfolgungsbehörden haben bereits ihr Interesse am Zugriff auf die durch smarte Lautsprecher erhobene Daten bekundet (Chazan 2019). Dies könnte einen *chilling effect* entfalten, der Grundrechte wie die Meinungsfreiheit untergräbt (EDPB 2021a).

3.3.2.2. Zur Rechtfertigung der Datenbearbeitung

Eine Persönlichkeitsverletzung ist widerrechtlich, wenn sie nicht durch Einwilligung des Betroffenen oder durch ein überwiegendes privates oder öffentliches Interesse oder durch Gesetz gerechtfertigt ist.[62] Im Vordergrund steht der Rechtfertigungsgrund der Einwilligung. Die Einwilligung muss für eine oder mehrere bestimmte Bearbeitungen nach angemessener Information freiwillig erteilt werden;[63] findet ein Profiling mit hohem Risiko statt oder werden besonders schützenswerte Personendaten bearbeitet, muss die Einwilligung ausdrücklich erfolgen. Die Einwilligung muss auf die bestimmten Zwecke gerichtet sein.

Wie oben ausgeführt, werden von smarten Lautsprechern unterschiedliche Personendaten genutzt. Dies beinhaltet Primärdaten, namentlich die Audioaufzeichnungen der Stimme wie auch Kontodaten (z.B. Name, Geburtsdatum, Telefonnummer, Wohnort, Zahlungsmodalitäten) oder die Suchhistorie, abgeleitete Daten aus der Nutzung des Geräts sowie dessen Verknüpfung mit weiteren Geräten (z.B. smarten Videotürklingeln) oder weiteren Diensten. Aus der Analyse dieser Daten können sich zudem weitere personenbezogene Daten er-

[62] Art. 28 Abs. 2 ZGB.

[63] Art. 6 Abs. 6 nDSG.

geben, wie Verhaltensweisen, der Gesundheitszustand, die Beziehung zwischen mehreren Personen, die mit dem Lautsprecher interagieren, etc. (EDPB 2021a).

In einigen Fällen handelt es sich um besonders schützenswerte Personendaten. Dies ist der Fall bei bestimmten Inhalten (z.B. wenn der intelligente Lautsprecher angewiesen wird, einen Arzttermin im Kalender zu vermerken) oder bei der Integration von Drittanbietern (z.B. einem Zyklustracker). Ebenfalls handelt es sich um besonders schützenswerte Personendaten, wenn Informationen über die Gesundheit aus den Stimmaufnahmen abgeleitet werden (*voice profiling*). Die Bearbeitung der menschlichen Stimme resp. der Audiodateien alleine stellt noch keine Bearbeitung besonders schützenswerter Daten dar. Wird die Stimmerkennung jedoch zum Zweck der Identifizierung eines Nutzers verwendet, handelt es sich um biometrische Daten i.S.v. Art. 5 Bst. c Ziff. 4 nDSG. Die Einwilligung zur Bearbeitung dieser Daten muss ausdrücklich erfolgen,[64] dies bedingt auch eine besonders transparente Information.

Aus der Analyse und Verknüpfung der Daten kann zudem ein Profiling resultieren. Ein Profiling ist jede Art der automatisierten Bearbeitung von Personendaten, die darin besteht, dass diese Daten verwendet werden, um bestimmte persönliche Aspekte, die sich auf eine natürliche Person beziehen, zu bewerten. Dazu gehören insb. Aspekte bezüglich Arbeitsleistung, wirtschaftlicher Lage, Gesundheit, persönlicher Vorlieben, Interessen, Zuverlässigkeit, Verhalten, Aufenthaltsort oder Ortswechsel.[65] Stellt das Profiling ein hohes Risiko dar, ist eine ausdrückliche Einwilligung erforderlich.[66] Dies ist der Fall, wenn das Profiling ein hohes Risiko für die Persönlichkeit oder die Grundrechte der betroffenen Person mit sich bringt, indem es zu einer Verknüpfung von Daten führt, die eine Beurteilung wesentlicher Aspekte der Persönlichkeit einer natürlichen Person erlaubt.[67] Ob im konkreten Fall ein Profiling mit hohem Risiko vorliegt, kann nur im Einzelfall anhand der erfassten Daten und der verwendeten Analysemethoden beurteilt werden. Jedenfalls ist aber festzustellen, dass durch die Vielzahl von erfassten Daten bei smarten Lautsprechern auch eine Beurteilung wesentlicher Aspekte der Persönlichkeit möglich ist; insofern sollte bei smarten Lautsprechern vorsichtshalber immer von hohem Risiko ausgegangen werden. Über diese Möglichkeit muss transparent und eindeutig informiert werden, anderenfalls liegt keine gültige Einwilligung vor (EDPB 2021a).

Gemäss den Datenschutzbestimmungen von Google Nest und Amazon Echo finden Datenübermittlungen in die Vereinigten Staaten statt (Google Inc. 2021e). Werden Personendaten ins Ausland übermittelt, sind die Vorschriften von Art. 16 ff. nDSG zu beachten. Sofern die betroffenen Personen nicht im Einzelfall ausdrücklich in die Bekanntgabe eingewilligt haben[68] oder ein anderer Ausnahmetatbestand[69] vorliegt, dürfen Personendaten nur ins

---

64 Art. 6 Abs. 7 lit. a nDSG.

65 Art. 5 Bst. f nDSG.

66 Art. 6 Abs. 7 lit. b nDSG.

67 Art. 5 Bst. g nDSG.

68 Art. 17 Abs. 1 lit. a nDSG.

69 Namentlich Vertragsabwicklung, überwiegende öffentliche Interessen oder Rechtsansprüche, Notfälle, allgemeine Zugänglichmachung oder gesetzlich vorgesehenes Register, Art. 17 Abs. 1 lit. b–f nDSG.

Ausland bekannt gegeben werden, wenn ein Angemessenheitsentscheid des Bundesrates vorliegt oder der Datenschutz durch völkerrechtlichen Vertrag, vertragliche Datenschutzklauseln, Standarddatenschutzklauseln oder unternehmensinterne Datenschutzvorschriften gewährleistet wird.[70] Dabei ist zu beachten, dass das EU-US Privacy Shield mit Urteil des EuGH vom 16. Juli 2020 für ungültig erklärt wurde[71] und dass vor diesem Hintergrund der Schweizerische Datenschutz- und Öffentlichkeitsbeauftragte zur Auffassung gelangt ist, dass auch der Privacy Shield CH-US kein adäquates Datenschutzniveau bietet (EDÖB 2020a). Folglich ist derzeit eine Datenübermittlung in die USA nur noch gestützt auf einen Vertrag, Standarddatenschutzklauseln oder auf eine ausdrückliche Einwilligung der Betroffenen zulässig.

Weitere Pflichten der Bearbeiter umfassen die datenschutzfreundliche Ausgestaltung der Datenbearbeitung («privacy by design and by default»), das Führen eines Verzeichnisses der Bearbeitungstätigkeiten, die Durchführung einer Datenschutzfolgenabschätzung sowie die Meldung von Verletzungen der Datensicherheit. Namentlich auf die Verpflichtung zur datenschutzfreundlichen Ausgestaltung der Datenbearbeitung soll hier kurz eingegangen werden. Standardmässig sollten Dienste eines smarten Lautsprechers, die keinen identifizierten Benutzer benötigen, keine Benutzer mit den Befehlen verknüpfen. Eine datenschutzfreundliche Ausgestaltung würde die Daten der Benutzer nur zur Ausführung der Benutzeranfragen bearbeiten und weder die aufgezeichneten Sprachdaten noch die ausgeführten Befehle speichern (EDPB 2021a).

Die Anbieter von smarten Lautsprechern müssen zudem die Rechte der betroffenen Personen garantieren. Dazu gehört die transparente Information, das Auskunftsrecht, die Rechte auf Berichtigung, Einschränkung der Bearbeitung sowie auf Löschung. Wie erwähnt, können mehrere Personen zu den Betroffenen der Datenbearbeitung eines smarten Lautsprechers gehören, was die Information sowie die Ausübung der Betroffenenrechte erschweren kann. Der EDSA empfiehlt, dass die Betroffenenrechte direkt über das Sprachinterface ausgeübt werden können, unabhängig davon, ob eine betroffene Person als Nutzer registriert ist oder nicht (EDPB 2021a).

Abschliessend bleibt festzuhalten, dass die datenschutzrechtliche Beurteilung smarter Lautsprecher sehr komplex ist. Dies ergibt sich in erster Linie aus der Vielzahl von möglichen Verantwortlichen sowie von möglichen Nutzenden bei einem einzelnen Gerät, ebenso aus der Mehrzahl möglicher Bearbeitungszwecke. Dies macht u.a. die transparente Information der Betroffenen sowie das Einholen von Einwilligungen zusätzlich schwierig. Die Nutzungsbedingungen und Datenschutzerklärungen der hier untersuchten Lautsprecher Google Nest und Amazon Echo halten u.E. den Anforderungen an Transparenz und Verständlichkeit nicht stand, zudem wird, soweit ersichtlich, keine Einwilligung von Drittpersonen eingeholt, welche den Lautsprecher neben dem registrierten Benutzer bedienen. Amazon weist auf seiner Webseite darauf hin, dass mit der Nutzung oder dem Zugriff auf Alexa insgesamt

---

[70] Art. 16 Abs. 1 und 2 nDSG.

[71] EuGH, Rs. C-311/18 – *Facebook Ireland und Schrems*.

sechs Nutzungsbedingungen zugestimmt wird.[72] Die ausdrückliche Einwilligung für die Bearbeitung besonders schützenswerter Personendaten sowie bei einem Profiling mit hohem Risiko ist damit nicht erfüllt.

3.3.3. Gesellschaftliche und ethische Herausforderungen

Kritische Diskurse rund um smarte Lautsprecher oder Sprachassistenten drehen sich überwiegend um die mit deren Nutzung einhergehenden datenschutzrechtlichen Herausforderungen. Nutzer smarter Lautsprecher äussern grundsätzlich sieben Arten von Datenschutzsorgen: (1) Hackerangriffe auf das Gerät, (2) das Sammeln von personenbezogenen Daten, (3) dauerhaftes Mithören, (4) die Aufzeichnung von privaten Gesprächen, (5) das grundsätzliche Nichtrespektieren der Privatsphäre, (6) Datenspeicherung und (7) «gruseliges» Verhalten der Geräte (Manikonda et al. 2018, S. 229). Die grössten Privatheitsbedenken lösen verbaute Kameras aus. Mikrofone folgen auf Platz zwei (Lau et al. 2018, S. 3). Im Folgenden werden einige darüber hinausgehende gesellschaftliche und ethische Herausforderungen diskutiert

3.3.3.1. Verantwortlichkeit und Rechenschaftspflicht

Bei Google ist unklar, wie und in welcher Form die Daten auf den Servern der Anbieter gespeichert werden – lediglich, dass die Server in der EU stehen. Amazon geht in einem Dokument etwas detaillierter auf die Speicherung ein und schreibt, dass die Daten in «verschiedenen Formen und für verschiedene Zwecke in verschiedenen Amazon Services, wie S3 und DynamoDB» (übersetzt) gespeichert werden (AWS 2018, S. 7–8). Weiter heisst es in dem Dokument, dass «verschiedene Typen von Daten», wie z.B. Konfigurationsdaten, welche auch den Standort, Zeitzone und gewählte Masseinheiten (z.B. Kilogramm oder Pfund) gespeichert werden. Diese Daten ermöglichen in gewissem Masse eine Zuordnung der Nutzer (AWS 2018, S. 6–7).

Weiter schreibt AWS (2018, S. 7), dass «sensible Kundendaten im Alexa System (wie Sprachaufzeichnungen) verschlüsselt in Datenbanken gespeichert und verschlüsselt übertragen» werden. Weiter werden Log-Daten für den Fall einer nötigen Fehlerbehebung und Sicherheits-Log-Daten erstellt und verschlüsselt gespeichert. Auch zu diesen haben nur die Personen Zugriff, welche diese zur Erfüllung der Aufgabe benötigen (AWS 2018, S. 7). Daten, die von Drittanbietern (Skill-Anbietern) erhoben werden, werden nicht von Amazon, sondern von den Drittanbietern selbst nach deren Datenschutzbedingungen erhoben und gespeichert (AWS 2018, S. 7).

Beim Thema Zugriffsverantwortlichkeit und welche internen Abteilungen wie Zugriff zu Daten erhalten, lassen sich zu Google keine Informationen finden. Nach Angaben von Amazon haben nur berechtigte Mitarbeitende Zugriff auf die Daten zur Erfüllung ihrer jeweiligen

[72] Es handelt sich dabei um die Alexa Nutzungsbedingungen, die Amazon.de Allgemeinen Geschäftsbedingungen, die Amazon Prime-Teilnahmebedingungen, die Nutzungsbedingungen für Amazon Music, die Lizenzvereinbarung und Nutzungsbedingungen für Amazon.de Kindle sowie die Audible.de Allgemeine Geschäftsbedingungen.

Aufgabe. Die Berechtigungen würden von Managern erteilt und die erfolgten Zugriffe von diesen überwacht. Zudem würden die Berechtigungen mindestens quartalsweise geprüft und ggf. entfernt (AWS 2018, S. 7).

Eine (Datenschutz-)Folgenabschätzung hat bislang keiner der beiden Anbieter durchgeführt. Daher fordert der EDSA die Durchführung einer Folgenabschätzung für smarte Lautsprecher und digitale Assistenten (EDPB 2021b, S. 32).

Sowohl bei Amazon als auch bei Google können Audioaufnahmen teilweise gelöscht werden. Die bei der Interaktion mit dem smarten Lautsprecher entstandenen Aufnahmen können durch die Nutzenden sowohl manuell gelöscht werden, als auch kann eine Frist zur automatisierten Löschung festgelegt werden (Google Inc. 2022a; Thornsby 2020). Aufnahmen, die durch Google für andere Zwecke (Voice Match oder die Verbesserung der Spracherkennungssysteme) verwendet werden, können jedoch trotz o.g. Löschmöglichkeit für einen nicht näher spezifizierten Zeitraum weiterhin gespeichert bleiben (Google Inc. 2022a). Der zwischenzeitlich bekannt gewordenen Transkription der Sprachaufnahmen durch Amazon-Mitarbeitende, die nach Amazon-Angaben der Verbesserung der Spracherkennung dienten, kann mittlerweile widersprochen bzw. eine Löschung dieser Daten erwirkt werden (Thornsby 2020).

Um dem möglichen Missbrauch personenbezogener Daten durch deren Hochladen ins Internet bzw. die Hersteller-Server zu begegnen, ist in den vergangenen Jahren eine Diskussion rund um das edge computing entstanden. Dabei findet die Bearbeitung der Daten bzw. Sprachbefehle auf dem Gerät selbst statt (Edu et al. 2021, S. 25). Bislang sind konkrete Schritte allerdings ausgeblieben. Denn eine Reihe von Gründen hält die Hersteller davon ab, grössere Teile ihrer Programmlogik und Algorithmen direkt in die Geräte zu implementieren. Zum einen können so mehr Sicherheitslücken direkt im Produkt entstehen, welche angegriffen werden können (Martin 2021, S. 92). Zum anderen können Unbefugte mittels sog. Reverse Engineerings, also der Untersuchung und Analyse von bestehenden Produkten (Khan 2020), Einblicke in die Programmlogik erhalten. Dies führt u.a. dazu, dass bestehende Sicherheitslücken in den Geräten sichtbar werden (Huang et al. 2020) und die Geräte teurer werden (Nguyen et al. 2021).

3.3.3.2. Sicherheit

Die Sicherheit der Lautsprecher wird nicht von unabhängigen Stellen evaluiert. Jedoch gibt es eine grosse Gemeinschaft von «White Hat Hackern»[73], die bereits regelmässig Schwachstellen gefunden und gemeldet haben (Brown 2019). Um die Sicherheit der Geräte stets möglichst gut zu gewährleisten, bieten die Hersteller regelmässig Updates und Patches an, die meist auch ohne das Zutun des Nutzenden installiert werden. Dennoch wird auf mehrere Sicherheitsgefahren hingewiesen (Bugeja 2021): etwa inwiefern die informationstechnische Resilienz der Geräte gegeben ist, ob Sabotage oder Missbrauch durch Mitarbeitende verhindert werden kann oder ob bspw. mittels Logging der Zugriff von Mitarbeitenden auf Kundendaten wirksam aufgedeckt werden kann (Zurechenbarkeit).

[73] Ethische Hacker mit guten Absichten, die Sicherheitsschwachstellen aufzeigen.

Alle Informationen, die nach Erkennung des Aktivierungsworts anfallen, werden mittels TLS 1.2 Verschlüsselung an einen Server von Amazon gesendet und dort durch den Alexa Voice Service bearbeitet (AWS 2018, S. 2). Dies entspricht üblichen IT-Sicherheitsstandards.

3.3.3.3. Transparenz und Erklärbarkeit der Softwareergebnisse

Amazon bietet auf der eigenen Wissenschafts-Webseite ausführliche technische Informationen darüber, wie die Erkennung funktioniert und wie mit den erfassten Daten umgegangen wird. Aussagen darüber, wie Entscheidungen zustande kommen, gibt es jedoch nicht. Diesbezügliche Aussagen sind eher oberflächlich. Bspw. wird erwähnt, dass bei dem Sprachbefehl «Play ‹Radioactive› by Magic Dragons» die Antwort «Playing ‹Radioactive› by Imagine Dragons» genauso wahrscheinlich sei. Was verstanden werde, hänge auch von weiteren Daten aus dem Profil des Nutzers ab, die ebenfalls hinzugezogen würden (Sarikaya 2019).

Grundsätzlich versuche Amazon laut eigenen Aussagen die Beteiligung von Menschen bei der Korrektur von falsch ausgeführten Befehlen so gering wie möglich zu halten (Amazon Science 2020; Sarikaya 2019). Eine Erklärung für die Auswahl eines Skills wird u.a. in Kim et al. (2018) beschrieben. Jedoch fehlen wissenschaftliche Untersuchungen dazu, wie es dann innerhalb des Skills zu einer Entscheidung und somit zu einer Reaktion von Alexa kommt.

Über die Anzahl und Qualität der Trainingsdaten schweigt sich Amazon aus. Jedoch wird erwähnt, dass mittels «aktivem Lernen»[74] bis zu 97 % weniger Daten zum Trainieren eines Algorithmus benötigt würden (Amazon 2018). Zur Generierung von Trainingsdaten wird zudem erwähnt, dass diese aus Audioaufzeichnungen erstellt würden. Aus dieser «Vorlage» könne durch Paraphrasieren ein beliebig umfangreicher Satz von Trainingspaaren generiert werden (z.B. durch verschiedenen Satzstrukturen oder Negativbeispiele) (Hardesty 2019).

Auch Google bietet wissenschaftliche Arbeiten zu seinen Assistenten an. Allerdings finden sich dort keine Informationen zu verwendeten Trainingsdaten oder der Erklärbarkeit von Entscheidungen.

3.3.3.4. Transparenz der Anwendung

Die seitens Amazon veröffentlichen Berichte zu Alexa richten sich eher an ein technisches Fachpublikum und dienen der Erläuterung (neuer) technischer Funktionsweisen (Amazon Science 2020). Zudem arbeiten beide Unternehmen daran, mittels gestaffelter Datenschutzerklärungen eine bessere Information der Betroffenen über wesentliche Aspekte zu erreichen, sodass sich deren Lesbarkeit besser als ein Grossteil der Konkurrenz schlägt (Litman-Navarro 2019). Generell stehen Datenschutzerklärungen jedoch weiterhin in der Kritik, unleserlich zu sein und von den Betroffenen letztlich nicht gelesen zu werden (Rossnagel et al. 2020, S. 20–23).

[74] Ein Sonderfall des «maschinellen Lernens», bei welchem ein Algorithmus einen Nutzer aktiv fragt, welche Datenpunkte in einer Ausgabe korrekt und erwünscht waren.

3.3.3.5. Gerechtigkeit, Fairness und Nicht-Diskriminierung

Smarte Lautsprecher werden damit beworben, den Nutzenden Erleichterung im Alltag zu verschaffen, indem sie die Erledigung einer wachsenden Anzahl an Aufgaben mittels Sprachbefehlen ermöglichen und so Tätigkeiten vereinfachen sollen. Zum Beispiel lässt sich Musik per Sprachbefehl binnen weniger Sekunden starten, sodass das manuelle Anschalten der Musikanlage und die händische Eingabe der gesuchten Musik entfallen können. Dem steht die Kritik gegenüber, die grossen Digitalkonzerne würden aus der Bearbeitung von Daten einen (Macht-)Gewinn erzielen, der in keinem Verhältnis zu den Vorteilen für die Nutzenden stehe. Demnach dienten Produkte wie smarte Lautsprecher der (heimlichen) Sammlung personenbezogener und anonymisierter Daten, die zur Extraktion von Verhaltensdaten aus allen menschlichen Erfahrungen verwendet werden, ohne dass sich die Betroffenen darüber im Klaren seien, welche Aussagekraft ihre Alltagsdaten hätten. Diese Verhaltensdaten würden dann mittels KI-Einsatzes in Verhaltensprognosen umgewandelt und an Unternehmen weiterverkauft, um das Verhalten von Nutzerinnen und Nutzern zu analysieren und im Sinne der Unternehmen zu manipulieren (Zuboff 2018; Mühlhoff 2021).

Noch ist es Werbetreibenden nicht möglich, gegen Gebühr Werbung auf smarten Lautsprechern zu schalten. Dies liegt womöglich daran, dass Nutzerbefragungen ergeben hatten, dass ein Grossteil der Nutzenden sich gegen das Abspielen von Werbung auf smarten Lautsprechern ausspricht (Meyers 2019). Sollte die Schaltung von Werbung auf smarten Lautsprechern zukünftig möglich werden, würde sich allerdings die Möglichkeit der Schaltung von Werbung ergeben, die auf die personenspezifischen demografischen, religiösen oder politischen Aspekte des Publikums abgestimmt wäre. Besondere Brisanz könnte das Aussenden politischer Werbung (political micro-targeting) über smarte Lautsprecher beinhalten, weil darüber eine Beeinflussung der politischen Meinung der Bevölkerung möglich werden könnte. Unabhängig von der Aussendung der Werbung über smarte Lautsprecher fliessen die über die Geräte gesammelten personenbezogenen Daten allerdings ohnehin in das entsprechende Google-, Amazon- usw. -Profil des Nutzers ein. Insofern handeln die Anbieter bereits heute mit den Daten der Nutzenden, die auch zu Werbezwecken verwendet werden können (EDPS 2018).

Des Weiteren stehen smarte Lautsprecher insofern in der Kritik, als sie auf moralisch fragwürdige Fragen keine adäquaten Antworten lieferten. Eine weitere in diesem Zusammenhang diskutierte Konsequenz ist das sog. «**social-deskilling**», die Gefahr, dass Menschen die Fähigkeit verlieren, selbst Entscheidungen treffen zu können und zu wollen (Vallor 2015).

3.3.3.6. Bias-Vermeidung

Da keine Informationen über die Trainingsdaten vorhanden sind, kann auch keine Aussage über mögliche Verzerrungen (Bias) getroffen werden. Jedoch zeigen Forschende immer wieder, dass auch Spracherkennungssysteme stark von Verzerrungen beeinflusst sind (Chambers 2020).

3.3.4. Zwischenfazit

Wie die Ausführungen der vergangenen Unterabschnitte gezeigt haben, fokussieren die Diskussionen zu smarten Lautsprechern v.a. auf Datenschutz- und IT-Sicherheitsaspekte. Grundsätzlich gilt die Spracherkennung smarter Lautsprecher als technisch weitgehend zuverlässig. Allerdings ermöglicht die fortschrittliche Erkennungsleistung zugleich die Aufzeichnung und Analyse der Stimme und Sprache auch von Personen, die den smarten Lautsprecher nicht bewusst nutzen. Dadurch, dass smarte Lautsprecher zumeist in Privaträumen aufgestellt werden, können bei der bewussten oder unbewussten Nutzung Daten anfallen, die Rückschlüsse über Interessen, Tagesabläufe oder Gewohnheiten und damit Informationen über wesentliche Aspekte der Persönlichkeit der in einem Haushalt lebenden Menschen erlauben. Ob und inwiefern derartige Daten bzw. Analysen zu anderen geschäftlichen Zwecken genutzt werden, ist nicht bekannt. Die öffentlich-mediale Debatte über smarte Lautsprecher fokussierte insb. auf das unbeabsichtigte Aktivieren der Geräte und die Nicht-Löschung von Sprachaufzeichnungen. Sprachaufzeichnungen werden mittlerweile nur mit der Einwilligung der Betroffenen erstellt und auch an der Beseitigung der unbeabsichtigten Aktivierung wird gearbeitet.

3.4. Gesichts- und Spracherkennung im öffentlichen Raum seitens polizeilicher Stellen

Die Gesichtserkennung seitens polizeilicher Stellen kann zu unterschiedlichen Zwecken eingesetzt werden. Schaut man sich etwa die Einsatzzwecke an, die in einer einschlägigen Broschüre der US-amerikanischen Security Industry Association (SIA) – unter deren Dach zahlreiche Anbieter von Gesichtserkennungstechnologien versammelt sind – benannt werden, wird sie insb. für die folgenden Zwecke eingesetzt (Security Industry Association 2021):

- Grenzkontrollen (auf Flughäfen wie an See- und Landgrenzen)
- Ermittlungsverfahren in versch. Bereichen, insb. zu den Zwecken:
  - Verhinderung von Menschenhandel,
  - Auffinden von vermissten Kindern,
  - Auffinden von gesuchten Kriminellen oder Verdächtigen jeglicher Art, insb. in den Fällen Mord, Vergewaltigung, Terrorismus, Identitätsdiebstahl.

Detaillierte Informationen über die Einsatzzwecke liegen jedoch nicht vor, da die polizeiliche Nutzung von Gesichtserkennung selbst in Ländern, wie den Vereinigten Staaten, die auf langjährige Erfahrungen mit der Technologie zurückblicken können, häufig im Verborgenen stattfindet. Dass bspw. landesweit mehr als 2400 US-amerikanische Strafverfolgungsbehörden die umstrittene Software Clearview AI nutzten, wurde erst im Rahmen eines Interviews mit dem CEO von Clearview bekannt (Lopatto 2020).

Das zentrale Argument der Befürworter der Gesichtserkennung durch Strafverfolgungsbehörden bezieht sich auf deren höhere Effektivität im Vergleich zu den klassischen Methoden, die seitens Strafverfolgungsbehörden eingesetzt werden (McLaughlin und Castro 2020).

Unter Bezugnahme auf den Face Recognition Vendor Test des NIST wird etwa darauf verwiesen, dass die besten Gesichtserkennungsalgorithmen inzwischen der Gesichtserkennung durch Menschen überlegen seien. Heute noch vorhandene Mängel bei der Erkennung unterschiedlicher Ethnien würden mit der laufenden Verbesserung der Algorithmen schon bald der Vergangenheit angehören. Damit zusammenhängend ist die Steigerung der Effizienz klassischer Systeme ein weiteres wichtiges Argument. Selbst wenn eine Maschine ein Gesicht nur gleich gut wie ein Mensch erkennen kann, seien Maschinen deutlich schneller und trotz der hohen Erstanschaffungskosten wirtschaftlicher (McLaughlin und Castro 2020). Besonders populär ist dieses Argument bspw. im Kontext der Einführung automatisierter Passkontrollsysteme. Daher wächst der Umsatz und der Markt mit Gesichtserkennungstechnologie. Während 2020 noch 4,2 Milliarden Franken erzielt wurden, gehen Schätzungen von 9,4 Milliarden für das Jahr 2025 aus (Luchetta 2021b).

3.4.1. Geschichte der polizeilichen Gesichtserkennung

Im Zentrum der folgenden Ausführungen steht die Gesichtserkennung seitens Strafverfolgungsbehörden im öffentlichen Raum. Dazu werden zunächst die Anfänge der polizeilichen Gesichtserkennung weltweit und im Anschluss die Entwicklungen in der Schweiz beschrieben.

3.4.1.1. Die Anfänge der polizeilichen Gesichtserkennung

Mit Fortschritten im Bereich der Gesichtserkennungstechnologie kamen im Laufe der 1990er-Jahre erste Anbieter mit entsprechenden Systemen auf den Markt und insb. Sicherheitsbehörden wurden zunehmend auf die Technologie aufmerksam (Wright 1998, S. 9). Als mögliche Einsatzgebiete galten Stadtviertel mit erhöhtem Kriminalitätsniveau, Flughäfen, aber auch Sportstadien, Banken, Casinos und Führerscheinstellen – meist mit dem Ziel der automatisierten Identifikation von Straftätern, aber bspw. auch dafür, Identitätsdiebstahl mittels Abgleich des Führerscheins mit Personalausweisdaten festzustellen (Meyer 2020, S. 62). Den ersten Pilot-Testlauf eines 1-zu-N-Systems markierte der Einsatz des Systems «Mandrake», das ab November 1998 von der britischen Polizei im östlichen Londoner Stadtbezirk Newham eingesetzt wurde (Omega Foundation 2000, S. lviii). Die anfangs beworbene Erkennungsrate von 80 % konnte von der eingesetzten Software FaceIT des US-amerikanischen Anbieters Visionics nicht ansatzweise erreicht werden. Pressemeldungen zitierten Polizisten dahin gehend, dass das System allenfalls durch die allgegenwärtige Präsenz der Kameras abschreckend auf potenzielle Kriminelle wirken und dadurch womöglich zu einer Senkung der Kriminalitätsrate führen könne, die Gesichtserkennung jedoch nutzlos sei und regelmässig zu Falschalarmen geführt habe (Meek 2002). Ebenso problematisch wie die Falscherkennungsraten des Systems war die öffentliche Werbekampagne rund um die Technologie: Ohne nachprüfbare Evidenz war die um 34 % gesunkene Kriminalitätsrate dem Videoüberwachungssystem zugerechnet worden, woraufhin solche Systeme auch in anderen Städten (z.B. Birmingham) eingeführt wurden (Business Wire 2001). Eine Informationsfreiheitsanfrage der American Civil Liberties Union (ACLU) enthüllte, dass die Software in Tampa (Florida) weder zu einer Verhaftung geführt hatte noch einen einzigen korrekten

Treffer gelandet hatte. Zudem wurde bekannt, dass das System regelmässig und an einigen Tagen sogar ausschliesslich, falsch-positive Meldungen ausgegeben hatte (Stanley und Steinhardt 2002, 3 f.). Der für das Programm verantwortliche Polizeidetektiv Bill Todd hatte die Wahl des FaceIT-Systems zuvor mit dessen Erfolg beim FRVT 2000 begründet, bei dem die Software besser abgeschnitten hatte als andere kommerziell erhältliche Systeme (Gates 2011). Schliesslich versuchte die Polizei – obwohl die Einführung in Tampa öffentlichkeitswirksam zelebriert worden war –, über das Scheitern und die darauffolgende Abschaltung des Programms im August 2003 Stillschweigen zu bewahren (Gates 2011). Auch ein grossangelegter Testlauf an verschiedenen US-Flughäfen im Jahr 2003 bestätigte den Misserfolg der Gesichtserkennung. Am Hauptschauplatz des Tests, dem Logan Airport in Boston, konnte die eingesetzte Software innerhalb der Testdauer von zwei Jahren nicht einen erfolgreichen Treffer verbuchen (Willing 2003).

Ähnliche Ergebnisse lieferte auch der im Jahr 2006 gestartete Test des deutschen Bundeskriminalamts (BKA) am Mainzer Hauptbahnhof. Das von den Firmen L1-ID/Bosch Sicherheitssysteme, Cognitec und Crossmatch/Vitronic Dr. Stein entwickelte System lieferte während der viermonatigen Testphase statt der anvisierten Trefferrate von 80 % durchschnittlich lediglich eine Trefferrate von 30 %, woraufhin das BKA das Scheitern einräumte und die Beendigung des Projekts bekannt gab. Als Hauptgründe für die mangelhafte Erkennung wurden wechselhafte Lichtverhältnisse und die Schwierigkeit bei der Identifizierung von Gesichtern genannt, die nicht frontal in Richtung der Kameras blickten (Schulzki-Haddouti 2007; BKA 2007).

Deutlich erfolgreicher war derweil die Verifikation (1-zu-1-Abgleich). Polizeistellen testeten die Gesichtserkennung zur Bestätigung der Identität Festgenommener oder Verstorbener seit der Jahrtausendwende mit Erfolg. Bald wurden auch Tests an Flughäfen in unterschiedlichen Ländern begonnen und lieferten ebenfalls erste positive Resultate (NZZ 2002).

3.4.1.2. Aktueller Stand der Gesichtserkennung durch öffentliche Stellen weltweit

Deutliche Fortschritte bei der Gesichtserkennung wurden hingegen in den letzten Jahren gemacht. Zahlreiche Staaten weltweit setzen inzwischen regelmässig Gesichtserkennungstechnologien zu Ermittlungszwecken, immer mehr Staaten aber auch zum Zwecke der Echtzeiterkennung verdächtiger oder gesuchter Personen. Eine Studie der Carnegie Endowment for International Peace zur weltweiten Nutzung von KI-basierten Überwachungstechnologien aus dem Jahr 2019 kam zu dem Ergebnis, dass weltweit inzwischen 64 von 176 untersuchten Staaten, darunter auch die Hälfte aller freiheitlich-liberalen Demokratien, Gesichtserkennungstechnologie für Überwachungszwecke einsetzen (Feldstein 2019). Im Folgenden sollen ein paar herausstechende Anwendungen der Gesichtserkennung seitens staatlicher Behörden in anderen Staaten kurz skizziert werden, um das Einsatzspektrum der Technologie besser überblicken zu können.

3.4.1.3. Gesichtserkennung durch öffentliche Stellen in Russland

Von der russischen Polizei in Moskau wird Echtzeitgesichtserkennung seit 2017 eingesetzt, nachdem der Einsatz zuvor mehrere Jahre lang vorbereitet worden war (Vincent 2017). Zunächst wurde die Software zum Auffinden vermisster Menschen und gesuchter Krimineller verwendet. Dazu greift die Polizei auf die Aufnahmen von mehr als 13 Millionen Kameras zu, die über das ganze Land verstreut und in zunehmendem Masse per Internetanbindung miteinander vernetzt sind, davon knapp 200.000 Überwachungskameras allein in Moskau. Auf 1000 russische Einwohner kommen so 93 Überwachungskameras, womit das Land – nach den USA und China – auf Platz 3 der Länder mit den meisten Kameras pro Einwohner liegt. 59 % aller Kameras werden von Unternehmen, etwa zur Überwachung der Eingänge von Apartment-Komplexen, betrieben, 33 % von staatlichen Stellen und 9 % von Privatpersonen (Foltynova 2021). Dadurch, dass die seitens kommerzieller Stellen installierten Kameras 95 % der Moskauer Hauseingänge, aber auch alle Metrostationen und sonstigen öffentlichen Plätze abdecken, ist eine flächendeckende Überwachung der Moskauer Bevölkerung möglich. Dass die Gesichtserkennung in der Praxis noch nicht flächendeckend eingesetzt wird, liegt eher an der durchschnittlichen Bildauflösung und den Betriebskosten der Software (Vincent 2017). Die verwendete Software FindFace Security stammt vom russischen Anbieter NtechLab (Vincent 2020). Durch die flächendeckende Überwachung der Hauseingänge war es dann nach Beginn der Coronapandemie auch möglich, die Einhaltung der Pandemieregeln zu überwachen. Neben der Erkennung unerlaubter Menschenansammlungen konnten so auch Menschen identifiziert werden, die gegen ihre Quarantäneauflagen verstiessen (futurezone 2020). Schon 2017 war zudem bekannt geworden, dass die Gesichtserkennung gegen die politische Opposition verwendet wird, um Teilnehmer von Demonstrationen zu identifizieren (Rebiger 2017). Anfang 2021 wurde dann bekannt, dass die Gesichtserkennung auch dazu eingesetzt wird, «verdächtige Personen» von der Teilnahme von Kundgebungen abzuhalten, indem sie während der Anfahrt in Polizeigewahrsam genommen wurden (Stolyarov und Tétrault-Farber 2021).

3.4.1.4. Gesichtserkennung durch öffentliche Stellen in China

Im Gegensatz zu den USA oder UK wurde der Grossteil der Überwachungskameras in China erst in den letzten Jahren installiert (Gan 2020). Je nach Quelle schwankt die Anzahl der in China installierten Überwachungskameras zwischen 200 Millionen (Foltynova 2021) und weit über 300 (Gan 2020) Millionen, womit das Land in der Liste der meisten Kameras pro Land führend ist und in der Liste der Kameras pro Einwohner je Quelle den zweiten oder dritten Platz belegt. Auf 1000 chinesische Einwohner fallen so knapp 144 Überwachungskameras. Zudem belegen chinesische Städte in Bezug auf die Dichte an Überwachungskameras die ersten fünf Plätze, erst dann gefolgt von London, das über die höchste Kameradichte in westlichen Ländern verfügt (Baltrusaitis 2019).

Gesichtserkennung wird in China im Kontext einer umfassenden KI-basierten Digitalisierungs- und Überwachungsstrategie eingesetzt und soll künftig die flächendeckende Überwachung und Kontrolle der chinesischen Bevölkerung ermöglichen (Hoffman 2017). Dabei stellt die Nutzung der Gesichtserkennung nur einen Aspekt der chinesischen Strategie dar, Daten aus möglichst vielen Quellen zusammenzuführen, um den Kontrollgrad zu maximie-

ren. Ein zentrales Element dieser Bemühungen stellt das chinesische Sozialkreditsystem dar: Ausgehend von ursprünglichen Überlegungen, die bis in die 1990er-Jahre zurückreichen und basierend auf einem Vorschlag des Staatsrats im Jahr 2007 zur Verbesserung der ökonomischen Vertrauenswürdigkeit von Unternehmen und Bürgern, wurden die Ziele des Systems seit dessen Start im Jahr 2014 deutlich erweitert. Das Verhalten von Bürgern und Unternehmen soll nicht mehr nur im Hinblick auf deren ökonomische Kreditwürdigkeit, sondern in Bezug auf rechtliche, moralische und berufsethische Elemente bewertet werden. In dieser zu schaffenden neuen Datenbank sollen alle über eine Person verfügbaren Daten aus möglichst vielen Bereichen ihres Lebens zusammengeführt und bewertet werden, um die «gesellschaftliche Vertrauenswürdigkeit» dieser Person mittels eines Scores zu ermitteln und insb. für staatliche Stellen und Unternehmen transparent zu machen. Indem Personen, die über einen positiven Score verfügen, belohnt und solche mit einem niedrigen Score bestraft werden, soll die Bevölkerung zu ethischerem Handeln motiviert werden (Liang et al. 2018). Neben Kritik an der chinesischen Regierung und dem Verüben von Straftaten listet das System eine Vielzahl von individuellen Handlungsweisen in der Kategorie unerwünschter Handlungen, etwa auch das seltene Besuchen der eigenen Eltern und die unzureichende Gewährleistung ihrer Versorgung (Dpa 2018).

Mittels Überwachungskameras, deren Daten in Echtzeit mittels Gesichtserkennungstechnologien, aber auch Objekt- und Bewegungserkennungstechnologien analysiert werden, wird die Bevölkerung grossflächig überwacht. Verstösse gegen moralische und rechtliche Gebote in allen Bereichen, die mittels Gesichtserkennung überwacht werden, etwa das Ignorieren einer roten Ampel, führen zu Konsequenzen am persönlichen Score (Simonite 2018). Chinesische Polizisten testen seit 2018 Datenbrillen mit integrierter Gesichtserkennung (Chin 2018). Ende 2020 wurde zudem bekannt, dass die Technologieunternehmen Huawei und Megvii an der Nutzung von Gesichtserkennung zur Bestimmung der ethnischen Zugehörigkeit geforscht haben, um mittels eines sog. Uiguren-Alarms Behörden benachrichtigen zu können (Harwell und Dou 2020). Ein angeblich an der Installation von Überwachungskameras beteiligter Ingenieur gab Mitte 2021 bekannt, dass auch Emotionserkennung eingesetzt werde (Wakefield 2021).

Zu den führenden chinesischen Überwachungstechnologieanbietern zählen neben Huawei, Hikvision, Dahua und ZTE (Feldstein 2019, S. 13–17) in zunehmendem Masse auch jüngere, auf Erkennungstechnologien fokussierte Unternehmen wie SenseTime und Megvii (Mozur 2018, inside). Die umfangreichen Bemühungen zum Aufbau der Überwachungsinfrastruktur in China haben zudem längst zu einem chinesischen Export-Boom geführt: Nach Angaben der Carnegie Endowment for International Peace war China bereits im Jahr 2019 der weltgrösste Lieferant von Überwachungstechnologien und hat damit in diesem Bereich zuvor führende westliche Länder wie das Vereinigte Königreich, die Vereinigten Staaten, Deutschland und Frankreich abgelöst (Feldstein 2019, S. 14). Bereits im Jahr 2017 demonstrierten chinesische Polizeibehörden die Effektivität chinesischer Gesichtserkennungstechnologie: Die zu Demonstrationszwecken erfolgte Identifikation und Verhaftung des damaligen BBC-Korrespondenten in Peking, zu der sich dieser freiwillig bereit erklärt hatte, erfolgte in nur sieben Minuten (Russell 2017).

3.4.1.5. Gesichtserkennung durch öffentliche Stellen in den USA

In absoluten Zahlen sind in China zwar mehr Kameras installiert, bei der Überwachungskamera-Dichte pro Einwohner liegen die USA mit 153 Kameras je 1000 Einwohnern aber knapp vor China (Foltynova 2021).

Trotz des Bestehens einer Reihe von Regularien auf unterschiedlichen Ebenen gleicht die Regulierung von Gesichtserkennungstechnologien in den Vereinigten Staaten insgesamt eher einem Flickenteppich (Rashida 2021). Dies zeigte sich auf besonders eindringliche Weise im Rahmen des Clearview-Skandals. Anfang 2020 war bekannt geworden, dass ohne das Wissen der Öffentlichkeit und ohne vorherige Diskussion und Regulierungen seit 2018 mehr als 600 US-Polizeibehörden die fortschrittliche Gesichtserkennungssoftware des US-amerikanischen Unternehmens Clearview AI eingeführt hatten. Der Dienst, der sowohl per Desktop-PC als auch per App auf Mobilgeräten gesteuert werden kann, ermöglicht die Personenidentifikation mittels Gesichtserkennung. Wenn ein Treffer gefunden wird, zeigt die Software den Fundort an, der ggf. Namen und weitere Angaben beinhaltet. Besonders hervorgehoben wurde in dem Zusammenhang, dass die Software selbst dann zuverlässig arbeitet, wenn das Inputfoto nicht ideal ist, weil z.B. Teile des Gesichts verdeckt sind. Neben der Befürchtung, dass mit der Nutzung von Clearview AI auf intransparente Weise eine flächendeckende Überwachung aller Bürgerinnen und Bürger eingeführt wird, wurde insb. die gewaltige Bilddatenbank des Dienstes kritisiert. Diese bestand aus insgesamt über 3 Milliarden Bildern, die per Web-Scraping und höchstwahrscheinlich grösstenteils illegal von Plattformen wie Facebook und Youtube sowie von Millionen anderen Webseiten gesammelt worden waren (Hill 2020). Durch ein internes Datenleck bei Clearview AI wurde im Frühjahr 2020 ausserdem bekannt, dass die vorherigen Angaben des Unternehmens zu den Kunden deutlich untertrieben waren. Die Auswertung der geleakten Dokumente zeigte, dass weltweit in mindestens 26 Staaten mehr als 2200 Institutionen Clearview AI nutzen, darunter Polizeibehörden, Unternehmen und andere Institutionen wie Schulen und Universitäten. Unter den Abnehmerländern befanden sich demnach auch viele europäische Staaten, etwa Frankreich, Italien, Spanien, das Vereinigte Königreich und auch die Schweiz (Mac 2020). Obwohl der Software seitens der einsetzenden Stellen eine hohe Trefferrate attestiert wird, ist der zugrunde liegende Algorithmus nicht in das von NIST betriebene Bewertungsverfahren für Gesichtserkennungsalgorithmen eingegeben worden, sodass keine unabhängige Einschätzung möglich ist (Hill 2021).

Nachdem bereits in den Jahren vor den Enthüllungen über Clearview AI regelmässig über die Fehleranfälligkeit und mangelnde Zuverlässigkeit von Gesichtserkennungssystemen diskutiert, aber trotzdem Tausende Institutionen die Software zu nutzen begonnen hatten, entstand eine Diskussion rund um die Regulierung von Gesichtserkennung. Angesichts dessen, dass die an der Nutzung von Gesichtserkennungstechnologien interessierten Akteure begonnen hatten, Tatsachen zu schaffen, ohne sich zuvor einer öffentlichen Debatte zum Thema zu stellen, rückten die Kritiker der Gesichtserkennung insb. die Einführung von Moratorien, also zeitweiser vollständiger Verbote des Einsatzes von Gesichtserkennung durch sowohl staatliche als auch private Stellen, in den Mittelpunkt der Debatte (Fight for the Future 2021). Google, Facebook und Co. erliessen einstweilige Verfügungen und untersagten Clearview AI die Nutzung der in ihren Diensten gespeicherten Fotos. Gegenwärtig muss sich das Unternehmen in den Vereinigten Staaten vor Gericht verantworten (Hill 2021).

3.4.1.6. Videoüberwachung und Gesichtserkennung durch öffentliche Stellen in der Schweiz

Gesichtserkennung seitens öffentlicher Stellen wird in der Schweiz zum einen an Flughäfen zur Identitätsverifikation und zum anderen zur Verbrechensaufklärung verwendet. Während an Flughäfen einzig für diesen Zweck festinstallierte Kameras verwendet werden, wird bei der Nutzung zur Verbrechensaufklärung v.a. auf privat betriebene Kameras zurückgegriffen. Eine Echtzeitgesichtserkennung, wie sie etwa in China verwendet wird, findet zwar nicht statt. In einigen Polizeibehörden wird allerdings durchaus offen kommuniziert, dass der Einsatz der Echtzeitgesichtserkennung aus ihrer Sicht wünschenswert ist (Luchetta 2021b).

Ein erster Vorstoss zur Einführung der Gesichtserkennung an Flughäfen erfolgte bereits kurz nach der Jahrtausendwende am Zürcher Flughafen. Dabei wurde beabsichtigt, alle Einreisenden mittels Gesichtserkennung zu erfassen, um illegale Einreisen zu verhindern. Der damalige Datenschutzbeauftragte des Kantons Zürich, Bruno Baeriswyl, bezweifelte bei dieser Form der Überwachung sowohl das öffentliche Interesse (illegale Einreise ist keine schwere Straftat) als auch die Verhältnismässigkeit (der Überwachung 22 Millionen unverdächtiger Personen stünden jährlich 200 potenziell illegal eingereiste Migranten gegenüber) (Datenschutzbeauftragter des Kantons Zürich, S. 10). Trotz dieser Einwände startete die Pilotphase des Projekts mit etwas Verzögerung im Jahr 2002, nachdem die gesetzlichen Grundlagen angepasst worden waren. Eine definitive Einführung der Gesichtserkennung erfolgte zu diesem Zeitpunkt jedoch noch nicht (Datenschutzbeauftragter des Kantons Zürich, 29 f.).

Ein weiterer Einsatz biometrischer Gesichtserkennung seitens der Polizei am Flughafen Zürich begann Ende 2004 (Mäder 2005). Das Pilotprojekt «Secure Check» zwischen der Checkport Schweiz AG und Swissport Schweiz AG in Zusammenarbeit mit SWISS International Airlines und unter Beteiligung des EDÖB lief zwischen Dezember 2004 bis April 2005. In seinem Schlussbericht stellte der EDÖB dem Pilotprojekt eine überwiegend positive Beurteilung aus und formulierte lediglich kleinere Verbesserungsvorschläge für einen künftigen Einsatz, so etwa im Hinblick auf die Transparenz der Datenbearbeitung und im Hinblick darauf, keine unzulässige Weitergabe der erhobenen Daten an andere Stellen zu betreiben (EDÖB 2005). Die definitive Umsetzung der automatisierten Passkontrolle am Flughafen Zürich erfolgte schliesslich erst im Frühjahr 2018 nach erfolgreichem Abschluss einer halbjährigen Pilotphase seit Herbst 2017. Die automatisierte Passkontrolle ist freiwillig und Hinweisschilder vor Ort weisen auf den Einsatz der Gesichtserkennung hin. Zudem stehen herkömmliche Passkontrollschalter weiterhin zur Verfügung. Bei der Kontrolle erfolgt ein Vergleich zwischen der Live-Aufnahme des Gesichts vor Ort und dem im Reisepass gespeicherten biometrischen Foto. Zudem werden die Passdaten im Fahndungsregister überprüft. Sofern alles in Ordnung ist, werden die vor Ort erfassten personenbezogenen Daten nach Verlassen der Schleuse gelöscht (Regierungsrat des Kantons Zürich 2018). Seit Ende 2019 wird die automatisierte Passkontrolle z.B. auch am Genfer Flughafen getestet (Geneve Aeroport 2019).

Abgesehen vom Einsatz der Gesichtserkennung an Flughäfen verwenden aktuell zwei Kantonspolizeien Gesichtserkennungssoftware: die Kantonspolizei Aargau sowie die Kantonspolizei St. Gallen (Mac 2020). Andere Korps nutzen bislang entweder keine Gesichts-

erkennungssoftware, testen sie (Kantonspolizei Schaffhausen) oder setzen stattdessen auf menschliche Gesichtserkenner, sog. «Super Recognizer» (Glaus 2019). Der Einsatzbereich der Gesichtserkennung bei den drei genannten Korps beschränkt sich derzeit auf die Aufklärung von Verbrechen im Nachhinein. Die Basler Polizei hat zudem bereits Polizeiwagen von Tesla im Einsatz, in denen jeweils acht Kameras verbaut sind. Noch werden diese Kameras nicht zur Gesichtserkennung genutzt, eine Umstellung und Verknüpfung mit Gesichtserkennungssoftware wäre aber angesichts der Netzwerkfähigkeiten moderner Fahrzeuge ein leichtes (Fichter 2019).

Die folgende Betrachtung der technischen Grundlagen und Möglichkeiten bezieht sich auf die von den Kantonspolizeien Aargau und St. Gallen betriebene Gesichtserkennung. Bei der anschliessenden juristischen Betrachtung wird von den Anwendungsfällen abstrahiert und die anschliessende Analyse der gesellschaftlichen und ethischen Herausforderungen bezieht sich sowohl auf den allgemeinen Einsatz polizeilicher Gesichtserkennung als auch auf den spezifischen Einsatz beider Kantonspolizeien.

3.4.2. Technische Grundlagen und Möglichkeiten

Die bei der Aargauer Polizei eingesetzte Gesichtserkennungssoftware «better tomorrow» stammt von dem israelischen Unternehmen Anyvision. Die Software habe «in der Testphase zehn bis fünfzehn Prozent mehr mögliche Täterhinweise generiert» (Luchetta 2021b). Ob diese korrekt waren, wird jedoch nicht erwähnt, da das Team die Ergebnisse an die Ermittlungsbehörde vor Ort weitergibt, aber keine Rückmeldung bekommt (Luchetta 2021b). Eine objektive Bewertung der Erkennungsrate von aussen ist daher nicht möglich. Laut Hersteller könne «better tomorrow» Personen innerhalb von 0,2 ms mit einer Falscherkennungsrate von 0,1 % identifizieren (AnyVision 2021).

Unabhängige Tests zeigen indes, dass die beworbenen Erkennungsraten von den realen Erkennungsraten in Feldtests nicht erreicht werden: Mach et al. (2019, S. 238) setzten die Gesichtserkennungssoftware in einer realitätsnahen Umgebung zur Analyse von 13.000 Polizeifotos («mugshots») mit Männern und Frauen zwischen 15 und 80 Jahren aus verschiedenen Ethnien (wobei eine Überrepräsentation von Hellhäutigen bestand). Zudem wurden einige Bilder hinzugemischt, die in zufälligen Situationen aus unterschiedlichen Winkeln aufgenommen worden waren – so wie diese auch von Überwachungskameras aufgenommen würden. Hier zeigte sich, dass «better tomorrow» lediglich in 53 % der Fälle eine Person korrekt erkannte.

Die St. Galler Polizei setzt auf die laut Medienbericht von über 4000 Polizeibehörden weltweit eingesetzte Software «Analyze DI Pro» der schwedischen Firma Griffeye (Luchetta 2021b). Diese Software legt den Fokus stärker auf die Identifizierung und Kategorisierung von Objekten (Objekterkennung). So ist Gesichtserkennung nur ein Teil, wie die Erkennung von Fahrzeugen und Nummernschildern oder die Analyse von Bildmaterial zur Erkennung sexuellen Missbrauchs von Kindern (Griffeye 2021a). Hierzu werden u.a. auch Metadaten wie Ort und Zeit der Aufnahme mit einbezogen. Für eine Analyse können Daten (Bild oder Video) jedweder Überwachungskameras genutzt werden. Zudem können Open-Source-Daten mit in die Analyse eingebunden und eine Verbindung zu Datenbanken anderer Poli-

zeibehörden hergestellt werden (Griffeye 2021b). Für die Gesichtserkennung nutzt die Software die Luxand FACE technology (FaceSDK) (Forensic Focus 2019). Dabei handelt es sich um ein Software Development Kit, also eine Sammlung von Bibliotheken und Werkzeugen für Programmierer, die es ermöglicht, Anwendungen in verschiedenen Programmiersprachen zu erstellen. Laut Herstellerangaben könnten 70 Gesichtsmerkmale, Ausdrücke und Emotionen in Bild- und Videosequenzen, die Körpertemperatur sowie Alter und Geschlecht erkannt werden (Luxand 2021).

Im Gegensatz zu anderen Gesichtserkennungssoftwareanbietern setzt Griffeye auf einen Community-Ansatz. So gibt es ein Nutzerportal für den Austausch mit anderen Nutzern und Entwicklern. Zudem scheint die Software relativ offen zu sein, da es möglich ist, eigene Plug-Ins zu entwickeln bzw. einzubetten (Griffeye 2021a). Diese Eigenschaften sind auch der Grund dafür, dass die Software nicht nur bei Ermittlungsbehörden, sondern bspw. auch bei der Untersuchung des Flugzeugabsturzes von MH17 zur Anwendung kam, indem sie der Klassifizierung der Fotos von Wrackteilen diente (Gisolf et al. 2020, S. 2).

Vonseiten des Herstellers Griffeye gibt es keine Aussagen zu Erkennungsraten. Da dieser auf dem FaceSDK basiert, lassen sich jedoch Rückschlüsse ziehen. Im NIST Face Recognition Vendor Test (FRVT) (NIST 2020) erzielte der Algorithmus beim 1:1-Vergleich mit dem VISA-Fotos-Datensatz eine «false non-match rate» (also das Nicht-Erkennen einer Übereinstimmung) von 0,2814 bei einem Schwellenwert von ≤ 0,000001. Zum Vergleich: Der beste Algorithmus in dieser Kategorie erreichte einen Wert von 0,0018. Somit liegt der Algorithmus von Luxand auf Platz 186 von 195 und befindet sich über fast alle Datensätze hinweg im unteren Zehntel. Im FRVT 1:N-Vergleich findet sich der Algorithmus indes nicht (NIST 2021).

3.4.2.1. Untersuchung der Gesichtserkennung bei Kantonspolizeien

Zur Untersuchung des konkreten Einsatzes polizeilicher Gesichtserkennung in der Schweiz hat das Forschungsteam zu drei Kantonspolizeien, über die aus der Presseberichterstattung bekannt war, dass sie Gesichtserkennung einsetzen oder testen, Kontakt aufgenommen. Allen drei Polizeien wurde der Ethik-Fragekatalog Mitte 2021 (vgl. Tabelle 20 im Anhang) zugesandt. Im Folgenden werden zunächst die Antworten der Polizeien vorgestellt und diese anschliessend entlang der im Ethik- Fragekatalog formulierten Kategorien diskutiert und bewertet.

Die Schaffhauser Polizei teilte mit, dass sie derzeit keine Gesichtserkennungssoftware einsetzt, und konnte deshalb keine weiteren Auskünfte geben. Die Kantonspolizei St. Gallen und die Kantonspolizei Aargau beantworteten den Ethik-Fragekatalog. Die unten stehenden Ausführungen beziehen sich daher auf die Antworten dieser beiden Polizeistellen.

3.4.2.1.1. Kantonspolizei Aargau

Verantwortlichkeit und Rechenschaftspflicht:

Nach Angaben der Kantonspolizei Aargau werden die im Kontext der Gesichtserkennung verwendeten Daten auf einer zentralen Datenbank auf einem Server (Standalone-Server) gespeichert.

Die Frage der Verantwortungsadressierung ist klar geregelt und eine natürliche Person als Verantwortungsadressat bestimmt. Innerhalb der Kantonspolizei Aargau besteht zudem eine klare Regelung darüber, wer Zugang zu den Daten hat. Demnach haben ausschliesslich die Analysten der Kantonspolizei Zugang zu den Daten. Datentransfers an andere Institutionen erfolgen keine, sodass diesbezüglich keine Regelungen bestehen.

Der Kantonspolizei Aargau zufolge wurde vor der Inbetriebnahme des Gesichtserkennungsdienstes eine DSFA durchgeführt und dem kantonalen Datenschutzbeauftragten vorgelegt. Öffentlich zugänglich ist die DSFA nicht. Gegenstand der DSFA waren Fragen der Vertraulichkeit, Integrität, Verfügbarkeit, Zurechenbarkeit und Nachvollziehbarkeit. Zudem seien die Risiken für die Persönlichkeit und die Grundrechte betroffener Personen bewertet worden.

Für die Überprüfung der Sicherheitsmassnahmen, der gesetzlichen Vorgaben und der Reglemente, die der Gesichtserkennung zugrunde liegen, ist innerhalb der Polizei der ISO (Information Security Officer) der Kantonspolizei Aargau zuständig. Besonderen Wert lege der ISO einerseits auf die Datensicherheit und andererseits auf die Vermeidung von Risiken für die Persönlichkeit und die Grundrechte betroffener Personen. Von extern ist der kantonale Datenschutzbeauftragte für die Überprüfung der Einhaltung der Sicherheitsmassnahmen, gesetzlichen Vorgaben und Reglemente befasst.

Die Kosten der eingesetzten Gesichtserkennungssoftware belaufen sich auf CHF 3500 pro Jahr.

Da laut der Kantonspolizei Aargau keine softwarebasierte Personenidentifikation stattfinde, werden Betroffene auch nicht explizit über das Ergebnis der Gesichtserkennungssoftware informiert. Bestehende Beschwerdemöglichkeiten sind daher nicht auf die Gesichtserkennung im Speziellen bezogen, sondern wie in jedem Strafverfahren im Rahmen der Strafprozessordnung (StPO) geregelt. Zu den Haftungsfragen äusserte sich die Kantonspolizei Aargau nicht.

Sicherheit:

Den Angaben zufolge wurde die Sicherheit und Belastbarkeit des Systems im Rahmen der ex ante durchgeführten DSFA zuhanden des kantonalen Datenschutzbeauftragten überprüft. Dieser überprüfe die Einhaltung der Sicherheitsmassnahmen ex post.

Softwareupdates würden angeboten, dazu, ob diese auch regelmässig auf dem System der Kantonspolizei Aargau installiert werden, gab es keine Antwort. Der Grund für die Wahl eines Standalone-Servers für den Betrieb der Gesichtserkennungsanwendung diene dem Ziel der Gewährleistung der Sicherheit. Zugang zu den Analyseergebnissen und den Rohdaten habe nur autorisiertes Personal. Jeder Zugriff werde einzeln geloggt, um die Zurechenbarkeit zu gewährleisten. Dies sei mittels eines User Management-Tools möglich, das

drei verschiedene Berechtigungsstufen erlaubt, *Admin*, *Super User* und *User*. Alle zugangsberechtigten Personen verfügten demnach über einen eigenen Login und Passwort. Logfiles würden erstellt. Gemäss den Angaben des Softwareherstellers gäbe es für berechtigte Personen mit Admin-Rechten eine Möglichkeit, die Logfiles zu löschen.

Transparenz und Erklärbarkeit:

Den Angaben des Softwareherstellers zufolge war der dem NIST vorgeführte Algorithmus ein Derivat des von der Kantonspolizei Aargau verwendeten Algorithmus. Der Ursprung und der Inhalt der Trainingsdaten ist der Kantonspolizei Aargau nicht bekannt. Zudem ist der verwendete Algorithmus nicht frei verfügbar. Den Angaben des Softwareherstellers Anyvision zufolge ist die von NIST begutachtete Softwareversion verwandt mit der von der Kantonspolizei Aargau verwendeten Version, aber nicht dieselbe.

Transparenz der Anwendung:

Es erfolgt keine unabhängige Begutachtung der Leistungsfähigkeit des Algorithmus im Einsatzkontext der Kantonspolizei Aargau. Die Evaluation der Leistungsfähigkeit erfolgt seitens des Kantonspolizei-Personals.

Die der Gesichtserkennungssoftware zugrunde liegenden Daten stammen aus der Datenbank des Aargauer Polizeisystems und der Falldatenbank Picar. Im Bedarfsfall kann die Kantonspolizei Aargau im Rahmen eines Strafverfahrens Zugang zu privatem Bildmaterial einfordern und dieses über die Aufnahme in die üblichen Datenbanken in die Gesichtserkennungssoftware einspeisen. Stimm- und Spracherkennung wird bei der Kantonspolizei Aargau nicht eingesetzt.

Eine generelle Information der Betroffenen, deren Gesichter ggf. Gegenstand der polizeilichen Recherchen sind, findet nicht statt. Lediglich dann, sofern ein Sachbearbeiter den Hinweis auf eine mögliche Übereinstimmung von Bildern erhalten hat, kann dieser den Betroffenen mit dem Bildmaterial konfrontieren, in deren Zuge der Betroffene die Möglichkeit des Widerspruchs erhält.

Über den Einsatz der Gesichtserkennung, etwa darüber, wie viele Stellen auf Daten zugegriffen haben, wird weder gegenwärtig öffentlich Bericht erstattet, noch sei dies geplant. Diesbezüglich verweist die Kantonspolizei Aargau auf die regelmässige Überprüfung durch den kantonalen Datenschutzbeauftragten.

Gerechtigkeit, Fairness und Nicht-Diskriminierung, Bias-Vermeidung:

Den Angaben des Softwareherstellers Anyvision zufolge wird erheblicher Aufwand bezüglich der Vermeidung eines Bias sowohl im Bereich der Trainingsdaten als auch des Algorithmus betrieben. Auf die Anfrage des Forschungsteams wurde ausführlich geantwortet.

Die wechselhafte Qualität des Bildmaterials werde bei der Evaluation der möglichen Treffer stets berücksichtigt. Zur Vermeidung qualitätsbasierter Fehlentscheidungen werden die Softwareergebnisse stets von einem Analysten interpretiert und der Entscheid seitens des Analysten und nicht seitens der Software gefällt. Bekannter Bias werde während dieser Evaluation mitberücksichtigt.

Bei der Gestaltung der Gesichtserkennungsanwendung wurden keine Meinungen anderer Stakeholder (etwa der Betroffenen oder zivilgesellschaftlichen Akteuren) eingeholt.

Menschliche Kontrolle der Technik:

Nach Angaben der Kantonspolizei Aargau werden keine Entscheidungen nur auf Basis der Ergebnisse der Gesichtserkennungssoftware getroffen. Die Verwendung der Ergebnisse basiere auf zwei Schritten: Im ersten Schritt wertet ein Kantonspolizei-Analyst die Ergebnisse aus. Dazu vergleicht der Analyst die Softwareergebnisse mit anderen Bildern, die ihm zur Verfügung stehen (Videos, andere Winkel, andere Referenzbilder). Ein wichtiger Faktor bei diesem Schritt sei die Berücksichtigung der Bildqualität. Falls der Analyst schliesslich zu dem Ergebnis gelange, dass es sich um die gesuchte Person handeln *könnte*, würden die Ergebnisse zunächst von einer zweiten Person überprüft. Der Kantonspolizei Aargau zufolge handelt es sich bei dieser Feststellung nicht um eine Identifikation, da keine abschliessende Feststellung gemacht werde, sondern lediglich um die Erstellung eines Hinweises, demzufolge es sich um die gesuchte Person handeln könnte, dies aber nicht zwingend der Fall sein muss.

Im zweiten Schritt werde der generierte Hinweis an den Sachbearbeiter des Falles weitergegeben. Die Entscheidung, wie mit dem Hinweis umgegangen wird, fälle schliesslich der Sachbearbeiter, indem dieser den Hinweis gemeinsam mit den anderen Informationen, die ihm zur Verfügung stehen, evaluiert.

Bei den Analysten, die mit der Auswertung der Softwareergebnisse betraut sind, handele es sich um Personal, das hins. der korrekten Überprüfung und Interpretation der Ergebnisse geschult wurde.

3.4.2.1.2. Kantonspolizei St. Gallen

Verantwortlichkeit und Rechenschaftspflicht:

Nach Angaben der Kantonspolizei St. Gallen sind die im Kontext der Gesichtserkennung verwendeten Daten (Bilder) sowie die Gesichtserkennungssoftware auf einer Workstation gespeichert bzw. installiert. Dieser PC ist offline und nicht mit dem Internet verbunden. Der Datentransfer erfolge *vorwiegend* mit einem Speichermedium (USB-Stick, externe Festplatte usw.). Darüber, welche anderen Formen des Datentransfers stattfinden, wurde keine Auskunft erteilt.

Als Verantwortungsadressat sei eine natürliche Person bestimmt. Über Zugriffsrechte verfügten einerseits die IT-Forensiker, die mit der Datenaufbereitung befasst sind, andererseits – und anders als bei der Kantonspolizei Aargau – aber auch die für einen Fall zuständigen Sachbearbeiter.

Der Kantonspolizei St. Gallen zufolge sei vor Inbetriebnahme des Dienstes eine Datenschutzfolgenabschätzung durchgeführt worden. Öffentlich zugänglich ist die DSFA nicht und eine Veröffentlichung sei auch nicht vorgesehen. Eine sonstige, unabhängige und öffentliche Evaluation habe nicht stattgefunden. Die Softwareresultate würden ausschliesslich vom Polizeipersonal gesichtet und verifiziert.

Die zugrunde liegende Hardware und die Software hätten einmalige Kosten in Höhe von ca. CHF 11.000 (Hardware) bzw. ca. CHF 1750 (Software) verursacht. Die regelmässigen Betriebskosten betrügen CHF 400 für die Hardware und CHF 1750 für die Software.

Gemäss der Kantonspolizei St. Gallen fliessen die mittels Gesichtserkennungssoftware gewonnenen Erkenntnisse in die Ermittlungstätigkeit ein. Betroffene würden über den Einsatz der Gesichtserkennung nicht informiert. Die Entscheidung darüber, ob das Softwareanalyseergebnis brauchbar ist oder nicht, liege nach Aussagen der Kantonspolizei bei dem zuständigen Gericht, so wie es bei anderen Ermittlungsverfahren auch üblich sei. Entsprechend verweist die Kantonspolizei St. Gallen sowohl hins. Beschwerdemöglichkeiten als auch hins. Haftungsfragen auf die zuständigen Gerichte.

Sicherheit:

Die Sicherheit und Belastbarkeit des Gesichtserkennungssystems wurde der Kantonspolizei St. Gallen zufolge mittels interner Test-Cases überprüft, eine unabhängige Begutachtung habe nicht stattgefunden. Regelmässige Softwareupdates stünden sowohl zur Verfügung als auch würden diese regelmässig installiert.

Da das Gesichtserkennungssystem als nicht kritische Infrastruktur eingestuft worden sei, sei die Frage der Verfügbarkeit nicht relevant. Die Überprüfung der Echtheit der Ergebnisse erfolge mittels visueller Verifizierung seitens der zuständigen Ermittlerinnen und Ermittler. Die Resilienz des Systems sei dadurch gegeben, dass nur autorisierte Personen Zugriff auf das System bzw. die Räumlichkeiten hätten, in denen sich die Workstation befindet. Zur Gewährleistung der Vertraulichkeit werde ein entsprechender Zugang benötigt. Weil lediglich ein lokaler Login besteht, sei die Zuordbarkeit der Zugriffe lediglich über die zugeteilten Fälle nachvollziehbar. Zugleich existierten keine unlöschbaren Logfiles.

Transparenz und Erklärbarkeit:

Der zugrunde liegende Algorithmus ist der Kantonspolizei St. Gallen nicht bekannt. Ebenso sind der Ursprung sowie die Inhalte der Trainingsdaten nicht bekannt. Die Fragen, ob der Algorithmus frei verfügbar bzw. open source ist oder ob es Prozeduren zur unabhängigen Begutachtung des genutzten Algorithmus gibt, konnten von der Kantonspolizei St. Gallen ebenfalls nicht beantwortet werden.

Zudem erfolge auch keine unabhängige Begutachtung des verwendeten Algorithmus im konkreten Anwendungskontext der Kantonspolizei St. Gallen. Die der Gesichtserkennung zugrunde liegende Polizeidatenbank beinhalte sog. erkennungsdienstliche Personenbilder, deren Inhalt gem. der Kantonspolizei St. Gallen rechtmässig erworbene Bilder sind. Die Gesichtsfotos gesuchter Personen stammten von privaten Überwachungskameras und von öffentlichen Stellen betriebenen Kameras, die der Polizei ausgehändigt würden.

Transparenz der Anwendung:

Eine Information der Betroffenen über den Einsatz der Gesichtserkennung oder darüber, wer Zugriff auf die Bilder hat, erfolge nicht, weil eine rechtliche Grundlage für den Einsatz der Software vorhanden sei. Dabei sei es irrelevant, ob das menschliche Auge oder eine Software die Bilder durchsuche, solange es sich um rechtmässig erworbene Bilder handele. Auf die Frage, ob öffentlich über den Einsatz der softwarebasierten Gesichtserkennung be-

richterstattet wird, verwies die Kantonspolizei St. Gallen ebenfalls darauf, dass dies nicht erfolge, weil es irrelevant sei, ob das menschliche Auge oder eine Software die Bilder durchsuche, solange es sich um rechtmässig erworbene Bilder handele.

Gerechtigkeit, Fairness und Nicht-Diskriminierung:

Die Gesichtserkennungssoftware der Kantonspolizei St. Gallen wird als Hilfsmittel bei der Auswertung grosser Mengen Video- und Bilddateien zum Zwecke der Identifikation gesuchter Personen eingesetzt, um personelle Ressourcen zu entlasten. Der Kantonspolizei St. Gallen ist nicht bekannt, ob die Trainingsdaten oder der Algorithmus im Hinblick auf ein potenzielles Bias untersucht wurden. Daher konnte auch nicht über ein potenzielles Bias informiert werden. Hinsichtlich des Umgangs mit der wechselhaften Qualität des Inputmaterials bestehen bei der Kantonspolizei St. Gallen keine besonderen Prozeduren.

Auf die Frage, ob Mechanismen zur Evaluation der Daten und Betriebsprozesse in Kraft sind, erwiderte die Kantonspolizei St. Gallen, dass die Betriebsprozesse definiert und dokumentiert seien. Hinsichtlich des Einbezugs partizipativer Elemente bei der Gestaltung der Gesichtserkennung wurde darauf verwiesen, dass bei der Evaluation ein Pflichtenheft mit verschiedenen Stakeholdern erstellt worden sei.

Menschliche Kontrolle der Technik:

Die Ergebnisse der Gesichtserkennungssoftware würden bei der Kantonspolizei St. Gallen stets von einem/r Sachbearbeiter/in kontrolliert und interpretiert. Diese würden vor dem Einsatz einer speziellen Schulung unterzogen. Neben dieser Schulung sei aber auch Erfahrung in der Ermittlungsarbeit sehr wichtig.

3.4.2.2. Zusammenfassung und Bewertung

Zu begrüssen sind bei beiden Kantonspolizeien zahlreiche vorhandene Regelungen für einen ordnungsgemässen Betrieb, darunter die Verantwortungsadressierung, die Durchführung einer DSFA vor der Inbetriebnahme des Gesichtserkennungsdienstes, Vorkehrungen zur Gewährleistung der IT-Sicherheit, der Zugriff auf ausschliesslich gesetzlich erworbene Bilddaten. Bei der Kantonspolizei Aargau sind zudem ausführliche Kenntnisse über die verwendete Software und den Algorithmus sowie organisatorische Massnahmen zur Vermeidung von Bias positiv hervorzuheben.

Kritisch zu betrachten ist hingegen, dass beide Kantonspolizeien weder die durchgeführte DSFA veröffentlichen noch anderweitige unabhängige Begutachtungen der Leistungsfähigkeit des Systems durchgeführt und veröffentlicht werden sollen. Ebenso sollen keine Berichte darüber veröffentlicht werden, welche Stellen auf welche Daten zugegriffen haben. Zudem erkennt die Kantonspolizei St. Gallen auch keine Notwendigkeit der öffentlichen Berichterstattung des Einsatzes von Gesichtserkennungssoftware, weil es keinen Unterschied mache, ob das menschliche Auge oder eine Software die Bilder durchsuche.

Beide Kantonspolizeien teilen zudem die Meinung, dass eine Information der von Gesichtserkennung Betroffenen grundsätzlich nicht notwendig sei.

Bei der Kantonspolizei St. Gallen ist zudem keine technische Zuordnung der Datenzugriffe gewährleistet. Dadurch kann eventueller Missbrauch schwieriger nachvollzogen werden, als wenn mittels unlöschbarer Logfiles klare Evidenz darüber vorherrschte, welche Beamten zu welcher Zeit Zugriff auf welche Daten hatten. Problematisch ist auch, dass die Kantonspolizei St. Gallen weder wusste, welcher Algorithmus der Gesichtserkennungssoftware zugrunde liegt, noch Ursprung und Inhalte zu den Trainingsdaten oder zu einem möglichen Bias benennen konnte. Das Fehlen gesonderter Prozeduren zum Umgang mit der wechselhaften Qualität des Inputmaterials birgt das Risiko von Falscherkennungen.

Dass die Polizeien v.a. auf die Durchsicht privaten Videomaterials setzen, entspricht zudem einem generellen Trend, der in den vergangenen Jahren zu beobachten war. Schon bei der klassischen Videoüberwachung des öffentlichen Raumes, die ohne Gesichtserkennung auskommt, wurde in zunehmendem Masse auf die Unterstützung seitens privater Stellen gesetzt. Der Grund dafür liegt darin, dass polizeiliche Überwachung stets verhältnismässig sein muss, also nur so lange dauern darf, wie es für einen konkreten Zweck erforderlich ist und somit gesetzlich klar geregelt ist (etwa zur Überwachung bei Fussballspielen oder Demonstrationen mittels mobiler Kamerateams). Dieser Gesetzesvorbehalt erschwert die massenhafte Installation von Überwachungskameras in der Öffentlichkeit, die von öffentlichen Stellen betrieben werden. Sofern die Überwachung zudem nicht dauerhaft erfolgen darf, kommen zur Schwierigkeit des Gesetzesvorbehalts auch die Montage- bzw. Demontagekosten hinzu, sodass sich ein befristeter Einsatz womöglich bereits aus Kostengründen nicht lohnt. Dementsprechend greifen öffentliche Stellen im Bedarfsfall für Ermittlungszwecke zunehmend auf das Bildmaterial privat betriebener Überwachungskameras zurück (Ryser 2019). Zudem gibt es im Hinblick auf die Speicherung und Weitergabe der erfassten Daten keine allgemeingültigen und strengen Vorgaben, sofern Überwachungskameras von Privaten betrieben werden, was der polizeilichen Ermittlungstätigkeit insofern zugutekommt, weil die gespeicherten Daten somit einen grösseren Zeitraum abdecken können (Tobler 2019). Darüber, wie viele private Videokameras im Einsatz sind, wo diese eingesetzt werden und ob und inwiefern sie miteinander vernetzt sind, herrscht weitgehende Unkenntnis. Als bspw. 2016 in Zürich die Einführung eines Registers aller im öffentlichen Raum installierten Kameras zur Debatte stand, stimmte der Kantonsrat dagegen (Tobler 2019). Transparent ist nur der Bestand der von der Polizei selbst betriebenen Überwachungskameras. Demnach waren z.B. in Zürich Anfang 2021 nur 22 fest installierte Videoüberwachungskameras, von denen wiederum 18 nur bis zum 5. April 2021 betrieben werden durften, im Einsatz (Stadtpolizei Zürich 2021). Rückschlüsse auf den Kamerabestand sind aber auch über andere Wege möglich. Das Projekt «Surveillance under Surveillance» betreibt seit August 2016 eine Webseite, auf der Freiwillige installierte Überwachungskameras samt Ort, Typ und Blickwinkel melden können. Diese Zahlen geben zwar nicht den tatsächlichen Bestand wieder, sind aber eine hilfreiche Annährung (Reuter 2017). Stand 9. Mai 2022 waren in der Schweiz 1764 Überwachungskameras gemeldet, davon rund 200 allein in Zürich (Surveillance under Surveillance 2022). Dass diese Zahlen nur eine hilfreiche Annäherung darstellen und kein offizielles öffentliches Register ersetzen können, zeigen allerdings die Recherchen der Zürcher Lokalzeitung aus Anfang 2018. Demnach betreibt allein die Stadt Zürich nicht nur Hunderte, sondern Tausende Videokameras (Lokalinfo.ch 2018).

3.4.3. Juristische Bewertung

In gewissen Kantonen wird Gesichtserkennungstechnologie für die Identifizierung einer Person in Ermittlungen und Strafverfahren eingesetzt (Riniker 2021; Kühne 2022; Simmler und Canova 2021). Im nachgehenden Abschnitt soll aber insb. auf die verdachtsunabhängige Echtzeitgesichtserkennung eingegangen werden. Dies bedeutet, dass eine unbestimmte Anzahl von Personen ohne spezifischen Tatverdacht gefilmt wird und diese Daten zeitgleich ausgewertet werden. Typische Anwendungen sind die Suche von vermissten Personen oder das Aufspüren von Straftätern. Die Technologie kann aber auch für die (präventive) Überwachung grosse Menschenansammlungen wie etwa Demonstrationen eingesetzt werden. Da es sich hierbei für die Schweiz – soweit ersichtlich – um ein hypothetisches Szenario handelt, können die rechtlichen Schranken nicht umfänglich dargestellt werden. Es sollen die Rahmenbedingungen, die sich aus der Bundesverfassung ergeben, dargestellt und auf datenschutzrechtliche Vorschriften hingewiesen werden.

Eine Massenüberwachung stünde *per se* in einem gewissen Spannungsverhältnis zum demokratischen Rechtsstaat. Der Staat ist in seinem Handeln an das Recht gebunden. Hierfür sind die rechtsstaatlichen Grundsätze und Grundrechte massgebend; das staatliche Handeln hat insb. verhältnismässig zu sein (Epiney 2016).

3.4.3.1. Massenüberwachung: Eingriff in den Schutzbereich von Grundrechten

Bei einer Massenüberwachung mittels Echtzeitgesichtserkennung können verschiedene Grundrechte in ihrem Schutzbereich betroffen sein (Braun Binder et al. 2021, S. 59).

Vom Einsatz einer Massenüberwachung mittels Echtzeitgesichtserkennung könnte der Grundsatz der Menschenwürde betroffen sein. Dieser wirkt als eine Art Fundament für alle übrigen Grundrechte. Der Kerngehalt anderer Grundrechte lässt sich auf die Menschenwürde zurückführen (siehe unten).[75]

Das Diskriminierungsverbot ist eng mit der Menschenwürde verbunden. Es bietet Schutz gegen Herabwürdigung, Stigmatisierung oder soziale Ausgrenzung und Unterdrückung (Waldmann 2015). Die Person ist als Individuum wahrzunehmen und darf nicht aufgrund ihrer Zugehörigkeit zu einer bestimmten sozialen Gruppe stigmatisiert oder ausgegrenzt werden (Waldmann 2015). Gesichtserkennungssysteme ergaben in der Vergangenheit weniger gute Trefferquoten, speziell bei Menschen mit dunkler Hautfarbe, Frauen, Kindern und auch bei älteren Leuten. Diese Personen unterliegen daher eher einer Gefahr eines falschpositiven Resultats und einer damit verbundenen Personenkontrolle (FRA 2019). Aufgrund der erhöhten Kontrolldichte kann dadurch das Vertrauen in den Staat und die Polizei untergraben werden (FRA 2019).

[75] Art. 7 BV.

Aus dem Schutzbereich der Verfahrensrechte[76] lässt sich ableiten, dass sich Personen bezüglich der Anwendung einer Gesichtserkennungssoftware bewusst sein müssen (FRA 2019).

Das Bearbeiten von Gesichtsbildern stellt einen schweren Eingriff in das Recht auf Schutz der Privatsphäre dar, schliesslich ginge das Vertrauen einer weitgehenden Anonymität im öffentlichen Raum durch die Anwendung der Gesichtserkennungstechnologie verloren.[77] Ein solcher Eingriff besteht unabhängig davon, ob ein Treffer vorliegt oder die Daten direkt nach Datenabgleich gelöscht werden.

Beim Einsatz einer Gesichtserkennungssoftware werden Gesichtsbilder zum Zweck der Identifizierung von Personen gesammelt und ausgewertet, daher liegen biometrische Daten gemäss Art. 5 lit. c Ziff. 4 nDSG vor (FRA 2019). Biometrische Daten sind einzigartig, können kaum verändert werden und sind ein bedeutendes Merkmal der Persönlichkeit (FRA 2019). Das Gesichtsbild ist aufgrund seiner Einzigartigkeit und der Möglichkeit einer eindeutigen Abgrenzung gegenüber anderen Individuen ein bedeutendes Merkmal der Persönlichkeit.[78]

In der Bevölkerung könnten Befürchtungen bezüglich einer Zweckentfremdung der Daten aufkommen (*function creep*). Bei einer umfassenden Datenbearbeitung könnten demokratische Werte unterwandert werden, da die Privatheit ein inhärentes Recht der liberalen Demokratie und des pluralistischen Rechtsstaates ist und auch die Basis für die Ausübung anderer Grundrechte bildet (FRA 2019). Das Grundrecht der persönlichen Freiheit könnte in ähnlicher Weise eine Beeinträchtigung erfahren (Magnin 2017, S. 152).[79]

Die Meinungs- und die Versammlungsfreiheit sind weitere durch den Einsatz der Gesichtserkennungstechnologie betroffene Grundrechte. Die Meinungsfreiheit schützt das Recht jeder Person, ihre Meinung frei zu bilden, ungehindert zu äussern und zu verbreiten (Hertig 2015a).[80] Die Versammlungsfreiheit räumt Personen das Recht ein, Versammlungen zu organisieren, daran teilzunehmen oder fernzubleiben (Hertig 2015a).[81] Der Einsatz von Gesichtserkennungstechnologien könnte zum *chilling effe*ct führen (Verzicht auf Teilnahme an überwachten Kundgebungen aufgrund Befürchtung negativer Konsequenzen) (FRA 2019). Das effektive Funktionieren einer partizipierenden Demokratie könnte damit gefährdet werden (Braun Binder et al. 2021, S. 60; FRA 2019, S. 84).

3.4.3.2. Der absolut geschützte Kerngehalt

Der Kerngehalt eines Grundrechts erfährt gemäss Art. 36 Abs. 4 BV absoluten Schutz. Dies betrifft den Teil der Grundrechte, welcher fundamental ist und daher nicht eingeschränkt

---

[76] Art. 29 BV, Art. 6 EMRK und Art. 14 UNO-Pakt II.

[77] Siehe Art. 34 Abs. 2 nDSG.

[78] M.w.H. ECtHR, Guide on Article 8 of the European Convention on Human Rights Right to respect for private and family life, home and correspondence, Strasbourg, Council of Europe, 31 August 2020, § 153.

[79] Art. 10 Abs. 2 BV.

[80] Art. 16 Abs. 2 BV. Art. 10 EMRK.

[81] Art. 22 BV.

werden darf. Die Kerngehalte müssen für jedes Grundrecht einzeln eruiert werden; Hinweise bieten jedoch das Prinzip der Menschenwürde und die rechtsstaatlichen Grundsätze (Epiney 2015).

Der bewusst offengehaltene Begriff der Menschenwürde soll gemäss Bundesgericht das letztlich nicht fassbare Eigentliche des Menschen schützen und den Einzelnen in seiner Werthaftigkeit und individuellen Einzig- und allfälligen Andersartigkeit anerkennen.[82] Insb. verbietet das Gebot der Menschenwürde, den Einzelnen zum Objekt herabzusetzen (Belser und Molinari 2015). Wird Gesichtserkennungstechnologie eingesetzt, könnte dies u.U. einen Eingriff in das Gebot der Achtung der Menschwürde bedeuten. Die Verwendung von Gesichtserkennungstechnologie im öffentlichen Raum könnte zum diffusen Gefühl eines ständigen Überwachtwerdens, mithin einem Gefühl des «gläsernen Bürgers» führen. Dazu schreibt die *Grundrechteagentur der EU* (FRA) treffend, eine Massenüberwachung könnte bewirken, dass sich Personen während ihres Aufenthalts auf öffentlichen Plätzen unwohl fühlen. Als Folge davon würden Personen ihr Verhalten u.U. entsprechend anpassen und Orte, die überwacht werden (wie Bahnhöfe oder kulturelle oder soziale Veranstaltungen), nicht mehr frequentieren. Abhängig vom Einsatz und der Umsetzung der Gesichtserkennungstechnologie im Bereich Massenüberwachung könnte dies fundamentale Auswirkungen auf Einzelne haben und in diesem Sinne würdevolles Leben beeinträchtigen (FRA 2019). Hinzu kommt, dass als unmenschlich empfundene Personenkontrollen, die aufgrund eines Treffers durch das System herbeigeführt werden, für die Betroffenen – insb. im Falle der Häufung von Personenkontrollen aufgrund falsch-positiver Treffer – grossen Stress bedeuten können (FRA 2019).

Die Gesichtserkennungstechnologie ermöglicht einzelnen Personen, das Verhalten von Tausenden auf Unregelmässigkeiten hin zu untersuchen und vermeintliche Störer zu identifizieren, aber auch nach bestimmten Einzelpersonen zu suchen und daraus genaue Bewegungs- resp. Persönlichkeitsprofile zu erstellen. Diese Asymmetrie erinnert an die Ausgestaltung eines Überwachungsstaates. Bei undifferenziertem und weitreichendem Einsatz der Technologie im öffentlichen Raum wäre daher die Menschenwürde tangiert und die Umsetzung unzulässig. Die gesichtserkennungstechnologische Registrierung aller wäre ebenfalls unzulässig, da das Individuum nicht mehr als Subjekt, sondern als Objekt behandelt werden würde (Held, S. 63–65; Stettner, S. 141–143). Die routinemässige Registrierung und weitreichende Überwachung wäre daher mit der Menschenwürde nicht vereinbar und könnte auch nicht mithilfe einer gesetzlichen Grundlage legitimiert oder aufgrund eines öffentlichen Interesses gerechtfertigt werden.

Das Recht auf persönliche Freiheit/Privatsphäre oder das informationelle Selbstbestimmungsrecht könnten ebenfalls in ihrem Kerngehalt getroffen werden. Gemäss der bundesgerichtlichen Rechtsprechung greift die punktuelle und gewöhnliche Videoüberwachung im öffentlichen Raum nicht in den Kerngehalt ein.[83] Dieses Urteil bezog jedoch die Möglichkeit der Verknüpfung mit Gesichtserkennungstechnologie nicht mit ein. Bei einer umfassenden Überwachungs- und Identifizierungsmöglichkeit im öffentlichen Raum besteht hingegen die

[82] BGE 132 I 49 E. 5.1.

[83] BGE 132 V 241 E. 2.5.1; BGE 137 I 327 E. 5.1.

Gefahr, dass dem Einzelnen diese Überwachung jederzeit bewusst ist und er deswegen sein Verhalten und seine Willensbildung anpassen würde, was den Kerngehalt ebenfalls verletzen würde. Zudem wäre eine Fortbewegung in Anonymität nicht mehr möglich. Aus diesen Gründen kann davon ausgegangen werden, dass der Kerngehalt des Rechts auf Privatsphäre tangiert wäre. Wohl auch hoch wäre die Angst vor Missbrauch der erfassten Gesichtserkennungsdaten in der Bevölkerung, da der Staat die Identifizierung auch für andere Zwecke als die vorgegebenen verwenden könnte (*function creep*).

Der punktuelle Einsatz der Gesichtserkennungstechnologie in spezifischen Situationen soll nachgehend auf seine Rechtfertigung geprüft werden.

3.4.3.3. Gesetzliche Grundlage

Damit ein Grundrechtseingriff gerechtfertigt werden kann, benötigt es ein hinreichend bestimmtes Gesetz im formellen oder materiellen Sinn, welches eine generell-abstrakte Regelung enthält (Epiney 2015). Verlangt wird sowohl die genügende Normdichte als auch die genügende Normstufe. Bei schweren Eingriffen in Grundrechte, wie sie hier vorliegen, hat ein Gesetz im formellen Sinn vorzuliegen. Nebst den bereits erwähnten grundrechtlichen Schranken für den Einsatz einer Gesichtserkennungstechnologie ergibt sich die Notwendigkeit einer formell-gesetzliche Grundlage auch aus dem Datenschutzgesetz (Bundesrat).

Der Grad der Bestimmtheit einer Norm lässt sich nicht abstrakt festlegen. Vielmehr sind die Anforderungen an die Bestimmtheit im Einzelfall zu prüfen (Epiney 2015). Die Normdichte hängt u.a. von der Vielfalt der zu ordnenden Sachverhalte, der Komplexität und der Vorhersehbarkeit der im Einzelfall erforderlichen Entscheidung, den Normadressaten und der Schwere des Eingriffs ab.[84] Bei fehlender Bestimmtheit ist für den Einzelnen nicht erkennbar, unter welchen Voraussetzungen die Grundrechtseinschränkungen erlaubt sind. Insb. im Polizeirecht, aber auch im Bereich des Staatsschutzes und bei eher technischen Materien können die Anforderungen der Bestimmtheit herabgesetzt werden. Dennoch sind das Ermessen und die Tragweite so zu beschreiben, dass die Voraussehbarkeit des staatlichen Handelns für den Einzelnen gewahrt bleibt (Epiney 2015). Da es sich bei der Gesichtserkennung um einen schweren Eingriff in die Grundrechte handelt, müssten in einem Gesetz im formellen Sinn zumindest die Grundzüge festgelegt und allenfalls in einer Verordnung genauer spezifiziert werden (CoE 2021). Das Gesetz soll sicherstellen, dass die Behörde anhand verschiedener Faktoren, u.a. Ort und Zeit, im Einsatz der Technologien auf das Erforderliche und Verhältnismässige begrenzt bleibt (CoE 2021).

Der Europarat hat in Auslegung der Konvention 108+ Richtlinien zur Gesichtserkennung veröffentlicht (CoE 2021). In diesen Richtlinien wird betont, dass aufgrund des Ungleichgewichts zwischen Privaten und öffentlicher Hand eine Einwilligung nicht als Rechtfertigung für den Einsatz einer Gesichtserkennungstechnologie dienen kann (CoE 2021). Vielmehr ist im Gesetz Folgendes festzuhalten (FRA 2019):

[84] BGE 131 II 271 E. 6; BGE 135 I 169 E. 5.4.1.

- Klare Definition des Ziels und Zwecks des Einsatzes der Gesichtserkennungstechnologie (z.B. Verfolgung bestimmter Straftaten) (European Commission 2021).
- Art und Umfang der Datenbearbeitung, insb. Zeitpunkt, Ort und überwachte Personen (z.B. Personen auf der Watchlist) (FRA 2019).
- Beteiligte Behörden bzw. Zugriffsberechtigte, insb. die verantwortliche Stelle
- Kategorien der bearbeiteten Daten
- Regelung der Aufbewahrung und Löschung der Daten
- Gewährleistung der Rechte der Betroffenen
- Technische Umsetzung des Überwachungssystems, minimale Zuverlässigkeit und Genauigkeit des Algorithmus, Rückverfolgbarkeit des Prozesses, Sicherheitsmassnahmen und die Systemverantwortlichen (CoE 2021).

3.4.3.4. Öffentliches Interesse

Sowohl öffentliches Interesse als auch Schutz der Grundrechte Dritter können Einschränkungen in die Grundrechte rechtfertigen. Es gibt keine abschliessende Auflistung der öffentlichen Interessen, vielmehr steht dem Gesetzgeber ein grosser Gestaltungsspielraum zu, solange er die verfassungsrechtlichen Wertentscheidungen respektiert. Das öffentliche Interesse an der Massenüberwachung besteht in der Wahrung der öffentlichen Sicherheit und Ordnung. Dabei geht es insb. um die Verhinderung oder Aufklärung von Terrorangriffen oder schwerer Delikte gegen hohe Rechtsgüter (Straftaten gegen Leib und Leben) (Sigrist 2014, S. 4–7). Der Schutz der öffentlichen Sicherheit und Ordnung sowie der öffentlichen Gesundheit sind sog. Polizeigüter und anerkannte öffentliche Interessen (Epiney 2015). Weitere öffentliche Interessen könnten der Grundrechtsschutz Dritter (Auffinden von entführten Personen) oder auch die Effizienz von Ermittlungsverfahren sein.

Es besteht aber auch ein öffentliches Interesse am Schutz der persönlichen Freiheit, der Privatsphäre und dem Schutz der personenbezogenen Daten und auch der Ausübung von anderen Grundrechten wie Versammlungs- und Meinungsfreiheit, da diese wichtig für eine funktionierende Demokratie sind.

3.4.3.5. Verhältnismässigkeit

Ein Grundrechtseingriff ist immer auch zu überprüfen bezüglich der Verhältnismässigkeit der Massnahme (Epiney 2015). Dabei stellt sich die Frage, ob der Eingriff geeignet, erforderlich und zumutbar ist.

Die Frage nach der **Eignung** betrifft das Erreichen des angestrebten Zwecks durch die eingesetzten Mittel. Hier wäre bei den im Vorfeld besprochenen Fällen mit hoher Fehlerquote und falsch-positiven Resultaten die Eignung der Systeme wohl zu verneinen. Es lässt sich nicht generell beantworten, welchen Präzisionsgrad die Systeme aufweisen müssten, um als geeignet zu gelten. Die Zielsetzung spielt dabei eine wichtige Rolle. Grundsätzlich aber erscheint der Einsatz von Gesichtserkennungstechnologien für die Auffindung von vermissten Personen oder auch von Straftätern oder zur Überwachung von Personen als geeignet.

Damit können nämlich innerhalb kurzer Zeit viele Datensätze ausgewertet und analysiert werden. Ohne den Einsatz der Technologie würde dies sehr viel mehr Zeit benötigen oder gar nicht umsetzbar sein.

Bezüglich der **Erforderlichkeit** stellt sich die Frage, ob zur Erreichung des Zwecks nicht ein milderes Mittel zur Verfügung steht. Herkömmliche Videokameras bedeuten im Vergleich zu einer Echtzeitgesichtserkennung einen milderen Eingriff. Zur Aufklärung von Straftaten können auch sog. Super Recognizer eingesetzt werden (m.w.H. Schindler, 2019). Die Erforderlichkeit ist für die jeweilige Situation, für welche der Einsatz der Echtzeitgesichtserkennung angedacht wird, einzeln zu prüfen. Auf jeden Fall ist der Einsatz immer zeitlich und geografisch auf das Minimum zu reduzieren (Desoi 2018).

Stellt sich noch die Frage der **Zumutbarkeit**, also inwiefern der Eingriff für die Betroffenen (hier wären es wohl alle Personen, die in einem solchen überwachten Bereich verkehren) im Verhältnis steht mit dem öffentlichen Interesse oder auch dem schützenswerten Interesse Dritter. Problematisch ist insb. die allgemeine Überwachung und das Abgleichen von Gesichtern von Personen (die sich dessen nicht bewusst sind) (FRA 2019). Die örtliche Begrenzung der Überwachung bedeutet für Betroffene nicht automatisch einen leichteren Eingriff. Genauso wenig wird durch eine grosse Anzahl von Betroffenen der individuelle Grundrechtseingriff nicht generell schwerer gewichtet.[85] Hingegen stellt eine längere Aufbewahrungsdauer, aufgrund erhöhter Gefahr einer missbräuchlichen Verwendung der Daten, einen schwerer wiegenden Eingriff dar.[86] Für die Beurteilung der Zumutbarkeit ist nicht nur die Aufbewahrungsdauer des Überwachungsmaterials, sondern sind auch die Zugriffsberechtigungen und die getroffenen Massnahmen zum Schutz vor unsachgerechtem Zugriff und missbräuchlicher Verwendung der Daten zu berücksichtigen.[87]

Ebenfalls in die Verhältnismässigkeitsprüfung einzubeziehen ist die Örtlichkeit der Überwachung. Es muss unterschieden werden zwischen Sport oder kulturellen Events und (politischen) Demonstrationen, die für das Funktionieren der Demokratie eine wichtige Rolle einnehmen. Der *chilling effect* könnte die Meinungs- und Versammlungsfreiheit beeinträchtigen, da sich Personen vor möglichen negativen Konsequenzen fürchten. Die *Grundrechteagentur der EU* kommt daher zum Schluss, dass bei Demonstrationen kaum Situationen vorstellbar sind, in welchen der Einsatz von Gesichtserkennungstechnologien tatsächlich erforderlich und verhältnismässig sei (FRA 2019). Auch unzulässig wäre eine Echtzeitgesichtserkennung ohne konkrete Gefahr für wichtige Rechtsgüter.

Die Europäische Kommission schlägt insofern auch vor, die Echtzeitgesichtserkennung für Strafverfolgung auf öffentlich zugänglichen Plätzen grundsätzlich zu verbieten (European Commission 2021).

Für jeden Einsatz von zeitechten biometrischen Identifikationssystemen im öffentlichen Raum zur Strafverfolgung sieht der Vorschlag der Kommission eine vorgängige Autorisie-

---

[85] BGE 133 I 77 E. 5.3.

[86] Art. 13 Abs. 2 BV; BGE 133 I 77 E. 5.3. Siehe EGMR, 30562/04, S. and Marper v. the United Kingdom, 4.12.2008; 45245/15, Gaughran v. the United Kingdom, 13.2.2020; 24029/07, M.M. v. the United Kingdom, 13.11.2012; 53205/13, Trajkovski and Chipovski v. North Macedonia, 13.2.2020.

[87] BGE 133 I 77 E. 5.4.

rung durch ein Gericht oder eine unabhängige Behörde vor (European Commission 2021). Bei grosser Dringlichkeit kann diese Autorisierung im Nachhinein eingeholt werden. Das Gericht oder die zuständige Behörde hat diese nachträgliche Autorisierung zu erteilen, wenn sie aufgrund der objektiven Beweise und Hinweise zur Überzeugung gelangt, dass der Einsatz des Systems nötig und verhältnismässig war (European Commission 2021).

3.4.3.6. Datenschutzrechtliche Vorgaben

Neben den verfassungsrechtlichen Vorgaben müssten auch gesetzliche Vorgaben – insb. die Vorschriften des Datenschutzgesetzes – beachtet werden. Bei jeder Bearbeitung von Personendaten haben Bundesorgane und Private die Datenschutzgrundsätze (Rechtmässigkeit, Treu und Glauben, Verhältnismässigkeit, Transparenz, Zweckbindung, Datensicherheit und Datenrichtigkeit der Datenbearbeitung) einzuhalten. Die Grundsätze werden im Zusammenhang mit dem Einsatz der Gesichtserkennungssysteme durch Private im nächsten rechtlichen Kapitel erläutert. Eine etwas andere Auslegung erfordert die Transparenzpflicht bezüglich Datenbanken von Strafverfolgungsbehörden, da diese beschränkt werden kann, wenn dies notwendig ist, um nicht den Zweck der Strafverfolgung zu vereiteln (CoE 2021).

Beim Einsatz einer Gesichtserkennungssoftware, die der Identifikation von Personen dient, werden biometrische Daten und damit besonders schützenswerte Personendaten bearbeitet. Es kommen erhöhte Anforderungen an die Datenbearbeitung zum Zuge, insb. wird eine formell-gesetzliche Grundlage verlangt.[88] Es ist ein Verzeichnis der Bearbeitungstätigkeit zu führen, die Informationspflicht ist zu beachten und die Rechte der Betroffenen sind sicherzustellen.[89] Das revidierte DSG führt ausgeweitete Informationspflichten ein. Die Information der Betroffenen soll angemessen sein und alle Informationen enthalten, die nötig sind, um ihre Rechte durchzusetzen. Der alleinige Hinweis, dass Gesichtserkennung eingesetzt wird, würde diesem Erfordernis nicht genügen (Christen et al. 2020; Braun Binder et al. 2022, S. 57). Weiter wäre vor dem Einsatz eine Datenschutzfolgenabschätzung durchzuführen, sowohl aufgrund Art. 22 Abs. 2 lit. a (wegen der umfangreichen Bearbeitung von besonders schützenswerten Personendaten) als auch nach lit. b nDSG (systematisch umfangreiche Überwachung öffentlicher Bereiche). Das Datenschutzgesetz des Bundes ist auf kantonale Behörden nicht anwendbar. Die kantonalen Datenschutzgesetze sind aber weitgehend ähnlich, da diese – wie das Datenschutzgesetz des Bundes – verfassungs- und völkerrechtskonform zu sein haben.

Der Bundesrat kam in seiner Stellungnahme vom 11.8.2021 zur Interpellation Glättli (Regulierung der Gesichtserkennung im öffentlichen Raum) so auch zum Schluss, dass das neue DSG ausreichenden Schutz für die Bearbeitung von Daten mittels Gesichtserkennungstechnologie gewährleiste. Ein absolutes Verbot oder ein entsprechendes Moratorium auf Bundesebene sei nicht geplant (Bundesrat).

---

[88] Art. 34 Abs. 2 lit. a nDSG.

[89] Art. 12, Art. 19 und 25 nDSG.

3.4.4. Gesellschaftliche und ethische Herausforderungen

Kritik an Gesichtserkennung bezieht sich zumeist auf deren Einsatz durch polizeiliche Stellen. Dabei kann zwischen zwei Formen der Kritik unterschieden werden: die Kritik an heutigen Anwendungen der Technologie auf der einen und der grundsätzlichen Kritik an ihrer Missbrauchsanfälligkeit auf der anderen Seite.

Unter die erste Kategorie fällt insb. die Kritik an der immer noch **mangelnden Zuverlässigkeit der Gesichtserkennung**. Demnach minderten trotz technologischer Fortschritte hohe falsch-positiv und falsch-negativ-Erkennungsraten weiterhin die Praktikabilität der Gesichtserkennung. Ebenso würden schnelle Bewegungen, die Kopfhaltung und die Beleuchtungsverhältnisse die Ergebnisse verzerren. Hinzu komme eine noch weiter verminderte Zuverlässigkeit bei Minderheiten, so insb. bei Frauen dunkler Hautfarbe (Brühl 2020). Daneben wird auf Schwierigkeiten bei der Information und ggf. auch Einholung der Einwilligung der Betroffenen hingewiesen (EDPB, 2020, FRA 2020). **Betroffene würden Warn- bzw. Informationszeichen nicht wahrnehmen** und **könnten sich der Überwachung faktisch nur schwer entziehen**. Menschen, die mit den Massnahmen nicht einverstanden seien, würden zudem vor die Wahl gestellt, entweder die überwachten Bereiche zu meiden oder mit dem Gefühl des Überwachtwerdens und den ggf. damit einhergehende Konsequenzen leben zu müssen (Berle 2020). In diesem Zusammenhang wird immer wieder auch auf **IT-Sicherheitsrisiken bei der Datenspeicherung** verwiesen, die es Angreifende regelmässig möglich machen, Datensätze, die sensible biometrische Daten enthalten, zu entwenden (Behring 2019). Gegenüber der genannten Kritik folgt seitens der Befürworter der Technologien der Einwand, dass die genannten Punkte behebbare Fehler seien, die mit zunehmender Technologiereife und besserer administrativer Prozesse eine zunehmend geringere Rolle spielen würden (Charlesworth 2018). Nach Angaben der Einsatzbefürworter gab es in den vergangenen Jahren grosse Fortschritte, in deren Ergebnis z.B. Falscherkennungsraten drastisch zurückgegangen seien und der konkrete Einsatz durch zahlreiche Regeln und Vorschriften verantwortungsvoll gestaltet worden sei (McLaughlin und Castro 2020; Security Industry Association 2021). Zwar ist eine deutliche Verbesserung der Erkennungsraten zu erwarten. Allerdings werden Gesichtserkennungstechnologien bzw. darauf basierende Anwendungen auf absehbare Zeit dennoch nicht in der Lage sein, eine 100-prozentige Trefferrate zu erzielen. Gründe hierfür können eine mangelhafte Aufnahmequalität, schlechte Licht- und Wetterbedingungen und Bias in Datenbänken sein (Matzner 2016).

Seit dem Beginn der Diskurse rund um Gesichtserkennung in den 1990er-Jahren wird zudem darum gerungen, wie der Einsatz ausgestaltet sein sollte, um eine vertrauenswürdige und verantwortungsvolle Nutzung zu ermöglichen. Entsprechende Gestaltungsvorschläge wurden bereits früh seitens Datenschutzaufsichtsbehörden (Information Commissioner 2000), später auch der Zivilgesellschaft (ACLU 2001) und der Privatwirtschaft vorgelegt (Security Industry Association 2020; Microsoft 2018). Angesichts der in den letzten Jahren erfolgten massiven und intransparenten Zunahme des weltweiten Einsatzes der Gesichtserkennung zu polizeilichen Zwecken und der dabei immer wieder aufgetretenen Fälle von Diskriminierung, gingen Bündnisse weltweiter zivilgesellschaftlicher Organisationen schliesslich dazu über, ein Moratorium hins. des Technologieeinsatzes seitens polizeilicher Stellen und das Verbot biometrischer Massenüberwachung zu fordern (Scott 2020; EDRi 2020). Das Verbot soll v.a. den **anlasslosen Einsatz von Gesichtserkennungstechno-**

logie betreffen, bei der die Gesichter einer beliebigen Population etwa zum Auffinden von gesuchten Straftätern oder vermissten Personen dauerhaft und in Echtzeit erfasst werden. Andererseits betrifft diese Kritik den aus der Perspektive der Kritikerinnen und Kritiker **zu schnellen Ausbau von Gesichtserkennung**, womit unter Umgehung eines notwendigen demokratischen Diskurses bereits Tatsachen geschaffen würden (ebd.). Diese Forderungen wurden teilweise auch im politischen Raum aufgegriffen: So wurde der behördliche Einsatz von Gesichtserkennung Mitte 2019 seitens des Stadtrats von San Francisco verboten (Reuter 2019a). Die Europäische Kommission hat in ihrem Vorschlag zur Regulierung künstlicher Intelligenz vorgesehen, dass der Einsatz von KI-Systemen zur biometrischen Identifikation im öffentlichen Raum in Echtzeit zu Strafverfolgungszwecken grundsätzlich verboten werden sollte (Geminn 2021). Und auch der Europarat befürwortete in seinen «Guidelines on Facial Recognition» die Möglichkeit eines Moratoriums, solange keine demokratische Debatte über den Einsatz stattgefunden hat (CoE 2021). Von den genannten Kritikpunkten zu unterscheiden ist die grundsätzliche Kritik an der Missbrauchsanfälligkeit weitreichender Überwachungstechnologien wie der Gesichtserkennung. Dies betrifft insb. das Dammbruchargument, die attestierte zunehmende Machtdifferenz zwischen Staat und Bevölkerung sowie mögliche Abschreckungseffekte der Überwachung.

3.4.4.1. Per Dammbruch zur anlasslosen Dauer- und Massenüberwachung

Dem Dammbruchargument (häufig auch als «function creep» bezeichnet; Pounder 2008) zufolge werde bei der Einführung neuer Überwachungsmassnahmen anfänglich stets der Weg des geringsten Widerstands gewählt, um die öffentliche Kritik so gering wie möglich zu halten. Sobald sich die Gesellschaft an die neuen Überwachungsmassnahmen gewöhnt habe, würden deren Einsatzzwecke jedoch sukzessive ausgeweitet. Wenn etwa zunächst aus guten Gründen die Abwehr von Mord- oder Terrorfällen im Vordergrund stehe, finde über die Jahre eine Ausweitung des Anwendungsbereichs auf unbedeutendere Delikte wie Taschendiebstahl statt oder der Nutzerkreis mit Zugang zu den Datensätzen werde erweitert (z.B. von Sicherheits- auf Migrationsbehörden). Dadurch könne eine anlasslose Dauer- und Massenüberwachung der Bevölkerung sozusagen durch die Hintertür eingeführt werden (Castelluccia und Le Métayer 2020; Selinger und Leong 2021). Ebenso könnten die angewandten Technologien erweitert werden: Statt einer «einfachen» Personenidentifikation könnten Stimm-, Sprach- und Gesichtserkennungstechnologien bzw. Videokameras auch sukzessive zur Klassifizierung von Menschen verwendet werden, indem aus der Analyse von Bild- und Tondaten oder über das Verhalten bzw. den Gang Rückschlüsse auf Emotionen, die Persönlichkeit und damit letztlich die von einer Person vermeintlich ausgehende Gefahr gezogen werden.

3.4.4.2. Zunehmende Machtdifferenz zwischen Staat und Bevölkerung

Vertreter dieses Arguments verweisen darauf, dass mit fortschreitenden technologischen Möglichkeiten die Überwachungsfähigkeiten des Staates sukzessive ausgeweitet würden, ohne dass diese von ausreichenden demokratischen Transparenz- und Kontrollmassnahmen begleitet werden. Im Wesentlichen dreht sich das Argument darum, dass ohne

eine demokratische Kontrolle und öffentliche Berichterstattung über Überwachungsmassnahmen die Bevölkerung letztlich nicht mit Sicherheit davon ausgehen könne, dass die staatliche Macht sich nicht gegen die Menschen wendet (Roberts 2015). Weitgehende Überwachungsmassnahmen wie die Gesichtserkennung könnten demnach zu einem ungerechtfertigten und unüberschaubaren Machtgewinn des Staates führen, dessen Kontrollmacht die Wahrnehmung der Bürgerrechte erschweren und so die liberale Demokratie selbst gefährden könne (Reclaim Your Face 2022). Im Hinblick auf diese Diskussion ist die entscheidende Frage, welche Grenzen seitens der Öffentlichkeit dem Einsatz der Technologie in der Schweiz gesetzt werden und wie die Einhaltung dieser Grenzen beaufsichtigt wird.

3.4.4.3. Abschreckungseffekte der Überwachung

Abschreckungseffekte (*chilling effects*) der Überwachung bezeichnen Nebeneffekte staatlichen Handelns. Demnach würden weitreichende staatliche Sanktions- und Überwachungsmöglichkeiten nicht allein potenzielle Straftäter von der Ausübung derartiger Taten abhalten, sondern eine darüber hinausgehende, generell abschreckende Wirkung auf die Ausübung von Grundrechten entfalten. Bürgerinnen und Bürger würden so zu konformistischem Verhalten und zur Selbstzensur gedrängt, wodurch der freie demokratische Meinungsaustausch und letztlich die Grundfesten der liberalen Demokratie beschädigt würden (Schreiber und Joss 2020; Staben 2016).

3.4.5. Zwischenfazit

Gesichtserkennungstechnologie wird inzwischen auch in der Schweiz eingesetzt. Die entsprechenden Kantonspolizeien nutzen Gesichtserkennung dabei als ein Instrument, das in der Ermittlungstätigkeit verwendet wird. Eine Echtzeitgesichtserkennung findet nach aktuellem Wissensstand nicht statt. An dieser Stelle ist auch eine Diskrepanz zwischen Teilen der öffentlichen Debatte und der tatsächlichen Nutzung von Gesichtserkennungstechnologien festzustellen; denn insb. die zivilgesellschaftliche Diskussion zur Technologie hat die Massenüberwachung im öffentlichen Raum mittels Echtzeitgesichtserkennung zum zentralen Thema. Dies mag aus Motiven der Verhinderung einer sukzessiven Ausweitung der Gesichtserkennung womöglich taktisch sinnvoll sein, der vorliegende Bericht rückt jedoch neben der Diskussion der Echtzeitmassenüberwachung auch den Ex-post-Einsatz von Gesichtserkennung in den Vordergrund.

Die Untersuchung der Gesichtserkennung bei Kantonspolizeien zeigt sowohl Bemühungen hins. eines ethischen Einsatzes der Technologie als auch blinde Flecken. Positiv hervorzuheben sind zahlreiche behördeninterne Vorkehrungen, die einen ordnungsgemässen Betrieb sicherstellen sollen. Demgegenüber sind ein Mangel an öffentlicher Transparenz über den polizeilichen Einsatz, die Nicht-Information von Betroffenen und teils auch unzureichende Prozesse hins. der technischen Zuordnung von Datenzugriffen und Unkenntnis über Algorithmen, Bias und oder Ursprung der Trainingsdaten kritisch hervorzuheben.

Zur flächendeckenden Echtzeitüberwachung zeigt die rechtliche Analyse, dass diese die Kerngehalte von mehreren Grundrechten verletzt und in der Schweiz nicht erlaubt wäre. Für einen allenfalls örtlich und zeitlich begrenzten Einsatz der Echtzeitgesichtserkennungstechnologie müsste eine hinreichend bestimmte gesetzliche Grundlage im formellen Sinn geschaffen werden. Die Diskussion der gesellschaftlichen und ethischen Herausforderungen zeigt, dass es sich bei der aktuellen Kritik an der mangelnden Zuverlässigkeit der Gesichtserkennung, der unzureichenden Information von Betroffenen und IT-Risiken bei der Datenspeicherung um Herausforderungen handelt, die mit der Zeit, also mit besserer Technik und verbesserten organisatorischen Prozessen um weitestgehend lösbare Probleme handelt. Eine schwieriger zu adressierende Herausforderung wäre hingegen die Verhinderung einer allmählichen Ausweitung von einzelnen Überwachungsmassnahmen zu einer anlasslosen Dauer- und Massenüberwachung, die zunehmende Machtdifferenz zwischen Staat und Bevölkerung sowie Abschreckungseffekte der Überwachung.

3.5. Authentifizierung via Stimme bei Banken

Telefonbanking wird bereits seit 1984 eingesetzt (The Telegraph 2015). Seitdem bieten viele Banken in Europa diese Dienste ihren Kunden an, wobei die Hochzeit des Telefonbankings in den 2000er-Jahren vorbei ist (Clark 2007, S. 41). Stattdessen ist Onlinebanking inzwischen das meistgenutzte Verfahren in der Schweiz, das im Jahr 2019 von 68,4 % der über 15-Jährigen genutzt wurde (Bundesamt für Statistik 2019). Dennoch bietet das Telefonbanking auch heute noch die Möglichkeit, rund um die Uhr Bankgeschäfte durchzuführen. Dies wird meist durch einen Sprachcomputer ermöglicht, der die Kunden Schritt für Schritt durch das Banking-Angebot navigiert. Nur in bestimmten Fällen wird der Kunde zu einem Mitarbeiter weitergeleitet. Für die Verwendung muss sich ein Kunde vorab registrieren und erhält eine Zugangsnummer, ein Passwort (PIN) sowie häufig auch eine Transaktionsnummer-Liste (TAN-Liste) postalisch zugesandt. Mittels Zugangsdaten, wie Kundenummer und PIN, können sich Kunden dann am Telefon authentifizieren (Oettinger 2010). Dies birgt allerdings den Nachteil, dass die Entwendung des PINs und der TAN-Liste meist den vollen Zugriff auf alle Banktransaktionen ermöglicht (t-online 2012). Daher setzen einige Banken seit 2018 auf eine Authentifizierung per Stimme (Theunissen 2019, S. 159).

Ein Beispiel ist die PostFinance. Diese erkennt in der Authentifizierung per Stimme Vorteile im Hinblick auf eine erhöhte Sicherheit und Zeitersparnis (PostFinance 2021). Wenn ein Kunde den Service wünscht, wird von diesem zur Registrierung ein Stimmabdruck über das Telefon angefertigt und für spätere Authentifizierungen gespeichert. Ein Opt-Out ist laut Bandansage möglich, indem der Kunde bzw. die Kundin den Bankmitarbeitenden zu Beginn des Gesprächs darauf hinweist, dass er/sie keinen Stimmabdruck möchte. Zusätzlich ist es online im PostFinance «E-Finance» möglich, die Authentifizierung mit Stimmerkennung zu aktivieren oder zu deaktivieren.

Auch die Migros Bank nutzt Stimmbiometrie und verwendet dazu Stimmerkennungstechnologie der Firma Spitch. Die stimmbiometrische Authentifizierung ist aktuell allerdings nur für französischsprechende Kunden möglich. «Hochdeutsch, Schweizerdeutsch und Italienisch sollen im nächsten Schritt folgen» (IT Finanzmagazin 2020). Das Verfahren zur Authenti-

fizierung läuft hier völlig im Hintergrund ab, während der Kunde sein Anliegen vorträgt. Dies geschieht jedoch auch nur auf Wunsch des Kunden, sodass die Vorgaben des Schweizer Datenschutzbeauftragten erfüllt würden (IT Finanzmagazin 2020). Laut Spitch konnte bei der Bank dadurch die Gesprächsdauer um 20 % reduziert werden (Spitch 2019, S. 1). Bisher wird das System nur beim Telefonbanking eingesetzt. Jedoch bestehe die Möglichkeit, dies auch für virtuelle Assistenzsysteme oder als Unterstützung bei Beratungsgesprächen einzusetzen (IT Finanzmagazin 2020). Softwareanbieter von Stimmauthentifizierungssystemen bieten in zunehmendem Masse Module zur Erkennung der Emotionen der Anrufenden an (Stark und Hoey 2020). In Bezug auf die Migros Bank und die Postfinance ist jedoch nicht bekannt, ob Emotionserkennungssysteme eingesetzt werden.

3.5.1. Technische Grundlagen und Möglichkeiten

Das Verfahren ist bei beiden Banken grundsätzlich ähnlich. Als Erstes muss der Kunde für die Spracherkennung registriert werden. Zuvor wird er oder sie anhand einiger Fragen zur eigenen Person und Kontodaten befragt und somit authentifiziert (Jones 2018). Danach wird über das Telefon ein Stimmabdruck angefertigt, indem Merkmale der Stimme wie Sprechtempo, Frequenz und Lautstärke erfasst werden. Daten über den Inhalt werden nicht erfasst (PostFinance 2021).

Laut der PostFinance werden die Stimmabdrücke nur auf Servern in der PostFinance-Sicherheitszone in der Schweiz gespeichert und ausschliesslich zu Authentifikationszwecken verwendet (PostFinance 2021). Die Erkennung des Kunden erfolgt innert weniger Sekunden (Theunissen 2019, S. 159), noch bevor ein Bankmitarbeiter persönlich mit der Kundin spricht.

Bei der Migros Bank erfolgt die Authentifizierung durch die Software von Spitch parallel in dem Zeitraum, während der Kunde sein Anliegen vorträgt. Auch hier reichen bereits wenige Sekunden, um den Kunden zu erkennen (Spitch 2020). Erst danach gelangt er oder sie zu einem Mitarbeitenden der Bank (IT Finanzmagazin 2020). Bei der Migros-Bank ist «die gesamte Lösung […] auf bankeigenen Servern vor Ort installiert und kommt ohne Cloud-Komponenten aus» (IT Finanzmagazin 2020).

Grundsätzlich sei die Authentifizierung per Stimme sehr einfach und sicher per Telefon möglich. So argumentiert Theunissen (2019, S. 159), dass es für ein Telefon «dank der hohen Sicherheit der biometrischen Stimmerkennung keine sicherere Erkennung als die menschliche Stimme» gibt. Dass dies jedoch zweifelhaft ist, zeigte sich bspw. bereits 2012 durch Inthavisas und Lopresti (2012, S. 46), die einen Algorithmus zum Angriff auf sprachbiometrische Authentifizierungssysteme mittels Vorlagen (*templates*) präsentierten. Sie schlugen daher vor, zusätzlich ein Passwort zu nutzen. Beide Verfahren zusammen seien somit sicher gegen das willkürliche Ausprobieren eines Passworts (Brute Force) und demonstrierten ähnliche Fehlerraten (Falscheingaben der Nutzer) wie bei herkömmlichen passwortbasierten Systemen (Inthavisas und Lopresti 2012, S. 46). Generell gilt die Authentifizierung mittels Biometrie nur so lange als sicher, wie die biometrischen Zugangsdaten geschützt bleiben. Einmal kompromittiert, kann eine Person ihre biometrischen Daten, also Stimm- oder Gesichtsabdruck, nicht mehr verändern (Jones 2018). Bereits Tonschnipsel aus einem

Youtube-Video können reichen, um ein solches System zu infiltrieren (Softjourn 2021; Jones 2018). Mittels Generative Adversarial Networks, die zur Generierung von Deepfake-Audio verwendet werden, verstärkt sich dieses Problem. Zudem zeigte ein BBC-Reporter, dass sich sein Zwilling mittels der Sprachauthentifizierung der HSBC-Bank Zugang zum Bankkonto verschaffen konnte (Simmons 2017).

Bernd Martin, der Deutschland-Verantwortliche der Spitch AG, betont indes, dass für «die Stimmerkennung […] ein Mikrofon, wie es jedes Telefon, jedes Handy und jedes Smartphone hat[,] [genügt], während bei anderen Verfahren zusätzliches technisches Equipment benötigt wird bzw. nicht jedes Smartphone damit ausgestattet ist» (IT-I-Ko 2020). Offen bleibt jedoch, ob hierfür auch die Mikrofone älterer Smartphones oder Telefone ausreichen. Wie bereits erwähnt, ist die Daten-Input-Qualität ein entscheidender Faktor der Stimm- und Spracherkennung, die mit der Qualität der verwendeten Hardware und insb. des Mikrofons zusammenhängt. Zudem liegt der Frequenzbereich bei der analogen Telefonübertragung zwischen 300 Hz und 3,4 kHz und somit werden einige Frequenzbereiche der menschlichen Stimme «abgeschnitten» (Mathelitsch und Verovnik 2016, S. 83). Bei moderneren digitalen Übertragungen liegt der Wert bei etwa 8 kHz (Plassmann 2016, S. 1259). Hinzu kommt, dass das übertragene Frequenzspektrum je nach Auslastung des Funkmastes variiert und die übertragenen Frequenzbereiche ggf. nicht ausreichen könnten, um eine sichere Authentifizierung zu ermöglichen. Um eine Überlistung des Systems bspw. mittels Tonschnipseln zu Beginn eines Anrufs zu verhindern, erfolge die Analyse der Stimme bei Spitch dauerhaft während des Gesprächs (Spitch 2020).

3.5.1.1. Untersuchung der Stimmauthentifizierung durch die PostFinance

Zur Untersuchung des konkreten Einsatzes von Stimmauthentifizierungssoftware wurde Mitte 2021 Kontakt mit der PostFinance aufgenommen und um Beantwortung des Ethik-Fragekatalogs (vgl. Tabelle 20 im Anhang) gebeten.

3.5.1.1.1. Verantwortlichkeit und Rechenschaftspflicht

PostFinance speichere alle Sprachdaten in gehashter Form. Dies bedeutet, dass keinerlei Sprachdaten als Audiodatei gespeichert würden, sondern nur ein eindeutiges Abbild. Backups der Daten würden erstellt. Unklar ist der Speicherort der Daten und auch, ob ältere Sicherungskopien gelöscht werden.

Der Zugang zu den Daten sei durch ein Need-to-know-Prinzip beschränkt. Dies bedeutet, dass auch dann, wenn Personen durch ihre Freigabe Zugriff zu den Daten haben, auf diese nur dann zugreifen dürfen, wenn es hierfür einen betrieblichen Grund gibt. Zudem gab PostFinance an, dass es verschiedene Rollenkonzepte und Kontrollprozesse gäbe, um sicherzustellen, dass das Need-to-know-Prinzip auch korrekt umgesetzt wird.

Vor der Inbetriebnahme sei eine interne Sicherheitsbeurteilung durchgeführt und der Entscheid zur Einführung vom Management abgeholt worden. Öffentlich einsehbar ist diese Beurteilung nicht. Zudem führe PostFinance Sicherheits- und Penetrationstests in unregelmässigen Abständen durch. Offen bleibt, was unregelmässig bedeutet und in welcher Weise

die Sicherheits- und Penetrationstests durchgeführt werden. Aus IT-Security-Perspektive sinnvoll wäre solch ein Test mind. nach jeder Veränderung am System (Installation neuer Komponenten oder Updates).

3.5.1.1.2. Sicherheit

Die Sicherheit und Belastbarkeit des Systems würden von unabhängiger Stelle evaluiert. Patches zum Schliessen von Sicherheitslücken würden durch den Hersteller regelmässig geliefert und nach einer internen Prüfung bei PostFinance auch installiert. Die drei Schutzziele der Informationssicherheit (Vertraulichkeit, Integrität, Verfügbarkeit) sowie Schutz gegen Sabotage (Resilienz) und die Zurechenbarkeit von Zugriffen auf das System sind laut PostFinance gegeben. Durch die detaillierte Speicherung der Zugriffe (Logs) auf einem zentralen Logserver seien Zugriffe auf das System durch Mitarbeitende kontrollierbar (Verbindlichkeit). Angaben zur Überprüfbarkeit (Authentizität) machte PostFinance keine.

3.5.1.1.3. Transparenz und Erklärbarkeit

Bei der von PostFinance verwendeten Software handelt es sich um die weltweit vertriebene Real Time Authentication-Software des israelischen Unternehmens Nice (Cmm360 2018). Nach Angaben der PostFinance ist der verwendete Algorithmus ihnen nicht bekannt. Entsprechend sei auch nicht bekannt, ob der Algorithmus unabhängigen externen Tests unterzogen wurde und welche Resultate er dort ggf. liefert. Die genutzte Software werde mittels firmeneigener Sprachbeispiele trainiert, deren Inhalte ausschliesslich PostFinance bekannt sind. Die Zuverlässigkeit der Ergebnisse würden dann von einem unabhängigen Anbieter und PostFinance überprüft.

3.5.1.1.4. Gerechtigkeit, Fairness, Nicht-Diskriminierung und Bias-Vermeidung

PostFinance gibt an, dass die Trainingsdaten nicht auf potenzielle Bias untersucht würden, dafür aber der Algorithmus selbst (z.B. im Hinblick darauf, ob Menschen, die mit Dialekt sprechen, nicht so gut erkannt werden). Am Ende dieser Untersuchungen sei kein Bias festgestellt worden, weswegen auch nicht über ein potenzielles Bias kommuniziert werde. Hier stellt sich die Frage, wie der Algorithmus untersucht werden kann, wenn unklar ist, ob die Daten ein Bias enthalten. Bei der Gestaltung des Stimmauthentifizierungssystems habe PostFinance teilweise auf partizipative Elemente gesetzt, nannte jedoch nicht, inwiefern dies erfolgte.

3.5.1.1.5. Menschliche Kontrolle der Technik

Die Ergebnisse der Stimmerkennung würden laufend von Mitarbeitenden mittels Stichproben kontrolliert, um einen ordnungsgemässen Betrieb sicherzustellen.

3.5.1.2. Zusammenfassung

Laut Angaben der PostFinance werden umfangreiche und regelmässige Kontrollen durchgeführt, um den sicheren und zuverlässigen Betrieb der Stimmauthentifizierung zu gewährleisten. Die getroffenen Sicherheitsmassnahmen entsprechen damit gängigen IT-Sicherheitsstandards. Allerdings basiert diese Beurteilung lediglich auf den Angaben des Unternehmens, weil keine öffentlich einsehbare externe Begutachtung erfolgt. Marktstudien sprechen stimmbasierten Authentifizierungsverfahren, wie dem von Nice oder auch des Schweizer Unternehmens Spitch, eine hohe Zuverlässigkeit zu (Miller et al. 2020). Mangels unabhängiger Evaluierungen ist eine Verifizierung derartiger Angaben jedoch ebenfalls nicht möglich. Generell zeigt sich beim Thema der Stimmauthentifizierung, dass nur wenige Studien deren technische Zuverlässigkeit prüfen. Insb. das Fehlen eines grossangelegten Anbietervergleichs, wie die NIST-Studien zur Gesichtserkennung, erschwert eine externe Beurteilung der Technologie.

3.5.2. Juristische Bewertung

In diesem Kapitel wird zuerst auf die rechtlichen Aspekte der Authentifizierung von Bankkunden eingegangen.

Erhebt eine Bank bei Telefonaten mit Kunden einen Stimmabdruck, um diese zu authentifizieren, liegt eine Bearbeitung von Personendaten durch Private vor und es sind die entsprechenden Datenschutzvorschriften anzuwenden (Emmenegger und Reber 2019, S. 166).[90] Dies gilt sowohl für die Bank als Verantwortliche der Datenbearbeitung als auch für die allfällig dafür beauftragte Firma. Der Stimmabdruck zur Identifikation einer Person wird mithilfe eines spezifischen technischen Verfahrens gewonnen;[91] es handelt sich also um die Bearbeitung von biometrischen und damit besonders schützenswerten Personendaten.[92]

3.5.2.1. Die Datenschutzgrundsätze bei Authentifizierung via Stimme

Der Grundsatz von Treu und Glauben umfasst die *Transparenz* der einzelnen Datenbearbeitungsschritte (Rosenthal 2020).[93] Gemäss dem Grundsatz der Erkennbarkeit müssen die Erhebung der Daten und dessen Zweck für die betroffene Person ersichtlich sein. Dies ist beim Erstellen des Stimmabdruckes am Telefon oder einem allfälligen Abgleich mit einem bestehenden Stimmabdruck nicht per se der Fall. Deshalb sind betroffene Kunden über eine Datenbearbeitung immer ausdrücklich in Kenntnis zu setzen. Bei den untersuchten Anwendungsfällen geschieht dies durch eine automatische Ansage (PostFinance 2021) oder durch den Kundenberater (Migros Bank 2021). Die Beschaffung des Stimmabdruckes gilt dann als

---

[90] Art. 2 Abs. 1 lit. a nDSG. Auch die im obigen Beispiel genannte PostFinance ist eine privatrechtliche Aktiengesellschaft mit einer Mehrheitsbeteiligung der SchweizerPost AG. Sie erfüllt private Aufgaben und tritt nicht hoheitlich auf, womit sie als Privatperson zu qualifizieren ist.

[91] BBl 2017, 7020.

[92] Art. 5 lit. c nDSG.

[93] Art. 6 Abs. 3 nDSG.

erkennbar. Inwiefern weiter gehende Informationspflichten eingehalten werden, wird unten noch eingehender besprochen.

Das Prinzip der *Verhältnismässigkeit* gibt vor, dass die Datenbearbeitung mit denjenigen Mitteln erfolgen muss, welche im Hinblick auf den Bearbeitungszweck geeignet, erforderlich und zumutbar sind (EDÖB 2017). Die Stimmauthentifizierung scheint grundsätzlich ein geeignetes Instrument zu sein, es besteht jedoch – wie im technischen Teil ausgeführt (oben 3.5.1) – eine gewisse Fehleranfälligkeit. Als Vorteil, im Vergleich zu anderen Methoden (Prüfung der Identität anhand von Sicherheitsfragen bspw.), ist insb. die kürzere Dauer der Authentifizierung zu nennen, so konnte bspw. die erforderliche Gesprächsdauer in einer Bank um circa 20 % verringert werden.[94] Sowohl die Bank als auch der Kunde haben ein Interesse an der Effizienz eines Telefonats. Es besteht jedoch das Risiko der Verletzung der Datensicherheit und insb., dass schützenswerte Personendaten in die Hände von unbefugten Dritten gelangen könnten. Die Datenbearbeitung muss nicht nur geeignet, sondern auch objektiv notwendig für die Erreichung des gewünschten Zwecks sein. Die gewählte Massnahme hat im Hinblick auf das zu erreichende Ziel die mildeste Massnahme zu sein (Epiney und Nüesch 2015). Als mildere Massnahme könnte hier z.B. eine dezentrale Speicherung von biometrischen Erkennungsmerkmalen gelten, etwa die Speicherung auf dem Mobiltelefon des Kunden (Emmenegger 2019). Weiter ist zu prüfen, ob die Authentifizierung via Stimme für die Betroffenen hins. des Zwecks und der verwendeten Mittel zumutbar ist. Es stellt sich die Frage, ob ein angemessenes Verhältnis zwischen dem Bearbeitungszweck und der möglichen Beeinträchtigung der Persönlichkeit besteht (Epiney und Nüesch 2015). Dies ist abhängig von der Art der Umsetzung der Technologie und dem spezifischen Einsatz im Einzelfall, wobei die Zumutbarkeit hier zumindest fraglich erscheint.

Der Grundsatz der *Zweckbindung* schreibt vor, dass Personendaten nur für einen bestimmten und für die Betroffenen erkennbaren Zweck beschafft und nur auf eine mit diesem Zweck vereinbare Art und Weise bearbeitet werden dürfen.[95] Wenn Daten für den Stimmabdruck – zum Zweck der Authentifizierung der Kunden im Bankenwesen – erhoben werden, dürfen diese nur für diesen vorgegebenen Zweck verwendet und nicht etwa an Dritte bekannt gegeben oder für weitere Analysen – wie bspw. eine Auswertung der Stimmung der Kunden – genutzt werden (EDÖB 2017). Auf der Website der PostFinance wird aufgeführt, dass der Stimmabdruck ausschliesslich zu Authentifizierungszwecken verwendet werde (PostFinance 2021). Auch gemäss den Richtlinien der Migros Bank werden keine Merkmale zur Stimmung der Person oder ähnlichem erhoben. Ziel und Zweck der Stimmbiometrie sei nur die deutlich kürzere und vereinfachte Identifizierung (Migros Bank 2021).

Der Grundsatz der *Datenminimierung* verlangt, dass Personendaten vernichtet oder anonymisiert werden, sobald sie nicht mehr für den ursprünglichen Zweck benötigt werden.[96] Auch sollen nur erforderliche Informationen bearbeitet werden. Beim Zweck der Authentifizierung via Stimme ist nicht zusätzlich der Gesprächsinhalt zu speichern.[97] Daten sind zu löschen,

---

[94] Siehe oben technischer Teil.

[95] Art. 6 Abs. 3 nDSG.

[96] Art. 6 Abs. 4 nDSG.

[97] PostFinance speichert dementsprechend den Stimmabdruck auch nur in Form eines Codewertes, ohne den Gesprächsinhalt (PostFinance 2021).

sobald sie nicht mehr benötigt werden. Bei der Migros Bank erfolgt die Löschung des Stimmprofils drei Monate nach der Auflösung der Bankbeziehung automatisch. Die PostFinance macht dazu auf ihrer Website keine Angaben, die Notwendigkeit einer automatischen Löschung der Personendaten nach Beendigung der Bankbeziehung ergibt sich aber bereits aus dem Grundsatz der Datenminimierung (Migros Bank 2021). Bei Weiterbestehen der Bankbeziehung ist ein Widerruf bezüglich Stimmabdruck möglich und er wird auf Wunsch (gemäss Angaben der Banken) jederzeit gelöscht (Migros Bank 2021; PostFinance 2021).

Der Grundsatz der *Datenrichtigkeit*[98] verlangt, dass der Verantwortliche sicherstellt, dass die Person, welche ihre Stimme zum Abdruck gibt, auch mit derjenigen, die sie behauptet zu sein, übereinstimmt. Vor Erstellen des Stimmabdruckes wird dies in der Regel mittels Sicherheitsfragen geprüft.[99] Aber auch nach Abgabe des Stimmabdrucks könnte es zu einem Missbrauch, etwa im Rahmen von Deepfake-Audio kommen. Daher wird in der Literatur auch empfohlen, die Authentifizierung via Stimme immer mit anderen Überprüfungsfaktoren zu ergänzen (Emmenegger 2019, S. 182).

Gemäss dem Grundsatz der *Datensicherheit* sind die Vertraulichkeit, die Integrität und die Verfügbarkeit von Personendaten sicherzustellen.[100] Der Verantwortliche muss geeignete technische und organisatorische Massnahmen treffen, um eine dem Risiko angemessene Datensicherheit zu gewährleisten.[101] Die Daten sind insb. vor unerlaubtem Zugriff zu schützen (EDÖB 2017).

Positiv zu werten ist, dass die Daten in beiden besprochenen Fällen auf einem bankeigenen Server in der Schweiz gespeichert werden (Migros Bank 2021; PostFinance 2021). Die zentrale Speicherung sei hier gemäss EDÖB zulässig, da an der Verifikation ein starkes Interesse bestehe. Aufgrund der zentralen Speicherung sind die Anforderungen an die Datensicherheit jedoch auch erhöht (EDÖB 2017).

3.5.2.2. Persönlichkeitsverletzung nach Art. 28 ZGB

Gemäss der Rechtsprechung und Lehre gehört die Stimme in den geschützten Bereich der Persönlichkeit.[102] Dieser Schutz umfasst das Beschaffen, Weiterverbreiten oder Verfälschen der Stimme durch Dritte ohne spezifischen Rechtfertigungsgrund (Meili 2018). Da bereits das Aufnehmen einer individualisierbaren Stimme zum Schutzbereich der Persönlichkeit gehört, gilt dies umso mehr für den biometrischen Stimmabdruck (Emmenegger 2019, S. 171). Die Stimme ist ein eng und dauerhaft – von Geburt bis zum Tod – mit der Person verbundenes Merkmal (EDÖB 2016) und stellt einen einzigartigen Aspekt der Persönlichkeit dar. Im Unterschied zu einem Passwort kann dieses Merkmal bei unbefugter Entwendung nicht einfach neu erstellt werden. Deshalb handelt es sich bei diesem Eingriff

---

98 Art. 6 Abs. 5 DSG.

99 So auch bei der PostFinance (PostFinance 2021).

100 Art. 6 Abs. 5 nDSG.

101 Art. 8 nDSG. PostFinance ohne nähere Angaben (PostFinance 2021).

102 Art. 28 ZGB; BGE 110 II 411 E. 3b.

in den höchstpersönlichen Bereich um eine Persönlichkeitsverletzung – auch i.S.v. Art. 28 ZGB (Emmenegger 2019, S. 171, 183).

3.5.2.3. Zur Rechtfertigung der Datenbearbeitung

Es liegt eine widerrechtliche Persönlichkeitsverletzung vor, wenn die Verletzung nicht aufgrund einer Einwilligung der betroffenen Person, durch ein überwiegendes privates respektive öffentliches Interesse oder durch ein Gesetz gerechtfertigt ist.[103] Als möglicher Rechtfertigungsgrund ist insb. die Einwilligung zu prüfen. Damit diese gültig ist, muss sie nach angemessener Information freiwillig und auf einen bestimmten Zweck gerichtet erteilt werden.[104] Da bei der Authentifizierung via Stimme eine Bearbeitung von besonders schützenswerten Personendaten vorliegt, hat die Einwilligung ausdrücklich zu sein.[105] Für die Betroffenen hat die Tragweite (etwaige Risiken) ihrer Einwilligung nachvollziehbar zu sein. Die Information muss verständlich sein und vor der Erteilung der Einwilligung erfolgen (Emmenegger 2019; Belser et al. 2011). Rufen Kunden der PostFinance bei der Hotline an, hören sie die folgende telefonische Ansage: «Dieses Gespräch wird zu Sicherheits- und Wiedererkennungszwecken aufgezeichnet. PostFinance erstellt aus der Aufnahme einen Stimmabdruck, um Ihre Identität bei jedem Anruf anhand Ihrer Stimme zu verifizieren. Wünschen Sie keinen Stimmabdruck, bitten wir Sie, dies dem Kundenbetreuer mitzuteilen.» Widersprechen Schweizer Kunden nach der automatischen Ansage nicht, wird im Anschluss an das Gespräch ein Stimmabdruck gespeichert. Kunden mit Wohnsitz im Ausland werden nach der Ansage vom Kundenberater explizit gefragt, ob sie einen solchen Stimmabdruck wünschen (PostFinance 2021). Für die Kunden in der Schweiz besteht somit eine Opt-out- und für Kunden aus dem Ausland eine Opt-in-Lösung. Dies ist im Hinblick auf die Ausdrücklichkeit der Einwilligung von Bedeutung und wird deshalb unten im Text noch eingehender diskutiert. Die Kunden können zudem die Stimmerkennung auch online aktivieren oder deaktivieren (PostFinance 2021). Während gewisse Informationen wie der Bearbeitungszweck und die Art der Bearbeitung sowie der dafür Verantwortliche erwähnt werden, bleiben Risiken (ist ein solcher Stimmabdruck einmal erstellt, könnte er überall eingesetzt werden) unangesprochen (Emmenegger 2019, S. 176). Auf der Website finden sich zusätzliche Informationen, aber es scheint eher unwahrscheinlich, dass Personen die Website konsultieren, bevor oder während sie die Hotline anrufen (Emmenegger 2019, S. 177). Es bleibt daher fraglich, ob Betroffene transparent informiert werden.

Bei der Migros Bank werden die Kunden vom Kundenberater auf die Möglichkeit eines Stimmabdruckes aufmerksam gemacht und nur wenn der Kunde explizit zustimmt, wird dieser erstellt (Migros Bank 2021). Welche weiteren Informationen im Gespräch mit dem Kundenberater erteilt werden, ist unklar. Auf der Website lassen sich ebenfalls Informationen zur Stimmbiometrie finden, so werde diese einerseits anhand von physischen Merkmalen wie der Form des Kehlkopfes oder der Nasenpassage und anderseits aufgrund von Verhaltensmerkmalen wie Tonhöhe, Rhythmus oder Akzent bestimmt (Migros Bank 2021).

---

[103] Art. 31 Abs. 1 nDSG.

[104] Art. 6 Abs. 6 nDSG.

[105] Art. 6 Abs. 7 lit. a nDSG.

Die Freiwilligkeit einer Einwilligung liegt vor, wenn für Betroffene auch eine gangbare Alternative bereitsteht. Diese ist in beiden vorgestellten Szenarien vorhanden. Bei der PostFinance hat der Kunde die Möglichkeit, der Erstellung des Stimmabdrucks zu widersprechen, und er wird dann anhand von Sicherheitsfragen authentifiziert. Denkbar ist aber, dass sich Kunden in einer Überforderungssituation befinden, da sie möglicherweise bereits gestresst sind – etwa aufgrund eines Kreditkartenverlusts (Emmenegger 2019, S. 177–178) oder auch aufgrund der sehr beschränkten Zeit, sich über die Möglichkeit und die Auswirkungen des Stimmabdruckes Gedanken zu machen. Auch könnte die Überwindung, die Massnahme mit einem expliziten «Nein» abzulehnen, höher sein, als einfach zuzustimmen. Die Opt-in-Variante – auf welche die Migros Bank setzt und die bei der PostFinance für Personen mit Wohnsitz im Ausland zur Verfügung steht – wäre unter diesem Gesichtspunkt vorzuziehen. Betroffene können verneinen und dann darüber nachdenken, ob sie die Erstellung tatsächlich möchten.

Weiter wird die Ausdrücklichkeit einer Einwilligung verlangt. Eine konkludente Einwilligung – etwa Schweigen – erfüllt dieses Erfordernis nicht und ist als nicht gültige Einwilligung zu verstehen (Vasella 2015, S. 21–24; Emmenegger 2019, S. 178). Während bei der Migros Bank eine solche Ausdrücklichkeit vorliegt, mangelt es diesbezüglich beim Prozedere mit Schweizer PostFinance-Kunden, anders bei Kunden mit Wohnsitz im Ausland, bei welchen der Kundenberater noch nachfragt, ob man mit dem Stimmabdruck einverstanden ist. Nur in letzterem Fall ist das Erfordernis der ausdrücklichen Einwilligung gegeben (Emmenegger 2019, S. 178).

Für den Fall, dass keine Einwilligung vorliegt, wären auch die anderen Rechtfertigungsgründe zu prüfen. Für den Einsatz eines Stimmabdruckes für die Authentifizierung von Bankkunden besteht keine gesetzliche Grundlage.[106] Liegt keine spezialgesetzliche Regelung vor, erscheint der Rechtfertigungsgrund des überwiegenden öffentlichen Interessen meist als nicht gegeben, da dieser bei privaten Akteuren zurückhaltend zu bejahen ist (Emmenegger 2019, S. 174; Rampini 2014, N 47). Zu prüfen wäre noch das Vorliegen eines privaten Interesses. In Betracht fällt dabei nicht nur das Interesse der Bank, sondern auch das Interesse der Kunden.[107] Das Interesse an einer effizienten Kundenauthentifizierung erscheint als ein berechtigtes privates Interesse, da weniger Zeit benötigt wird als bei einer Identifikation mittels Sicherheitsfragen; auch weil daraus eine kürzere Gesprächsdauer resultiert. Wirtschaftliche Interessen genügen im Regelfall aber nicht für eine Rechtfertigung (Emmenegger 2019, S. 181; Meili 2018, N 49).[108] Hier wären insb. auch Alternativen denkbar, weshalb dem wirtschaftlichen Interesse kein massgebliches Gewicht zukommt.

Es hat sich gezeigt, dass die datenschutzrechtlichen Anforderungen durch die PostFinance nicht erfüllt werden. So fehlt es insb. an einer ausdrücklichen Einwilligung. Keine widerrechtliche Persönlichkeitsverletzung liegt bei der Migros Bank vor, hier erteilt der Kunde seine ausdrückliche Einwilligung (EDÖB 2017).

---

[106] Gemäss Art. 171^{quiniques} StGB ist die Stimmaufnahme bei Bestellungen, Aufträgen und Reservationen nicht strafbar. Die Erstellung eines Stimmabdruckes wird durch die Bestimmung jedoch nicht erfasst. Siehe Emmenegger (2019, S. 174).

[107] BGE 138 II 346 E. 10.3.

[108] BGE 138 II 346 E. 10.4 ff.

3.5.3. Gesellschaftliche und ethische Herausforderungen

Stimmerkennungssysteme zur Stimmauthentifizierung werden in der einschlägigen Literatur eher am Rande behandelt. Die mit ihr verknüpften Chancen und Herausforderungen entsprechen weitestgehend den bereits diskutierten Aspekten hins. des verantwortungsvollen und vertrauenswürdigen Einsatzes von KI-Systemen. Ein Aspekt, der heraussticht, betrifft die Möglichkeit der Nutzung biometrischer Stimmdaten zu weitergehenden Zwecken als der Authentifizierung. So können Unternehmen anhand von biometrischen Stimmdaten weiter gehende Informationen schlussfolgern, als den Betroffenen vielleicht bekannt ist und darüber Profile erstellen, die für Zwecke zielgerichteter Werbung verwendet werden können (Aichouni et al. 2019). Als ethisch besonders fragwürdig gilt hierbei die Emotionserkennung anhand von Stimme und Sprache (Stark und Hoey 2020).[109] Zudem wird weiterhin über die Frage gerungen, ob es sich bei biometrischen (Stimm-)Daten um eine sichere Authentifizierungsmethode handelt (Goode 2018; Access Now 2018). Schliesslich wird das Anwendungsfeld der Stimmauthentifizierung zur Erledigung von Bankgeschäften aufgrund der Angst vor einer Kompromittierung von Bankkonten als besonders sensibel wahrgenommen, wie der Widerstand bei der Einführung der Stimmerkennung in der Schweiz demonstriert hatte (SRF 2019).

3.5.4. Zwischenfazit

Die Untersuchung der Stimmauthentifizierung hat zweierlei gezeigt: Die Diskussion der technischen Grundlagen und Möglichkeiten betont, dass biometriebasierte Authentifizierungssysteme generell der Gefahr der Kompromittierung der biometrischen Daten ausgesetzt sind. Wie sich dies in der Praxis auf die Nutzung von Stimmauthentifizierung bei Banken auswirken mag, ist jedoch nicht vollständig beantwortet: Hersteller von Authentifizierungssoftware setzen vermehrt auf moderne KI-Sicherheitsverfahren, um die Imitation von Stimmen zu durchschauen und weiterhin die Sicherheit ihrer Systeme zu gewährleisten. Ob und inwiefern dies gelingt, wird Gegenstand weiterer Forschung sein. Die rechtliche Diskussion zeigte hingegen, dass die Bearbeitung von Personendaten zur Authentifikation eine ausdrückliche Einwilligung durch die Betroffenen erfordert. Weiter ist die angemessene Information der Betroffenen in der Praxis zu bezweifeln. Eine Debatte zu weitergehenden ethischen Herausforderungen der Stimmauthentifizierung ist bislang ausgeblieben.

3.6. Gewaltprävention und -aufklärung in Sportstadien

Das Thema Gewalt beim Sport ist seit vielen Jahren ein Dauerthema in der Öffentlichkeit. Darüber, wie genau Gewalt zu definieren und statistisch zu erheben ist, wird ebenso intensiv diskutiert wie über Möglichkeiten zur Eindämmung der unterschiedlichen Gewaltformen (Zick 2014). Den Sicherheitsbehörden stehen grundsätzlich die Möglichkeiten des Stadion- und Rayonverbots, Meldeauflagen, Polizeigewahrsam sowie Ausreisebeschränkungen und

[109] Vgl. diesbezüglich die Ausführungen zur Emotionserkennung in Abschnitt 3.8.3.

Massnahmen wie die räumliche Trennung von Fan-Gruppen zur Verfügung (Stadtpolizei Zürich 2021).

Im Zusammenhang mit dem Thema der Gesichtserkennung steht insb. die Herausforderung der Durchsetzung von Stadion-, teils auch Rayonverboten, im Zentrum der Debatte. Bis heute gilt deren Durchsetzung als schwierig, weshalb Stadionbetreiber und Sicherheitskräfte nach alternativen Lösungen suchen. Klassischerweise werden diese nämlich mittels manueller Massnahmen durchgesetzt: indem z.B. keine Tickets mehr an bekannte «Hooligans» und andere mit einem Stadionverbot belegte Personen verkauft werden oder indem das Sicherheitspersonal daraufhin trainiert wird, die Gesichter dieser Personen an den Eingängen zu erkennen und diese herauszuziehen. Beide Methoden werden allerdings hins. ihrer Wirksamkeit kritisiert: Den Blicken des Sicherheitspersonals können die Personen etwa durch das Ändern ihres Aussehens entgehen und alternative Wege, um trotz eines Verbots an Tickets zu gelangen, gibt es auch viele. Ein Dauerstreitthema hierbei sind die mit einer manuellen Personenerkennung verbundenen Personalkosten. In diesem Zusammenhang gilt die automatisierte Gesichtserkennung aufseiten der Befürworter als eine willkommene Methode sowohl zur effektiven als auch kostengünstigen Durchsetzung derartiger Verbote. Zusätzlich könnten moderne Gesichtserkennungssysteme auch zur Gewaltaufklärung genutzt werden, indem die Identität von Straftätern im Stadion ermittelt wird (Dave und Hume 2016).

3.6.1. Geschichte der Gesichtserkennung bei Grossveranstaltungen

Im Folgenden wird in die Anfänge der Debatten rund um Gesichtserkennung in Sportstadien weltweit und in der Schweiz eingeführt. Zur besseren Veranschaulichung der Betrachtung der technischen Grundlagen und Möglichkeiten wird anschliessend auf die Stadion-Gesichtserkennung der Schweizer Firma Deep Impact Bezug genommen. Schliesslich folgen die rechtliche Betrachtung und die Diskussion der gesellschaftlichen und ethischen Herausforderungen.

3.6.1.1. Die Anfänge der Gesichtserkennung bei Grossveranstaltungen

Die Idee, Grossveranstaltungen mittels Gesichtserkennung zu überwachen, reicht mehrere Jahrzehnte zurück. Das erste Videoüberwachungssystem, das Gesichtserkennung zum Zwecke der Überwachung von Grossveranstaltungen einsetzte, war das von der britischen Polizei in London/Newham eingesetzte System «Mandrake» des Anbieters Visionics. Dieses war im Jahr 2000 dahin gehend modifiziert worden, im System registrierte Hooligans während der Einlasskontrolle zu erkennen und so vom Betreten des Fussballstadions abzuhalten (Stern 2000). Über die Wirksamkeit dieses Anwendungsfalls wurde seitens der britischen Polizei zwar nicht Bericht erstattet, Presseberichte zur Effektivität von Mandrake legten jedoch die grundsätzliche Untauglichkeit des Systems, Gesichter effektiv zu erkennen, offen (1:n-Abgleich) (Meek 2002).

Anfang 2001 konnte Viisage, ein konkurrierender Anbieter, Gesichtserkennungssoftware beim Super Bowl in Tampa (Florida) einsetzen. Dabei wurden unter Rückgriff auf den Ei-

genface-Algorithmus, dessen Rechte sich Viisage zuvor gesichert hatte, die Gesichter von 70.000 Super-Bowl-Besuchern mit polizeilichen Fahndungslisten abgeglichen – ohne zuvor eine öffentliche Debatte über die Einführung zu führen, die Zuschauer vor Ort zu informieren oder deren Einwilligung einzuholen (Meyer 2020, S. 62). Die örtliche Polizei gab später bekannt, dass zwar 19 Personen vom System identifiziert worden waren, dies aber zu keinen Festnahmen geführt hatte. Nachprüfbare Informationen darüber, ob die angegebenen 19 Treffer korrekt waren oder ob es sich womöglich um falsch-positive Meldungen handelte, lagen auch in diesem Fall nicht vor (Canedy 2001). Die ACLU protestierte nach Bekanntwerden des Vorfalls dagegen, dass die Gesichtserkennung ohne die vorherige Information der Betroffenen erfolgt war (ACLU 2001).

3.6.1.2. Gesichtserkennung in Schweizer Stadien

Erste Debatten rund um die Videoüberwachung in Schweizer Stadien begannen ebenfalls um die Jahrtausendwende. Der damalige Datenschutzbeauftragte des Kantons Zürich, Bruno Baeriswyl, äusserte sich in dieser Zeit vorsichtig positiv gegenüber dem Einsatz der Videoüberwachung zur Verhinderung gewalttätiger Ausschreitungen, da ein öffentliches Interesse an der Verhinderung vorhanden sei. Als Vorbedingung zur Verhältnismässigkeit nannte er, dass zuvor weniger weitgehende Massnahmen zur Anwendung gekommen sein müssten und sich gezeigt haben müsste, dass diese unzureichend sind. Die Überwachung selbst müsste wiederum auf einer gesetzlichen Grundlage basieren und weitere Rahmenbedingungen einhalten, etwa die Information der Zuschauenden mittels schriftlicher Ankündigungen und über Lautsprecherdurchsagen sowie die Löschung des Bildmaterials bei Nichtgebrauch (Datenschutzbeauftragter des Kantons Zürich, S. 10).

Die Gesichtserkennung in Schweizer Stadien wurde erstmals Ende 2005 seitens des Schlittschuhclubs Bern (SCB) in der Bern-Arena getestet. Die nötige technische Infrastruktur für die Gesichtserkennung wurde von der Herstellerfirma Unisys bereitgestellt. Rund 100 SCB-Fans stellten ihre Gesichtsdaten freiwillig zur Verfügung. Der Testlauf beim SCB sollte zudem als Vorbereitung für den flächendeckenden Einsatz der Gesichtserkennung bei der Fussballeuropameisterschaft 2008 dienen, deren Belieferung der Anbieter Unisys anstrebte (Mäder 2005). Am Ende der mehrmonatigen Testphase verkündete der Firmenchef, dass es gelungen sei, 80 % der in der Datenbank enthaltenen Personen zu erkennen und dass die Fehlerquote bei unter einem Prozent gelegen habe (Bärtschi und Widmer 2006). Eine Verifikation der Ergebnisse von unabhängiger Seite fand allerdings nicht statt, sodass keine genaueren Aussagen über deren Gültigkeit möglich sind. Nach öffentlicher Kritik – u.a. waren der SCB und Unisys für den Schweizer Big Brother Award 2006 nominiert (Big Brother Awards 2006) – wurde schliesslich vom Einsatz des Systems während der Fussballeuropameisterschaft 2008 abgesehen (Handelszeitung 2008). Nachdem dieser erste Anlauf zur Einführung der Gesichtserkennung gescheitert war, wurde Mitte 2008 das Pilotprojekt «Sicherheit im Sport» vorgestellt. Vorgesehen war, dass personenbezogene Daten zwischen Vereinen, Stadionbetreibern und der Polizei ausgetauscht werden und Gesichtserkennung bei der Einlasskontrolle zum Einsatz kommen sollte (Venutti 2008a). Nachdem auch dieser Vorstoss auf teilweise erhebliche Kritik bei Vereinen, Stadionbetreibern, Staatsrechtlern und Fansozialarbeitern gestossen war, wurde die Gesichtserkennung aus dem Pilotprojekt ent-

fernt (Venutti 2008b). Stattdessen sollten die verstärkten Eingangskontrollen fortan durch mehr Personal erfolgen (NZZ 2009).

Später machte das Pilotprojekt «Focus One» der Swiss Football League (SFL) von sich reden. Dabei kam zwar keine Gesichtserkennung zum Einsatz. Mitte 2015 war jedoch bekannt geworden, dass eine private Firma, die XpertCenter AG (Bern), im Auftrag der SFL Besucher von Spielen seit März 2015 bei Auswärtsspielen auf den Fanmarschstrecken ausserhalb der Stadien per Video und Foto überwacht und daraus resultierende Dossiers an die Strafverfolgungsbehörden weitergeleitet hatte. Der Grund für das Filmen ausserhalb der Stadien war, dass sich die Fan-Gewalt zunehmend weg vom Stadion in die Stadt verlagert hatte. Der EDÖB war zwar vor Projektstart einbezogen worden, doch dessen Argumentation, wonach die private Überwachung auf öffentlichem Grund insb. gegen das Verhältnismässigkeits- und das Transparenzprinzip verstosse und daher nicht zulässig sei, war beim Start des Pilotprojekts unberücksichtigt geblieben (Burch und Rosch 2015). Etwa zeitgleich hatten auch Polizeistellen Pläne zur Kameraüberwachung potenzieller Gewalt-Hot-Spots ausserhalb der Stadien vorangetrieben. Allerdings setzte die Polizei weniger auf mobile Kamerateams und stattdessen auf fest installierte Kameras, da diese eine bessere Qualität lieferten und deutlich günstiger im Betrieb seien. Zudem war vorgesehen, dass Gesichtsaufnahmen der Fans aus verschiedenen Blickwinkeln gemacht werden, eine automatisierte Gesichtserkennung sollte jedoch nicht zum Einsatz kommen und die Kameras sollten nur während der Fanmärsche, also auf dem Hin- und Rückweg der Fans vom Stadion, in Betrieb sein (NZZ 2016; Grundrechte.ch 2015).

Abgesehen von diesen Fällen war ausserdem Ende 2017 bekannt geworden, dass das Sportamt der Stadt Zürich, das jahrelang Überwachungskameras ohne entsprechendes Reglement betrieben hatte, im neuen Reglemententwurf für elf Sport- und Badeanlagen vorgesehen hatte, Gesichtserkennung einzuführen. Nachdem dieser Vorstoss aufgrund gesellschaftlichen Widerstands scheiterte, gab das Zürcher Sportamt bekannt, dass mit Gesichtserkennung nicht das «elektronische Scannen von Gesichtern, [sondern] vielmehr eine Anforderung an die Bildqualität gemeint [war], die es erlaubt, von blossem Auge Gesichter zu erkennen» (Ledebur 2018).

Ein Blick auf die Angebote mehrerer Schweizer Gesichtserkennungstechnologieanbieter zeigt, dass der Anwendungsfall der Erkennung von Besuchern von Grossveranstaltungen in deren Angebot enthalten ist (Aptex AG 2021). Aus Sicht der Unternehmen mangele es v.a. an der gesellschaftlichen Akzeptanz, oder wie es Christian Fehrlin, Geschäftsführer der Firma Deep Impact, ausdrückt: «Weil die Unsicherheit im [sic] Bezug auf den Datenschutz so gross ist, wagt man es in Europa nicht einmal, mittels Gesichtserkennung Stadionverbote durchzusetzen» (Knupfer 2020b).

3.6.1.3. Neuere Entwicklungen über die Schweiz hinaus

Ganz anders sieht es aus, wenn die Entwicklung der Gesichtserkennung bei Grossveranstaltungen weltweit betrachtet wird. In verschiedenen Staaten hat deren Nutzung in den vergangenen Jahren erheblich zugenommen. Den Auftakt in Europa markierte das UEFA Champions-League-Finale in Cardiff (Grossbritannien) Ende Mai 2017 (Owen 2017). Dabei

waren die Gesichter von rund 170.000 Besuchern mit einer Polizeidatenbank bestehend aus 500.000 Gesichtern abgeglichen worden. Weil es sich um einen Testlauf handelte, hatte die Polizei zwar keine Verhaftungen auf Basis der Gesichtserkennung vorgenommen, doch die Software schnitt miserabel ab: Von 2470 Treffern waren 2297 falsch-positiv. Damit wies die Software eine Fehlerrate von 92 % auf. Unbekannt blieb, wie viele falsch-negative Treffer ausgegeben wurden (BBC News 2018). Zu ähnlichen Ergebnissen kam auch ein Bericht der Bürgerrechtsorganisation Big Brother Watch. Demnach lag die Trefferrate bei den Tests der Londoner Metropolitan Police in den Jahren 2016 und 2017 bei weniger als 2 % (Big Brother Watch 2018). Bei der South Wales Police, die auch für die Stadionüberwachung beim Champions-League-Finale verantwortlich zeichnete, lag die Fehlerrate bei allen 18 Tests der Gesichtserkennung zur Überwachung von Grossveranstaltungen zwischen Mai 2017 und März 2018 bei 91 %. 31 Menschen wurden auf Basis falsch-positiver Treffer angehalten und seitens der Polizeikräfte kontrolliert (ebd. 28–30). Zudem war bekannt geworden, dass auch die Gesichter der falsch-positiv erkannten Tausenden Personen für die Dauer von zwölf Monaten in die Datenbank der Polizei aufgenommen und nicht gelöscht wurden (ebd.). Unter den überwachten Veranstaltungen befand sich auch eine Demonstration vor einer Waffenmesse Ende März 2018. Deren Teilnehmer waren mittels automatisierter Gesichtserkennung erfasst worden und die daraus resultierenden falsch-positiven Treffer führten wiederum zu einer Aufnahme in die Polizeidatenbank (ebd. 31).

Auch die Schweizer Firma Deep Impact konnte sich bereits im Bereich der Gesichtserkennung in Sportstadien platzieren. Das Unternehmen bzw. dessen Spin-off Ava-X brachte die Software Sentinel beim Hochrisiko-Fussballderby zwischen Beşiktaş Istanbul und Fenerbahçe Istanbul zum Einsatz (Knupfer 2020a). Seit Beginn der Covid-19-Pandemie wurde das Gesichtserkennungssystem Sentinel zudem um eine Echtzeitfiebermessfunktion erweitert (Deep Impact 2021a).

Die Pandemie hat ohnehin den Anlass dazu geboten, verstärkt darüber zu diskutieren, wie verschiedene alltägliche Praktiken künftig berührungslos gestaltet werden könnten. Bereits beobachtbar ist bspw. eine Zunahme im Bereich der berührungslosen Zahlung in der Schweiz (Wernli 2020). Betroffen ist aber auch der Bereich der Kundenabwicklung bei (Gross-)Veranstaltungen. In verschiedenen Ländern laufen daher seit Beginn der Pandemie Vorbereitungen für den Umstieg auf kontaktlose Prozesse hins. der Zugangskontrollen und des «Bezahlens mit dem Gesicht», aber auch zur Kontrolle des Einhaltens von Mindestabständen (Olson 2020). Ehemals eingestellte Projekte zur Gesichtserkennung, etwa in der Amsterdamer Johan Cruijff Arena des Fussball-Vereins AFC Ajax, wurden seither neu belebt (Marti 2021).

Italien plant indes, einen Schritt weiter zu gehen und die Gesichtserkennung in Sportstadien mit Spracherkennungstechnologien zu verknüpfen. Diese soll zur Erkennung rassistischer Sprechchöre und der anschliessenden Identifikation der Sprechenden mittels Gesichtserkennung eingesetzt werden. Laut Medienberichten verzögerte sich der Einsatz nicht wegen der Menschenrechts- und Datenschutzbedenken, die z.B. seitens eines ehemaligen italienischen Datenschutzbeauftragten geäussert wurden, sondern wegen der Covid-19-Pandemie und der damit einhergehenden leeren Stadien (Chiusi 2020).

Bei der nun folgenden technischen Betrachtung dient der Anwendungsfall der Gesichtserkennung beim o.g. Fussballspiel zwischen Beşiktaş Istanbul und Fenerbahçe Istanbul

auf Grundlage der Software Sentinel als Beispiel, weil dieser Fall am ehesten dem in der Schweiz möglichen Nutzungsszenario der Gesichtserkennung in Sportstadien entspricht.

3.6.2. Technische Grundlagen und Möglichkeiten

Wie in Kapitel 2.1.1 beschrieben, besteht der Gesichtserkennungsprozess grundsätzlich aus den fünf Schritten: (1) Bilderfassung, (2) Gesichtsdetektion, (3) Merkmalsextraktion, (4) Datenbankabgleich und (5) Personenfeststellung (Kaur et al. 2020). Dies ist bei Gesichtserkennung in Stadien nicht anders. Es gibt jedoch Unterschiede bei der Bilderfassung. Im Falle von Sentinel wurden bei dem Fussballspiel der Istanbuler Mannschaften Bilder der Personen erst am Eingang des Stadions gemacht. Mittels 30 Kameras wurden die Gesichter aller 45.000 Zuschauer während des Einlasses eingescannt und in eine Datenbank aufgenommen (Deep Impact 2018), weil es zuvor keine Datenbank gab, in der Gesichtsfotos bekannter Hooligans gespeichert waren (Knupfer 2020c). Die Software könne mit externen Datenbanken, u.a. sogar direkt von Polizeibehörden, verbunden und dadurch mit weiteren Daten angereichert werden (Deep Impact 2021b).

Problematisch bei der Bilderfassung an Eingängen ist, dass manche Gesichter nur teilweise aufgenommen werden, von Sonnenbrillen und Schals verdeckt sind oder Kameras verschmutzt sein können (Gunther et al. 2017, S. 697). Anbieter von Gesichtserkennungssoftware nehmen sich derartiger Probleme in verstärktem Masse an und arbeiten an der Erkennung von geneigten bzw. teilverdeckten Gesichtern, indem bspw. aus mehreren Aufnahmen eines Gesichts das Beste für eine möglichst gute Erkennungsrate herausgefiltert bzw. zusammen Teilelemente zusammengeführt werden (Panasonic 2021).

Mit der Software Sentinel können einzelne Gesichter isoliert werden. Operatoren können ein beliebiges Gesicht aus dem Publikum auswählen, woraufhin die Software alle ähnlichen Gesichter in der Reihenfolge der Erkennungsgenauigkeit anzeigt. Über einen weiteren Regler kann die Erkennungsschwelle angepasst werden, sodass nur die über dem gewählten Schwellenwert liegenden Gesichter angezeigt werden. Nach Auswahl dieser Gesichter zeigt das System alle vorhandenen Videoaufnahmen der Person (Deep Impact 2019).

Laut Aussagen des Firmengründers benötige die Erkennung 0,3 Sekunden und sei bei dem Test für das Fussballspiel in der Lage gewesen, die gesuchten Gesichter mit einer Korrektheit von 99,8 % finden (Knupfer 2020c). Dabei wird keine spezielle Hardware benötigt, eine beliebige HD-Kamera, wie sie in jüngeren Videoüberwachungssystemen zum Einsatz kommt, reiche aus (Deep Impact 2021b). Eine unabhängige Evaluation (etwa im Rahmen des NIST-Vergleichs) der Ergebnisse liegt nicht vor, sodass keine Beurteilung der Trefferrate möglich ist.

3.6.3. Juristische Bewertung

Das Datenschutzrecht ergänzt und konkretisiert den durch das Zivilgesetzbuch in Artikel 28 verankerten Schutz der Persönlichkeit. Das Recht am eigenen Bild wird durch Art. 28 ZGB erfasst. Die Anfertigung eines Bildes mit einer erkennbaren und individualisierten Person stellt eine Persönlichkeitsverletzung dar (Keist 2019). Persönlichkeitsverletzungen sind wi-

derrechtlich, wenn kein Rechtfertigungsgrund vorliegt (siehe unten). Zweck des Einsatzes des Gesichtserkennungssystems durch den Stadionbetreiber besteht in der Identifizierung von bestimmten Personen. Die Gesichtsbilder, die in einem solchen spezifischen technischen Verfahren bearbeitet werden, sind als biometrische Daten zu qualifizieren. Damit liegt hier eine Bearbeitung von besonders schützenswerten Personendaten vor.[110]

3.6.3.1. Datenschutzrechtliche Grundsätze

Bei der Bearbeitung von Daten müssen Private die Datenschutzgrundsätze einhalten. Aus dem Grundsatz von Treu und Glauben kann auch der Grundsatz der Transparenz jeglicher Datenbearbeitungsschritte abgleitet werden (Rosenthal 2020). Der Grundsatz der Erkennbarkeit der Datenbeschaffung ist damit eng verbunden.

Eine Besonderheit der Gesichtserkennungstechnologie besteht darin, dass sie ohne Beteiligung und ohne Wissen der Betroffenen zur Erhebung biometrischer Daten führen kann. Der Europarat betont denn auch in der Auslegung des – für die Schweiz verbindlichen – Art. 8 der Konvention 108+ die Wichtigkeit einer transparenten Datenbearbeitung beim Einsatz einer solchen Technologie. Der Verantwortliche (hier der Stadionbetreiber) hat gemäss Art. 8 der Konvention 108+ alle nötigen Informationen über die Datenbearbeitung bekannt zu geben. Gemäss der Richtlinie zur Gesichtserkennung des Europarats sind für die Bestimmung der Transparenz folgende Punkte von Bedeutung (CoE 2021):

- Information der Betroffenen (FRA 2019; EDÖB 2009).[111]
- Bekanntgabe des Zwecks der Datensammlung.
- Informationen über die künftige Verwendung der Daten.
- Erläuterungen über die Auswirkungen der Erhebung, Verwendung und Weitergabe der Daten.
- Angaben zu bestehenden Rechten und Rechtsbehelfen.
- Mitteilung, ob und in welchem Umfang die Gesichtserkennungsdaten an Dritte übermittelt, werden können. (Falls ja, Informationen über den Vertragspartner)
- Informationen zur Aufbewahrung, Löschung oder Deidentifizierung der Daten.
- Angaben zu Kontaktstellen, an welche sich die Individuen mit Fragen zur Erhebung, Nutzung und Weitergabe der Gesichtserkennungsdaten wenden können.
- Im Falle einer Änderung der Erhebungs-, Nutzungs- oder Weitergabepraktiken muss diese durch die Unternehmen in Form der Aktualisierung ihrer Datenschutzrichtlinien oder anderweitig veröffentlicht werden (CoE 2021).
- Ob der Grundsatz der Verhältnismässigkeit bei der Bearbeitung von Personendaten eingehalten ist, kann nicht generell beantwortet werden, sondern ist im Einzelfall zu prüfen (Passadelis 2015). Es wird verlangt, dass die Datenbearbeitung für die Erreichung des verfolgten Zwecks geeignet ist. Die Eignung wird bestimmt durch die Genauigkeit, mit

[110] Art. 5 lit. c nDSG.

[111] Art. 19 Abs. 1 nDSG.

welcher das mit der Datenbearbeitung angestrebte Ziel erreicht wird (Passadelis 2015). Für die Erkennung von bereits gewalttätig gewordenen Besuchern von Sportveranstaltungen scheint das System geeignet. Auch Videoaufnahmen, die nachträglich ausgewertet werden, um allfällige Störer zu identifizieren, scheinen geeignet.

Es scheinen jedoch mildere Mittel als der Einsatz der Gesichtserkennungstechnologie denkbar. Insb. die Zugangskontrolle könnte auch mithilfe von personalisierten Tickets geschehen, zusammen mit einer Passkontrolle. Auch der Einsatz von spezialisierten Personen, welche darauf trainiert sind, die betreffenden Personen zu sichten, erscheint weniger eingreifend. Bei Ausschreitungen während des Spiels sollte es mit genügend Personal möglich sein, die Personen an Ort und Stelle zu identifizieren, womit eine nachträgliche Identifikation überflüssig erscheint. Möglich wäre auch eine Ausgestaltung mit Möglichkeit der Wahl zwischen einem personalisierten Ticket oder dem Betreten des Stadionbereichs mit einer Zugangskontrolle mit Gesichtserkennung. Die Erforderlichkeit des Einsatzes eines Gesichtserkennungssystems scheint deshalb zumindest zweifelhaft.

Weiter hat die Datenbearbeitung im Blick auf den Zweck und auf den Eingriff in die Grundrechte der Betroffenen zumutbar zu sein (Passadelis 2015). Die Zumutbarkeit eines solchen Systems ist fraglich. Gemäss den Richtlinien des Europarats dürfen Private Gesichtserkennungstechnologien in Umgebungen wie Einkaufszentren nicht einsetzen, wenn sie damit eine Identifizierung einer Person zu Marketingzwecken oder zu privaten Sicherheitszwecken verwenden möchten (CoE 2021).

Gemäss dem Grundsatz der Zweckbindung dürfen Personendaten nur für einen bestimmten und für die Betroffenen erkennbaren Zweck beschafft werden. Der Zweck muss ausdrücklich, spezifisch und rechtmässig sein (CoE 2021). Die Daten dürfen nur in der Weise bearbeitet werden, die mit dem Zweck vereinbar sind (Rosenthal 2020). Die biometrischen Daten dürften demnach im Falle eines Einsatzes zur Überwachung der Stadionsicherheit nicht etwa für personenspezifische Werbung oder Ähnliches verwendet werden (Desoi 2018).

Der Grundsatz der Datenminimierung schreibt vor, dass Daten zu vernichten oder zu anonymisieren sind, sobald sie zum Zweck der Bearbeitung nicht mehr benötigt werden.[112] Nur die erforderlichen Informationen sollen bearbeitet werden. So muss bspw. geprüft werden, ob eine Datenbearbeitung möglich ist, ohne die Daten zu speichern (Desoi 2018). Weiter hat der Verantwortliche eine Aufbewahrungsdauer festzulegen (welche nicht länger sein kann, als es für den spezifischen vorgesehenen Datenbearbeitungszweck erforderlich ist) und zudem sind die biometrischen Daten zu löschen, sobald der Zweck erreicht ist (CoE 2021; EDÖB 2022).

Beim Einsatz von Echtzeitgesichtserkennung hat der Verantwortliche gemäss den Richtlinien des Europarats unterschiedliche Aufbewahrungsdauern festzulegen:

- Liegt kein Treffer vor, müssen die Daten automatisch und sofort gelöscht werden.
- Liegt ein Treffer vor, können biometrischen Daten für eine festgeschriebene und beschränkte Dauer gespeichert werden.

[112] Art. 6 Abs. 4 nDSG.

- In jedem Fall müssen die Watchlist und die biometrischen Daten gelöscht werden, sobald der Zweck abgeschlossen ist (CoE 2021).

Der Verantwortliche muss gemäss dem Grundsatz der Datenrichtigkeit alle angemessenen Massnahmen treffen, um sich der Korrektheit der Personendaten zu vergewissern (Rosenthal 2020). Betreffend den Einsatz einer Gesichtserkennungssoftware ist insb. zu überprüfen, ob die bearbeiteten Daten korrekt sind. Die Datenrichtigkeit hat nicht absolut zu sein, sondern muss in einem angemessenen Verhältnis zum Bearbeitungszweck stehen. Die Technologie sollte eine hohe Wahrscheinlichkeit der korrekten Identifizierung erlauben (Braun Binder et al. 2022, S. 58). Ebenfalls ist während des Einsatzes der Software der Prozentsatz der Erkennungssicherheit zu überwachen und bei Bedarf ist ein erneutes Training oder entsprechende Anpassungen vorzunehmen (CoE 2021).

Datenqualität und biometrische Templates, die in der Watchlist aufgeführt werden, müssen zur Verringerung der Wahrscheinlichkeit von falschen Treffen kontrolliert werden. Auch ist sicherzustellen, dass die Bilder in der Watchlist datenschutzrechtskonform erhoben wurden. Falls es falsche Treffer gibt, muss der Verantwortliche alle angemessenen Massnahmen ergreifen, um dies zukünftig zu vermeiden und die Datenrichtigkeit sicherzustellen (CoE 2021).

Der Grundsatz der Datensicherheit verlangt, dass die Vertraulichkeit, Integrität und Verfügbarkeit von Personendaten sichergestellt wird (Rosenthal 2020). Wird die Datensicherheit nicht gewährleistet, kann es zu schwerwiegenden Konsequenzen für die Betroffenen kommen, etwa durch eine unautorisierte Bekanntgabe von sensiblen Daten an Dritte (CoE 2021). Dabei wird nicht verlangt, dass die Massnahmen einen absoluten Schutz bieten, vielmehr soll ein vernünftiges Verhältnis zwischen dem Risiko einer Verletzung und der Datensicherheit gefunden werden (Rosenthal 2020). Sicherheitsmassnahmen sollen angesichts neuer Gefahren laufend überprüft und angepasst werden (CoE 2021).

Liegt eine Persönlichkeitsverletzung vor, ist zu prüfen, ob diese durch die Einwilligung der betroffenen Person oder durch ein überwiegendes privates oder öffentliches Interesse oder aufgrund des Gesetzes gerechtfertigt werden kann.[113]

Der Rechtfertigungsgrund der Einwilligung ist nur gültig, wenn diese für eine bestimmte Bearbeitung und nach angemessener Information erfolgt sowie freiwillig erteilt wird (CoE 2021). Eine solche Einwilligung wäre für den Einsatz von Gesichtserkennungssoftware durch Private nötig; da besonders schützenswerte Personendaten bearbeitet werden, hat sie zudem ausdrücklich zu erfolgen.[114] Der Aufenthalt in einem durch Gesichtserkennungssoftware überwachten Ort kann nicht als eine ausdrückliche Einwilligung angesehen werden (CoE 2021). Die Freiwilligkeit der Einwilligung ist von hoher Bedeutung. Der Europarat empfiehlt denn auch, dass den Betroffenen eine alternative Lösung zur Gesichtserkennung anzubieten sei. Die Zugangskontrolle zum Stadion könnte auch mittels Ausweises durchgeführt werden. Es muss sich dabei um eine tatsächliche Alternative für die Betroffenen handeln, sie darf nicht viel länger dauern oder komplizierter sein (CoE 2021; Keist 2019).

[113] Art. 31 Abs. 1 nDSG.

[114] Art. 6 Abs. 7 lit. a nDSG.

Ein weiterer Rechtfertigungsgrund für den Einsatz der Gesichtserkennungstechnologie in Stadien wäre eine gesetzliche Grundlage. Die Bekämpfung von Gewalttätigkeiten anlässlich von Sportveranstaltungen wird in Art. 24 a des Bundesgesetzes über Massnahmen zur Wahrung der inneren Sicherheit (BWIS) aufgeführt. Es besteht ein elektronisches Informationssystem mit den Daten von Personen, die sich anlässlich von Sportveranstaltungen gewalttätig verhalten haben.[115] Organisatoren von Sportveranstaltungen können diese Daten erhalten, wenn dies für die Vorbeugung von Gewalttätigkeit anlässlich bestimmter Veranstaltungen nötig ist.[116] Nicht beschrieben wird jedoch der Einsatz einer Gesichtserkennungssoftware zur Erkennung dieser Personen. Es liegt somit keine gesetzliche Grundlage vor.

Zu prüfen wäre noch, ob allenfalls ein überwiegendes privates oder öffentliches Interesse am Einsatz der Technologie bestünde. Als öffentliches und auch privates Interesse kann hier die Sicherheit vorgebracht werden. Immer wieder finden gewalttätige Ausschreitungen in Zusammenhang mit Sportanlässen statt. Diese können für andere Zuschauer gefährlich werden. Auch der Stadionbetreiber hat ein hohes Interesse an der sicheren Durchführung der Sportveranstaltung, da u.U. Massnahmen wie etwa «Geisterspiele» getroffen werden können, die seine Einnahmen stark verringern. Der EDÖB bejaht das Vorliegen eines überwiegenden privaten oder öffentlichen Interesses an der Gesichtserkennung bei Personenkontrollen durch private Stadionbetreiber unter der Voraussetzung, dass die allgemeinen Datenschutzprinzipien eingehalten werden (EDÖB 2009; Keist 2019).

Der Einsatz eines Gesichtserkennungssystems als Zulassungsbeschränkung oder innerhalb des Sportstadions erfordert eine Datenschutzfolgenabschätzung, da aufgrund der umfangreichen Bearbeitung von besonders schützenswerten Personendaten ein erhöhtes Risiko besteht (CoE 2021).

3.6.4. Gesellschaftliche und ethische Herausforderungen

Seit den ersten Debatten über Videoüberwachungs- bzw. Gesichtserkennungstechnologien in den 1990er-Jahren stand insb. die Fehleranfälligkeit der Technologie in der Kritik (Canedy 2001). Ebenso wurde bemängelt, dass die Technologie ohne vorherige öffentliche Debatte, Information der Betroffenen und ohne deren Einwilligung eingesetzt wurde (Meyer 2020, S. 62).

In den darauffolgenden Jahren wurden diese Punkte angesichts der weltweiten Ausbreitung von Überwachungstechnologien im Nachgang der Terroranschläge vom 11. September 2001 auch auf den generellen Ausbau von Überwachungsmassnahmen und Videoüberwachung im Kontext von Sportveranstaltungen bezogen (Schimmel 2011; Eick 2011). Der Ausbau derartiger Überwachungsmassnahmen würde eine Hierarchisierung und Versicherheitlichung des urbanen Raums vornehmen (Aas et al. 2009, S. 65–66) und – gerade dann, wenn die Technologien korrekt funktionierten – Grundrechte und Freiheiten einschränken. Auf diese Weise würde die Pflicht, gesetzeskonform zu handeln, sukzessive vom Furcht-Imperativ einer allumfassenden Überwachung und Rechtsdurchsetzung ersetzt (Hale 2005, S. 151–152). In der Kritik steht hier weniger die Abwendung grösserer Gefahren, sondern die

[115] Art. 24 a Abs. 1 und 2 BWIS.

[116] Art. 24 a Abs. 8 BWIS.

Befürchtung, dass Crowd-Control-Massnahmen, wie die automatisierte Gesichtserkennung in Sportstadien, Teil einer immer mehr um sich greifenden Disziplinierung und staatlichen Steuerung der Bürgerinnen und Bürger würde und **das gegenwärtige Machtgefüge in Richtung privater und staatlicher Akteure verschiebe** (Hutchins und Andrejevcic 2021). Nicht nur in diesem Punkt sind Überlappungen zwischen den ethischen Herausforderungen der Gesichtserkennung in Sportstadien und jenen der polizeilichen Gesichtserkennung zu erkennen. Überschneidungen sind auch vorhanden im Hinblick auf die **Gefahr der Massenüberwachung, des Missbrauchs der Daten zur Diskriminierung politischer Gegner, mögliche Verdrängungs- und Abschreckungseffekte, IT-Sicherheitsrisiken, Probleme hins. der Zuverlässigkeit der Technologie** und **die Befürchtung, dass die Anwendungszwecke nach einer Einführung zunehmend ausgeweitet würden** (Samatas 2014, S. 115–118) – die jedoch unter Verweis auf das Kapitel zur polizeilichen Gesichtserkennung (vgl. 3.4.4) an dieser Stelle nicht wiederholt werden.

Ansonsten bezieht sich die Kritik (BdWi 2006) an Crowd-Control-Massnahmen zur Stadionüberwachung v.a. auf die datenschutzrechtlichen Implikationen, wie etwa die Erforderlichkeit und Verhältnismässigkeit, die im vorangegangenen rechtlichen Abschnitt (3.6.3) erörtert wurden.

3.6.5. Zwischenfazit

Dieser kursorische Überblick der Gesichtserkennung bei Grossveranstaltungen in der Schweiz zeigt, dass die Technologie bereits getestet wurde und seit Längerem im Gespräch ist, es aber bislang zu keiner Anwendung gekommen ist. Die Kapiteleinführung und die Betrachtung der technischen Grundlagen und Möglichkeiten haben verdeutlicht, dass auch diese Anwendungsform der Gesichtserkennung über Jahrzehnte mit unzuverlässigen Trefferraten konfrontiert war. Die rechtliche Untersuchung zeigt, dass zurzeit keine gesetzliche Grundlage für den Einsatz von Gesichtserkennungstechnologie durch Stadionbetreiber besteht. Allenfalls kann aber ein überwiegendes privates oder öffentliches Interesse vorliegen, ansonsten stellt die Einwilligung der Betroffenen den einzigen Rechtfertigungsgrund dar. Der Europarat empfiehlt für die Gewährleistung der Freiwilligkeit der Einwilligung, den Betroffenen einen alternativen Eintritt ohne Gesichtserkennung zu ermöglichen. Des Weiteren wäre der Verantwortliche in der Pflicht, die Betroffenen auf angemessene Weise über die Bearbeitung zu informieren und zahlreiche weitere Vorgaben einzuhalten.

In gesellschaftlicher und ethischer Hinsicht steht neben der Kritik an der mangelnden technischen Zuverlässigkeit v.a. die Befürchtung zur Debatte, dass die Einführung von Gesichtserkennung in Sportstadien nur ein erster Schritt in Richtung einer allgegenwärtigen Massenüberwachung aller Lebensbereiche seitens privater und öffentlicher Stellen sei.

3.7. Erkennung physischer und psychischer Krankheiten

Künstliche Intelligenz hat ihren Weg in das Schweizer Gesundheitswesen bereits gefunden (Perani 2018). Beispielsweise sollen Gesundheitssysteme weltweit durch den Einsatz von Stimm- und Spracherkennung effizienter, günstiger werden und damit insgesamt angenehmer für Patientinnen und Patienten werden, da Stimm- und Spracherkennung als schnell und

nicht invasiv gelten (Latif et al. 2020, S. 348). In einer Studie aus Deutschland aus dem Jahr 2017 gaben 64 % der Befragten im Gesundheitswesen an, dass KI das Gesundheitssystem grundlegend verändern werde, und 30 % der deutschen CEOs im Gesundheitswesen gaben an, bereits KI-Produkte einzusetzen (Burkhart und Huesman-Koecke 2021). Während Spracherkennung, insb. die Diktierfunktion, schon seit den 2000er-Jahren im Bereich der Medizindokumentation in Krankenhäusern eingesetzt wird (Zuchowski et al. 2020), können moderne Systeme inzwischen bereits selbstständig Krankheitsbilder anhand der Stimme von Patienten erkennen. Stimm- und Spracherkennung können zudem bei der Behandlung von Patienten helfen. So könnten Spracherkennungssysteme die Kommunikation zwischen Ärzten und Patienten, welche sich aufgrund ihrer Krankheit nicht korrekt oder verständlich ausdrücken können, erleichtern (Saz et al. 2009; Selouani et al. 2009). So beschreibt Nayar (2017) Sprachinterfaces für Kinder bei der Sprachtherapie. Darüber hinaus gibt es eine Vielzahl von möglichen Anwendungsfeldern, die überblicksartig für die Spracherkennung im Gesundheitswesen der Abbildung 8 entnommen werden können.

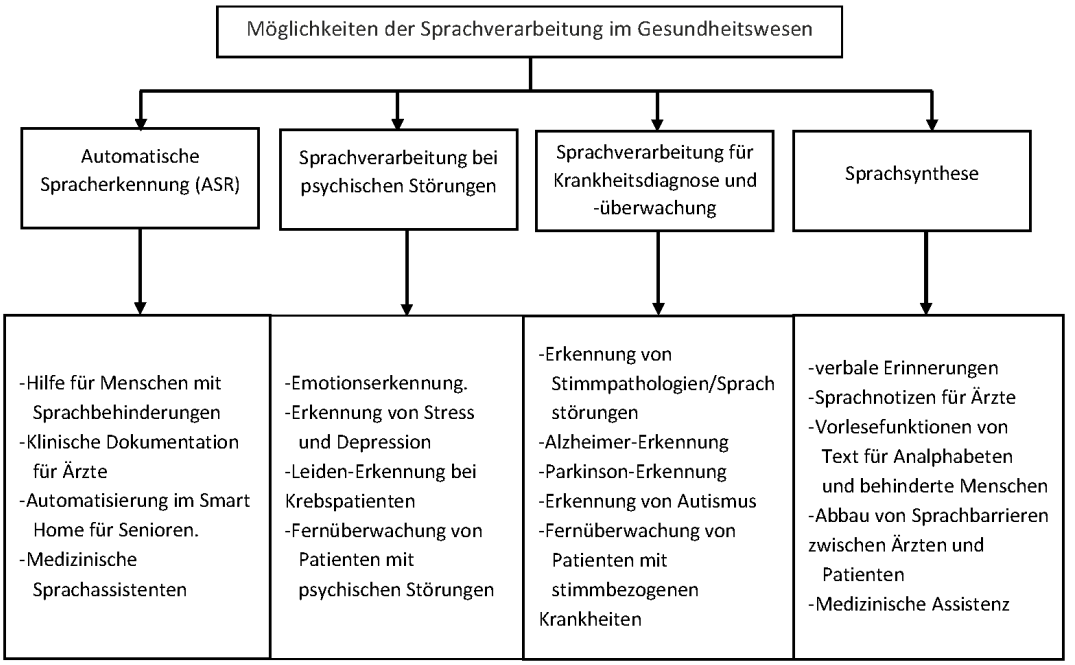

Abbildung 8: Nicht erschöpfende Übersicht über Einsatzmöglichkeiten von Sprachsystemen im Gesundheitswesen (übersetzt und in Anlehnung an Latif et al. 2020, S. 346)

Stimm-, sprach- und gesichtserkennungsbasierte Anwendungen im medizinischen Bereich können nach unterschiedlichen Gesichtspunkten kategorisiert werden, etwa Funktionalität, Nutzer- und Zielgruppe oder Anwendungsfeld (Evers-Wölk et al. 2018, S. 47–52). Im Folgenden erfolgt die Differenzierung zunächst nach Anwendungsfeld: Hier steht insb. die Erkennung und weniger die Therapierung von Krankheiten im Vordergrund. Eine zweite Differenzierung erfolgt nach physischen und psychischen Erkrankungen. Des Weiteren geht es um solche Applikationen, die entweder das medizinische Fachpersonal bei der Diagnosestellung unterstützen sollen oder Laien die Selbstdiagnose ermöglichen sollen.

Dementsprechend erfolgt im Folgenden zunächst die Betrachtung der Erkennung physischer Krankheiten und psychischer Krankheiten sowie der entsprechenden technischen Grundlagen und Möglichkeiten. Daran schliesst sich die Betrachtung der rechtlichen Rahmenbedingungen sowie der gesellschaftlichen und ethischen Herausforderungen an.

3.7.1. Erkennung physischer Krankheiten

Die menschliche Stimme entsteht in einem Zusammenspiel von vielen anatomischen Strukturen im Körper, welche u.a. 100 Muskeln in Anspruch nimmt. Das Gehirn regelt dieses Zusammenspiel (Geuter 2015, S. 297). Somit lässt sich aus der Stimme auf eine Vielzahl von möglichen Erkrankungen schliessen. Ähnlich ist es bei der Analyse von Gesichtsausdrücken: Beispielsweise können hängende Mundwinkel auf einen Schlaganfall hindeuten (Suchy et al. 2001; Xu et al. 2020a).

Schon vor der globalen Covid-19-Pandemie gab es Versuche, mittels Stimme, Sprache und Gesichtsausdrücken Krankheiten zu erkennen. So wurde u.a. erforscht, wie mittels Sprach- und Gesichtserkennung eine Früherkennung von Krankheiten erfolgen kann. So soll bspw. Parkinson durchschnittlich zwei Jahre vor einer klassischen ärztlichen Diagnose erkannt werden können (Schachtler 2019). Schon heute können Systeme zur Erkennung einer Reihe von Krankheiten eingesetzt werden: Mittels Analyse der Stimme können z.B. Alkoholismus und Ermüdung (Schuller et al. 2011), Parkinson (Godino-Llorente et al. 2017; Jeancolas et al. 2021; Schuller et al. 2015) und Erkältungssymptome (Schuller et al. 2017) erkannt werden. Mittels Analyse von Stimme und Sprache, also auch der Wortwahl, kann Alzheimer (Fraser et al. 2016) erkannt werden. Die Erkennungsraten liegen bei den Systemen meist zwischen 80 % und 95 %.

Seit Beginn der Covid-19-Pandemie wird weltweit an der Früherkennung von Symptomen mittels Analyse von Stimme und Husten geforscht, zum einen auf der Grundlage von Smartphone-Apps (Schuller et al. 2020; Stasak et al. 2021) und zum anderen mittels der Kombination von Spracherkennung, Schlagworterkennung und Hustenklassifikation (Wei et al. 2020). Ebenso arbeitet das Massachusetts Institute of Technology (MIT) an einer App, welche rein den Klang des Hustens analysiert. Dabei habe das System 98,5 % der Personen, die nachweislich mit Covid-19 infiziert sind, anhand der Hustengeräusche korrekt erkannt. Gleichzeitig sei eine Erkennungsrate von 100 % bei asymptomatischen (also Personen, welche an Covid-19 leiden, aber keine Symptome zeigen) erreicht worden, wobei die Falsch-Positive-Rate bei 16,8 % gelegen habe (Laguarta et al. 2020, S. 278). Derlei Ergebnisse sind allerdings kritisch zu betrachten, da die Daten per Crowdsourcing generiert und nicht gesondert wissenschaftlich überprüft wurden (Schneider 2020). Zudem ist die Falsch-Positive-Rate mit ca. 17 % relativ hoch, da fast jede fünfte gesunde Person fälschlicherweise als infiziert erkannt wurde. Generell sind die in diesem Bereich aktiven Unternehmen mit der Herausforderung der Sammlung brauchbarer Daten konfrontiert. Die bayrische Firma und Spin-off der TU München audEERING hat bspw. bereits eine App für iOS entwickelt, doch kann sie diese nicht über den Appstore vertreiben, weil nach Angaben der Gründerin die Firma Apple derzeit keine Apps dieser Art zulässt, die nicht direkt von der deutschen Bundesregierung kommen bzw. von ihr freigegeben wurden (Kohse 2020). So sucht die Firma weiterhin nach Covid-19-Positiven, die regelmässig Stimmproben spenden, um die Entwicklung der Audiosignale im Verlauf der

Krankheit zu untersuchen. Ziel seien 250.000 qualitativ hochwertige Datensätze, um die Diagnose-App für die Allgemeinheit als Hilfstool zur Verfügung zu stellen (Herzkammer 2021). Das israelische Start-up Vocalis hat bereits eine App auf dem Markt – sucht aber auch nach Stimmspenden. Das vorrangige Ziel von Vocalis ist es, nicht direkt Diagnosen zu liefern, sondern bei der Vorsortierung von Verdachtsfällen zu unterstützen (Anthes 2020).

Aber auch Krankheiten, die auf den ersten Blick nichts mit der Sprache zu tun haben, können erkannt werden. So gelang es bspw. Maor et al. (2018) und Pareek und Sharma (2016) Herzkrankheiten anhand von Stimmmustern zu erkennen. So korrelieren bestimmte Stimmfrequenzmuster mit schweren Erkrankungen der Koronararterien.

Im Bereich der Gesichtserkennung sind die aktuellen Anwendungsbereiche sehr viel kleiner. Jedoch wird auch in diesem Bereich breit geforscht. Im EU-Projekt «SEMEOTICONS» wurde bis 2016 an einer Art smartem Spiegel geforscht, der mit verschiedenen Sensoren und Kameras ausgestattet ist, um u.a. Herzkrankheiten und Diabetes zu erkennen. So zeigte sich, dass mittels Gesichtserkennung zusätzlich zu psychischen Faktoren auch die Fitness und der Ernährungszustand der Person erkannt werden könnten. Hierzu werden u.a. die Farbe der Haut und Schleimhäute, die Verteilung des Unterhautfettgewebes oder das Schwitzmuster analysiert (Colantonio et al. 2015, S. 2).

In der Diskussion ist auch, dass technische Erkennungssysteme v.a. bei der Erkennung von sehr seltenen Krankheiten nützlich sein könnten, wofür es nur sehr wenige Expertinnen und Experten gibt. Grundsätzlich lassen sich Infektionen, Gendefekte und Notfallerkrankungen (wie z.B. Schlaganfälle) in den Gesichtsausdrücken von Menschen erkennen (Kline et al. 2015). Rodríguez-Blanque et al. (2019) injizierten Testprobanden eine Form des e-Coli-Bakteriums, anderen eine Kochsalzlösung. Danach sollten andere Teilnehmer die Gesundheit der Testprobanden anhand der Gesichtsausdrücke auf einer zehnstufigen Skala bewerten. Die Gesichter der Probanden, die nur eine Kochsalzinjektion erhielten, wurden als deutlich gesünder eingestuft. Doch auch Maschinen seien hierzu in der Lage, wie das Forscherteam um Prof. Dr. Peter Krawitz der Universität Bonn erforscht (Gurovich et al. 2019). Dort wurde ein System namens «DeepGestalt» entwickelt, das seltene Krankheiten oder Gendefekte in einem Gesichtsfoto eines Patienten erkennen kann. Hierzu vergleicht die Software den Gesichtsausdruck mit einer grossen Datenbank an weiteren Fotos mit bestimmten diagnostizierten Krankheiten. Insgesamt sind in der Datenbank 17.000 Fotos enthalten, die mehr als 200 Syndrome repräsentieren (Gurovich et al. 2019, S. 60). So könne das Programm im Beispiel des Kindes mit blauen Augen und hängenden Mundwinkeln 130 Punkte im Gesicht mit 216 Syndromen vergleichen (MDR 2019). Das System gibt am Ende eine Top-10-Liste der möglichen Krankheiten aus. Die korrekte Krankheit tauche mit einer 90-prozentigen Wahrscheinlichkeit in dieser Liste auf (Gurovich et al. 2019, S. 60; Pantel et al. 2020). Die Forscher arbeiten daran, weitere sehr seltene Krankheiten zu erkennen und haben hierfür nun mit «GestaltMatcher» eine KI entwickelt, die im Gegensatz zu «DeepGestalt» unsupervised-learning (also Lernen ohne Vorgaben durch Menschen) nutzt. Somit ist es auch möglich, eine Krankheit zu erkennen, wenn hierzu keine Daten im Trainingsdatensatz vorhanden waren (Hsieh et al. 2021).

Lin et al. (2020) entwickelten eine KI, die Fotoaufnahmen von Patienten untersucht, um koronare Herzkrankheiten am Gesichtsausdruck zu erkennen. Während die Erkennung bei 80 % liegt, ist die Spezifität (also die Anzahl der Falsch-Positiven) mit 54 % allerdings extrem hoch (Lin et al. 2020).

3.7.2. Technische Grundlagen und Möglichkeiten bei der Erkennung physischer Krankheiten

Technisch gesehen, bestehen zwischen der Krankheitserkennung und der in Kapitel 3.8.1 beschriebenen Emotionserkennung grosse Gemeinsamkeiten. In beiden Anwendungsbereichen werden gefundene Stimm- oder Mimikmuster mit Eigenschaften verglichen, die lediglich in einem Bereich Emotionen und im anderen Bereich Krankheiten zugeordnet sind. Mittels verschiedener Algorithmen und neuronaler Netzen können Anomalien im Sprachmuster (Spektrogramm) oder Veränderungen der Mimik, z.B. bei Gesichtsteillähmungen oder anatomischen Veränderungen aufgrund von Gendefekten erkannt und zugordnet werden (Latif et al. 2020, S. 348).

Akustische Merkmale zur Erkennung sind bspw. spektral, prosodisch (z.B. Wort- und Satzakzente), cepstral (das Ergebnis einer Fourier-Analyse, um periodische Strukturen in Frequenzspektren zu erkennen), glottal (vom Kehlkopf kommend) und die berechnete Energie in einem Stimmsignal (Latif et al. 2020, S. 348). Im Bereich der Krankheitserkennung heissen diese Muster «Biomarker». Für jede Krankheit gibt es viele unterschiedliche Marker (Jeancolas et al. 2021). Diese werden dann mit Daten von gesunden Personen im gleichen Alter und des gleichen Geschlechts verglichen (Aronson und Bless 2011).

Im Falle der Parkinson-Erkrankung lassen sich Unregelmässigkeiten im Sprachsignal bei Vokalen erkennen. Hierzu reicht bereits eine sehr kurze Sprachsequenz von nur 20 Millisekunden (Frid et al. 2016, S. 4). Forscher untersuchten den «Aahh»-Laut von gesunden und an Parkinson erkrankten Menschen und fanden zehn akustische Charakteristika, die eine Parkinson-Diagnose am ehesten unterstützen. Dabei handelte es sich bspw. um die Atmung und um «zitternde» Schwingungen in Tonhöhe und Klangfarbe. So erreichten die Forscher nur anhand dieser zehn Biomarker eine Genauigkeit von fast 99 % (Tsanas et al. 2012). Alzheimerpatienten verwenden eher kürzere Wörter, ein kleineres Vokabular und abgehackte Sätze. Zudem wiederholen diese sich häufiger und nutzen häufiger Pronomen wie «es» oder «dieses» anstelle von Hauptwörtern (Fraser et al. 2016). Für eine Covid-19-Erkrankung liegen ebenfalls Biomarker vor. Das Augenmerk liegt hier auf Frequenzmustern in der Sprache, insb. der Vokale. Mittels Algorithmen des maschinellen Lernens werden mehrere Tausend Merkmale untersucht und daraus die Wahrscheinlichkeit für eine Infektion abgeleitet (Schuller et al. 2020).

Die Biomarker der gesuchten Krankheit werden anschliessend mittels Klassifikationsalgorithmen mit den Merkmalen in der Stimme oder Mimik des Patienten verglichen. Das Ergebnis ist ein Übereinstimmungswert (ein Wahrscheinlichkeitswert), der angibt, zu wie viel Prozent die extrahierten Merkmale mit den zuvor festgelegten Biomarkern einer Krankheit übereinstimmen. Diese Werte sind abhängig von dem zuvor festgelegten Schwellenwert. Ein Schwellenwert könnte bspw. vorgeben, dass mindestens vier Biomarker für die Krankheit übereinstimmen müssen. Der Klassifikationsalgorithmus würde dann alle Biomarker analysieren und als Ergebnis aussagen, dass Biomarker 1, 2, 3 und 4 zu 98,74 % zu Krankheit A passen. Im Bereich der Stimmerkennung haben sich Methoden durchgesetzt, die auf «Convolutional neural networks» (CNN; siehe Kap. 3.8.1) und sog. Long short-term memory (LSTM) setzen (Latif et al. 2020, S. 346). Bei LSTM handelt es sich um eine besondere Art von neuronaler Netzwerkarchitektur, bei der es eine direkte Feedback-Schleife in der Architektur gibt (Hochreiter und Schmidhuber 1997).

Eine grosse Herausforderung aufseiten der Forschenden und Anwendungsentwickler liegt in der Schaffung einer angemessenen Datenbasis (Schuller 2019, S. 7). So sind die Kosten für die Generierung von medizinischen Testdaten kostenintensiv, mit ethischen und rechtlichen Einschränkungen verbunden und der Zugriff auf viele Patienten mit ähnlichen Krankheitsbildern ist schwierig (Cummins et al. 2020, S. 142). Aktuell gibt es ein paar wenige Datenbanken für Stimmanalysen. Die Firma SurveyLex kann auf etwa 100.000 Personen zurückgreifen, deren Daten für Studien herangezogen werden können (Anthes 2020). Andere bieten Nutzern die Möglichkeit, eigene Stimmproben zu liefern, falls sie bspw. durch Kehlkopfkrebs ihre Stimme verlieren (Riccio 2019). Für Aufnahmen von Covid-19-Erkrankten bietet das «Sonde Health COVID-19 2020 dataset» Daten (Stasak et al. 2021; Sonde One 2021). Zwar erhoffen sich Forschende durch die weitverbreitete Nutzung von Smartphones eine grössere Datenbasis zu erhalten. Doch hier kommt das Problem der Qualität der Eingangsdaten zum Tragen. So sind die Unterschiede in der Erkennungsgenauigkeit zwischen Laborbedingungen und der Praxis sehr gross.

Ein Lösungsansatz für dieses Dilemma wäre der Einsatz von Generative Adversarial Networks (GAN), die Testdaten selbst genieren können (Cummins et al. 2020, S. 142). Ein Beispiel dieser Algorithmen findet sich in der digitalen Bild- und Videoverarbeitung. So gibt es im Programm Adobe Photoshop die Möglichkeit, Bildteile auszuschneiden und mit passendem Hintergrund zu füllen. Damit kann bspw. bei einem Foto ein auf einer Strasse fahrendes Auto aus dem Foto entfernt werden, ohne dass die Löschung erkenntlich würde. Hierzu muss der Algorithmus neue Inhalte erzeugen, da der echte Strassenbelag durch das Auto verdeckt und somit nicht zu erkennen ist. In diesem Fall werden Daten erzeugt, die dem schon existierenden Datensatz, also dem sonstigen Strassenbelag, statistisch ähnlich sind. Ähnlich ist dies auch mit Audiodaten möglich. Grundsätzlich stehen zur Generierung von neuen Daten zwei neuronale Netze im Wechselspiel bzw. im Wettbewerb zueinander. Das erste Netzwerk ist das «Generator Network» (GN) und erzeugt erste Daten. Diese bestehen anfangs noch aus reinen Zufallsdaten (Rauschen). Das zweite Netzwerk ist das «Discriminator Network» (DN). Es bewertet die vom GN generierten Daten. Damit das DN die Daten bewerten kann, wird es zuvor mit echten Daten (z.B. echten Stimmproben von Parkinson-Patienten) trainiert. Ziel des DN ist es zu erkennen, ob die Daten echt oder gefälscht sind. Das GN wiederum versucht in jeder Iteration ein Bild oder eine Sprachsequenz zu erstellen, welche das DN als «echt» ansieht. Durch das jeweilige Feedback (echt/gefälscht) des DN lernt das GN dazu, bis am Ende Daten erstellt wurden, die zwar nicht echt sind, sich von echten Daten jedoch nicht bzw. kaum mehr unterscheiden lassen (Goodfellow et al. 2014; Donahue et al. 2018). Dass dieses Vorgehen erfolgreich ist, zeigten u.a. Deng et al. (2017), die Trainingsdaten für Stimmerkennungssysteme zur Autismus-Erkennung von Kindern generierten. Ein anschauliches Beispiel für den Erfolg von GANs sind zudem sog. Deepfake Audio- und Videoinhalte, die sich mittlerweile mit blossem Auge teils kaum mehr von echten Aufnahmen unterscheiden lassen. Da für das Training von GANs auf reale Daten zurückgegriffen werden muss, ist dieser Ansatz der Generierung synthethischer Daten jedoch kritisch zu betrachten, weil dadurch eine Verzerrung (Bias) (re-)produziert werden kann (van Steenkiste et al. 2020, S. 309).

Forschende hinterfragen zudem die Aussagekraft von Biomarkern. So steht zur Debatte, ob die Biomarker eine direkte, typische Folge der betreffenden Krankheit sind oder ob diese schlicht Unterschiede zwischen Testgruppen deutlich machen, wie z.B. Alter, Geschlecht

oder Körpergrösse oder auch den Bildungsstand. Zudem wäre es denkbar, dass ein und dieselbe Krankheit sich bei unterschiedlichen Menschen in Gestalt verschiedener Biomarker äussert, was die Aussagekraft der gesamten zur Erkennung von Krankheiten verwendeten Methodik relativieren würde. Zudem haben auch Allergien Einfluss auf die Stimme, z.B. erkennbar an dauerhafter Heiserkeit, welche selbst wieder Einfluss auf die gefundenen Biomarker haben (Anthes 2020).

3.7.3. Erkennung psychischer Krankheiten

Angesichts der anhaltenden Zunahme psychischer Erkrankungen in Europa und der Schweiz (Fehr 2019) wird auch an Möglichkeiten zur automatisierten Erkennung dieser geforscht (Schär Gmelch 2019).

Die «Diagnostik und Therapie der Psychiatrie [ist] mit der Sprache verbunden» und dies mehr als «als andere Bereiche der Medizin» (Meyer-Lindenberg 2018, S. 863). Laut des Psychiaters Charles Marmar von der New York University ist die Stimme «enorm vielfältig dabei, unsere Emotionssignale zu vermitteln». So geben «Tempo, Rhythmus, Lautstärke, Tonhöhe, […] Aufschluss darüber, ob ein Patient niedergeschlagen und entmutigt ist, ob er aufgeregt und ängstlich, dysphorisch und manisch ist» (Anthes 2020). Daher lassen sich einige psychische Erkrankungen wie Depressionen (Valstar et al. 2016; Ringeval et al. 2017), Psychosen (Corcoran et al. 2018; Rosenstein et al. 2015), Burnout (Rodrigues et al. 2021) oder bipolare Störungen (Ringeval et al. 2018) mittels Stimmanalyse erkennen. Die Erkennungsraten liegen bei den Systemen meist zwischen 80 % und 95 %. Die gesprochene Sprache einer Person ist zudem ein guter Anhaltspunkt, um Suizidabsichten zu erkennen (Rana et al. 2019; Cummins et al. 2015). Aufgrund der Vielzahl möglicher Erkenntnisse sei die Spracherkennung bei der Behandlung von Depressionen zum Standard geworden (Latif et al. 2020, S. 348).

Beispielsweise könne Schizophrenie anhand von verbalen Gedächtnisdefiziten erkannt werden. Holmlund et al. (2020) entwickelten einen verbalen Gedächtnistest mit automatischer Bewertung und testeten diesen an 104 Studienteilnehmern (25 Patienten mit schweren psychischen Erkrankungen und 79 gesunde Freiwillige). Alle Teilnehmer hörten eine Geschichte und mussten diese zunächst sofort und nach Ablauf einer gewissen Zeit ein weiteres Mal wiederholen. Indem die rezipierte Geschichte mittels Spracherkennung mit dem Original verglichen wurde, konnten Aussagen über die Gedächtnisleistung abgeleitet werden. Dieser automatisierte Ansatz erzeugte dabei Transkripte der Patientenerzählungen mit einer Genauigkeit von 82 % (welche zu 99 % mit den Transkripten übereinstimmten, die ein Mensch aus den Erzählungen geschrieben hatte), wobei die Autoren zugeben, dass die Sprache von gesunden Menschen deutlich besser erkannt werden konnte als die von Kranken (6,2 % Fehlerrate im Vergleich zu 24,8 %) (Holmlund et al. 2020, S. 5). Zur Erkennung von Schizophrenie im Frühstadium nutzten Wu et al. (2012a) ebenfalls Spracherkennung. Dabei untersuchten sie, wie und ob Schizophrenie-Patienten auch in schwierigen Gesprächsumgebungen (sog. «Cocktail-Party-Situationen») durch verschiedene kognitive Hinweise, wie Gesprächsinhalte, die Aufmerksamkeit auf den Sprecher lenken können. Die Ergebnisse zeigten, dass Patienten mehr Schwierigkeiten beim Verständnis hatten, wenn

weitere Stimmen eingeblendet werden, als wenn zusätzliche Geräusche eingeblendet werden.

Forschenden ist es zudem gelungen, Autismus bereits im frühkindlichen Stadium im Alter von zehn Monaten anhand von vorsprachlichen Vokalisationen mit einer Genauigkeit von 75 % zu erkennen (Pokorny et al. 2017). Die Diagnose von Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung (ADHS) könne mit Stimmerkennungssystemen erfolgen, da Kinder mit dieser Störung schneller und lauter sprechen als andere Kinder (Anthes 2020; Söderlund und Jobs 2016). Marmar et al. (2019) untersuchten die Stimme von 129 männlichen ehemaligen Militärangehörigen und identifizierten 18 Stimmmerkmale, die auf eine posttraumatische Belastungsstörung (PTBS) schliessen lassen. Mittels dieser Indikatoren konnte das System eine fast 90-prozentige Treffergenauigkeit erreichen. Hierbei waren v.a. die Indikatoren für die langsame, gedämpfte und monotone Sprechlage sehr aussagekräftig (Marmar et al. 2019).

Sprach- und Gesichtserkennung können zudem genutzt werden, um das Asperger- und Down-Syndrom zu erkennen. So zeigen Patienten mit Asperger-Syndrom einen hohen Intelligenzquotienten (IQ), aber eine geringe emotionale Intelligenz (EQ), während dies bei Patienten mit Down-Syndrom genau umgekehrt der Fall ist. El-Seoud und Ahmed (2019) nutzen Sprach- und Emotionserkennung, um IQ und EQ von Patienten zu bewerten. Mittels eines Chatbots können Patienten mit dem System interagieren. Im EQ-Test werden die Stärken und Schwächen der Patienten analysiert, indem in verschiedenen Situationen (z.B. der Umgang mit anderen Menschen oder in Konversationen) die Gesichtsausdrücke, die Augen und die Stimme analysiert werden (El-Seoud und Ahmed 2019, S. 204).

Auch ADHS lasse sich anhand von Gesichtsausdrücken erkennen. So zeigten Pelc et al. (2006), dass es eine Korrelation zwischen zwischenmenschlichen Problemen und emotionalen Gesichtsausdruck-Dekodierungsstörungen für Ausdrücke, die Veränderung darstellen, gibt. Kinder mit Angststörungen zeigten zudem in verstärktem Masse angstvolle Gesichtsausdrücke (Ketamo und O'Rourke 2019).

Ziel vieler Erkennungssyteme ist es, mögliche Risikofaktoren für gewisse Krankheitsbilder herauszufinden, um schon früh eine Prophylaxe zu schaffen, bevor eine Erkrankung behandlungspflichtig in Erscheinung tritt (Hahn et al. 2017, S. 37).

3.7.4. Technische Grundlagen und Möglichkeiten bei der Erkennung psychischer Krankheiten

Die grundsätzliche technische Funktionsweise der Erkennung von psychischen Krankheiten mittels Stimm-, Sprach- und Gesichtserkennung gleicht der in Kapitel 3.7.2 beschriebenen Erkennung von physischen Krankheiten. Auch zwischen der Erkennung psychischer Krankheiten und der in Kapitel 3.8 untersuchten Emotionserkennung bestehen grosse Gemeinsamkeiten im Hinblick auf die technische Funktionsweise.

Ob gefundene Stimm- oder Mimikmuster mit Eigenschaften von Emotionen oder Krankheiten verbunden werden, ist technisch unerheblich. Dabei werden Biomarker mit Daten von gesunden Personen im gleichen Alter und des gleichen Geschlechts verglichen, um unter

Einsatz von Algorithmen und neuronalen Netzen Spektrogramme zu erstellen (Aronson und Bless 2011).

Für jedes Krankheitsbild sind unterschiedliche Algorithmen oder Merkmale der Stimme ausschlaggebend. So zeigten Scherer et al. (2013), dass suizidale Personen am besten mittels der prosodischen Merkmale der Stimme erkannt werden können. Depressionen erkennt man u.a. in der eher monotonen Tonhöhe, wie in Abbildung 9 beschrieben (Anthes 2020). Kinder mit ADHS sprechen hingegen schneller und lauter (Anthes 2020). Dennoch kommt bei der Erkennung solcher Krankheiten eine Vielzahl verschiedener Parameter zum Einsatz.

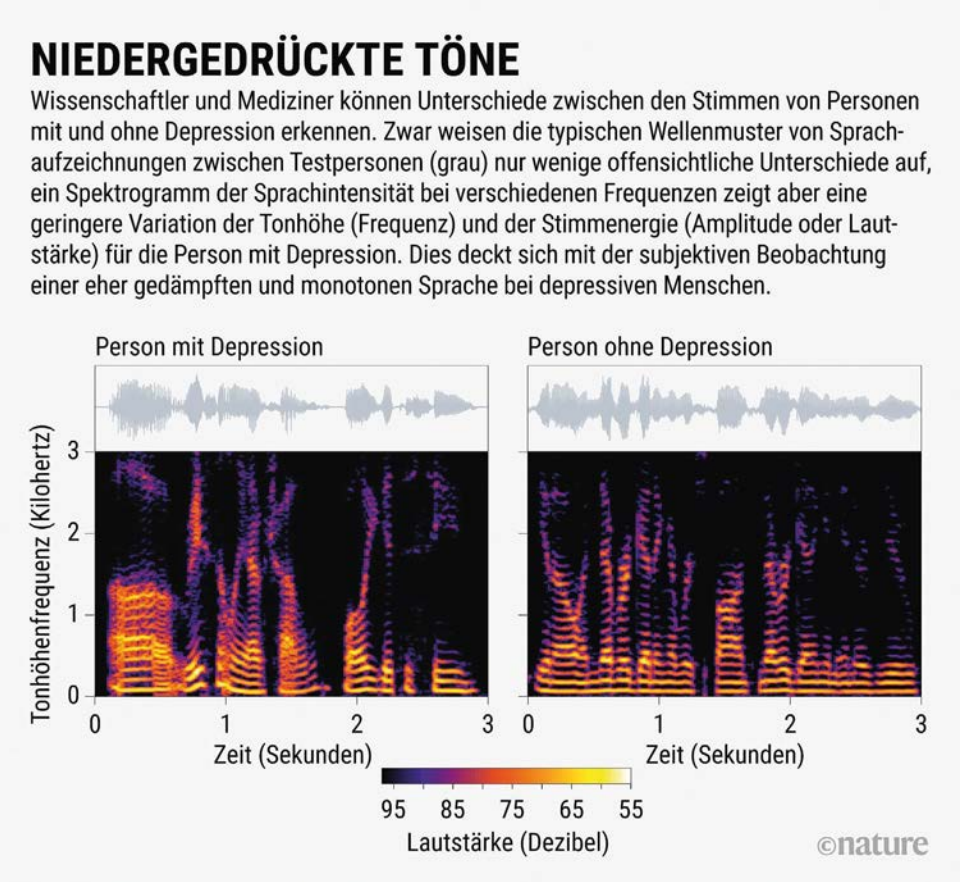

Abbildung 9: Depressionen können anhand der Tonhöhe und Stimmenergie erkannt werden (Anthes 2020)

Kulturelle Einflüsse müssen zudem mit in die Bewertung einbezogen werden. Wenn man bspw. die Lautstärke der Sprache von Kindern als Biomarker für ADHS heranzieht, dann sollte kein direkter Vergleich von Schweizer Kindern mit italienischen Kindern gemacht werden, da Letztere kulturell bedingt eine höhere Lautstärke vorweisen (Anthes 2020).

Im Hinblick auf die Datenverfügbarkeit gibt es zudem einen grossen Unterschied zwischen der Erkennung physischer und psychischer Krankheiten. Forschende verweisen hier insb. darauf, dass viele Menschen bei der Nutzung sozialer Medien viele Informationen öffent-

lich über sich selbst preisgeben, die zur Verbesserung sprachbasierter Erkennungssysteme genutzt werden könnten (Meyer-Lindenberg 2018, S. 864). Teilweise werden dabei zur Extraktion von Anhaltspunkten für psychische Krankheiten Sprachdaten mit anderen Daten, etwa Frequenz und Art von Internetaktivitäten, kombiniert (Meyer-Lindenberg 2018, S. 864). Siddiqui und Javaid (2020) zeigen, dass sich ein Grossteil der durchgeführten Studien nur auf die Kombination von Stimme und Gesicht beziehen. Nur sehr selten werden auch Körperbewegungen (Soleymani et al. 2012; Metallinou et al. 2016) oder EEG-Daten kombiniert (Koelstra et al. 2012). Die Erkennungsraten bei der Kombination von verschiedenen Eingangsdaten variieren dabei. Die Kombination mit EEG-Daten erreicht Werte zwischen 50 % und 65,1 % (Koelstra et al. 2012, S. 28). Die Kombination von Infrarotbildern und Aufnahmen des Gesichts ermögliche eine Genauigkeit von 85 % (Siddiqui und Javaid 2020, S. 11). Im Falle einer Kombination von Infrarotbildern, Gesichtsbildern und Sprache könne eine Genauigkeit von 79,41 % bis 100 % erreicht werden (Siddiqui und Javaid 2020, S. 13). Eine Kombination von Gesichtsaufnahmen, Sprache und geschriebenen Text erreiche Werte zwischen 7,8 % und 88,98 % (Yoon et al. 2018; Choi et al. 2018).

Derartige Studien basieren allerdings zumeist nicht auf Daten, die in realen Umgebungen erfasst wurden, sondern auf exemplarischen Datensätzen aus verschiedenen Bild- und Tondatenbanken und werden häufig in experimentellen, kontrollierten Umgebungen durchgeführt. Dementsprechend lassen sich die o.g. Trefferraten nicht auf reale Anwendungen von Stimm-, Sprach- und Gesichtserkennungstechnologien übertragen. Beispielsweise konnten Huang et al. (2019) Depressionen unter Einsatz eines professionellen Mikrofons mit einer Genauigkeit von ca. 94 % erkennen. Diese Genauigkeit sank jedoch auf weniger als 75 %, sobald die Stimmproben mit Smartphones in Alltagsumgebungen aufgezeichnet wurden.

3.7.5. Juristische Bewertung

Die im vorgehenden Kapitel behandelten Szenarien würden eine Krankheitsdiagnoseerstellung anhand von Stimmabdrücken oder Gesichtsbildern ermöglichen. Daten, die eine Diagnosestellung erlauben, gelten als Gesundheitsdaten. Gemäss Datenschutzgesetz handelt es sich dabei um besonders schützenswerte Personendaten und es kommen verschärfte Anforderungen zur Anwendung.[117] In diesem Abschnitt werden rechtliche Vorgaben für mögliche Szenarien untersucht; einerseits der Einsatz der Technologie durch ärztliches Fachpersonal (Kapitel 3.7.5.1), in Bewerbungsverfahren (Kapitel 3.7.5.2), durch Versicherungen (Kapitel 3.7.5.3) sowie andererseits durch die Betroffenen selbst (Kapitel 3.7.5.4).

3.7.5.1. Diagnosestellung durch medizinisches Fachpersonal

Bei der *Diagnosestellung durch ärztliches Fachpersonal* werden immer auch Personendaten bearbeitet. Abhängig davon, ob das Personal in einer Privatklinik oder in einem öffentlichen Spital tätig ist, liegt eine Datenbearbeitung durch Privatpersonen oder durch Behör-

[117] Art. 5 lit. c Ziff. 2 nDSG.

den vor.[118] Der Schutz von Daten bezüglich Weitergabe an Dritte wird grundsätzlich durch das Arztgeheimnis erfasst; für alle anderen Bearbeitungsschritte sind die entsprechenden Datenschutzgesetze massgebend (Bischof, S. 41–61; Büchler und Michel 2014, S. 86).

Das im Grundsatz von Treu und Glauben enthaltene Erfordernis der *Transparenz* muss für jeden Datenbearbeitungsschritt erfüllt werden. Hier erscheint insb. der Grundsatz der Erkennbarkeit wichtig; das heisst, dem Patienten muss bewusst sein, dass der Arzt ein zur Frage stehendes Diagnoseinstrument einsetzt und welche Daten damit bearbeitet werden. Auch wären zumindest die Identität, die Kontaktdaten des Verantwortlichen, der Bearbeitungszweck und eine allfällige Datenbekanntgabe an Dritte explizit zu benennen (Golla 2015, S. 197).[119] Im KI-Verordnungsvorschlag der EU-Kommission sind für den Einsatz dieser Diagnoseinstrumente weiter gehende Informationspflichten vorgesehen, z.B. verwendete Trainingsdaten und die Genauigkeit des Systems; entsprechende Regelungen bestehen in der Schweiz jedoch noch nicht (Braun Binder et al. 2021, S. 9–10).[120]

Der Grundsatz der *Zweckbindung* verlangt, dass es für betroffene Personen auch erkennbar ist, zu welchem Zweck die Daten erhoben werden. Diese sind auch nur so zu bearbeiten, dass sie mit dem vereinbarten Zweck korrespondieren. Das Fachpersonal hat den Patienten über die Aufnahme der Stimme oder des Gesichtsbilds zur Erstellung einer Diagnose aufzuklären.

Nach dem Grundsatz der *Datenminimierung* sind erhobene Daten nach erfolgter Diagnosestellung zu löschen. Dies steht im Widerspruch zum Wunsch, für die Datenqualität möglichst viele Datensätze zu sammeln und anhand derer das System zu verbessern (Braun Binder et al. 2021, S. 14).

Das Erfordernis der *Datenrichtigkeit* ergibt sich sowohl aus dem Datenschutzgesetz als auch aus dem Behandlungsvertrag. Eine Krankheitsgeschichte muss immer wahrheitsgetreu und vollständig sein (Uttinger 2015, S. 323).

Gemäss dem Grundsatz der *Datensicherheit* muss der Verantwortliche zudem Sicherheitsmassnahmen auf technischen und organisatorischen Ebenen vorsehen, um den Verlust oder unerlaubte Zugriffe auf die Daten zu verhindern. Abzuklären wäre es weiter, ob beim Einsatz der Software besonders schützenswerte Personendaten automatisch an Dritte weitergegeben werden oder eine Datenbekanntgabe ins Ausland erfolgt.[121]

Eine medizinische Behandlung ohne Einwilligung des Patienten stellt einen widerrechtlichen Eingriff in dessen Persönlichkeit dar. Es ist daher – zur Gewährleistung der Selbstbestimmung des Patienten im Blick auf eine medizinische Behandlung – grundsätzlich eine informierte Einwilligung nötig (m.w.H. Liang 2018, S. 71–73). Dies gilt auch für die Diagno-

---

[118] In einem Kantonsspital wären die kantonalen Datenschutzbestimmungen für staatliche Organe anwendbar, während bei der Behandlung in einer Hausarztpraxis die Bestimmungen für Private massgebend sind. Siehe Büchler und Michel (2014) sowie Vokinger 2012).

[119] Art. 19 nDSG.

[120] Siehe Art. 13 des Vorschlags für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union, Brüssel, den 21.4.2021, COM (2021) 206 final.

[121] Art. 30 Abs. 2 lit. c nDSG.

sestellung als Teil des Behandlungsvertrages. Das Recht des Patienten, eine Behandlung abzulehnen, ist absolut (Büchler und Michel 2014, S. 80). Auch aus dem Datenschutzgesetz ergibt sich, falls die Person nicht möchte, dass ihre Daten bearbeitet werden, ein Widerspruchsrecht. Als problematisch zu beurteilender Punkt von Assistenzsystemen ist der allfällig daraus resultierende übermässige Geltungsanspruch zu nennen. So wird befürchtet, dass diese Art der Diagnosestellung das Verhalten der Patienten beeinflussen könnte, wenn diese das Resultat als uneingeschränkte Wahrheit interpretieren. Auch Ärzte könnten sich auf das Resultat der Software verlassen (Golla 2015, S. 194) und möglicherweise ihrer Verpflichtung, die Behandlung sorgfältig und nach Regeln der ärztlichen Kunst durchzuführen, nicht gerecht werden (Büchler und Michel 2014, S. 80).

Im Falle eines Einsatzes der Software, die zu einer falschen Diagnose führt, besteht ein Diagnosefehler. Wenn die Diagnose jedoch durch das ärztliche Fachpersonal sorgfältig und die entsprechenden Mittel korrekt eingesetzt wurden, gilt der Irrtum als vertretbar, womit keine Haftungs- und Entschädigungsansprüche daraus resultieren. Weitere relevante rechtliche Vorschriften befinden sich im Heilmittelgesetz (HMG). Eine Software, welche eine physische oder psychische Krankheit erkennt und somit zur Diagnose benutzt werden kann, gilt als Medizinalprodukt (Klett und Verde 2016, S. 45–47). Demnach wären die Leistung und ein Konformitätsbewertungsverfahren nachzuweisen.[122] Auch das Produktehaftpflichtgesetz und die Produktesicherheitsbestimmungen fänden Anwendung (Klett und Verde 2016, S. 51).

3.7.5.2. Bewerbungsverfahren

Ein anderes denkbares Szenario wäre der Einsatz der erwähnten Technologien *in Bewerbungsverfahren*. Ein Arbeitgeber darf sich über künftige Arbeitnehmer ein Bild machen und zu diesem Zweck auch Angaben zur Person sammeln. Erlaubt sind Fragen, die zur Abklärung der Fähigkeit des Arbeitnehmers für die entsprechende Tätigkeit dienen oder zur Durchführung des Arbeitsvertrages erforderlich sind (Portmann und Rudolph 2020). Ein Verstoss dieser Bearbeitungsvorschrift von Art. 328b OR führt zu einer widerrechtlichen Persönlichkeitsverletzung. Vorbehalten bleibt das Vorbringen eines Rechtfertigungsgrundes nach Art. 13 DSG (Schmidlin 2022).[123] Setzt ein Arbeitgeber eine der oben beschriebenen Technologien im Bewerbungsverfahren offen ein, erscheint dies insb. aufgrund der Machtasymmetrie zwischen Arbeitgeber und Arbeitnehmer problematisch. Von der Lehre als zulässig betrachtet werden etwa Tests betreffend Charaktereigenschaften bei der Besetzung von Führungspositionen (Verantwortungsbewusstsein, Führungsverhalten und Zuverlässigkeit) sowie z.B. auch die Abklärung einer Alkoholabhängigkeit bei Berufschauffeuren oder einer Epilepsieerkrankung bei Piloten (von Kaenel 2006, S. 102). Entsprechend einem psychologischen Eignungstest oder einem grafologischen Gutachten wäre auch für die Analyse der Stimm- oder Gesichtsdaten eine Einwilligung der Betroffenen nötig (Papa und Pietruszak 2015; von Kaenel 2006).[124] Gemäss dem EDÖB kann eine Einwilligung aber

[122] Art. 45 Abs. 2 und Art. 46 Abs. 1 HEMG.

[123] BGer 4A_518/2020 vom 25. August 2021.

[124] BSK OR I-Portmann, Art. 328b N 27 f.

nur dann Persönlichkeitsverletzungen rechtfertigen, wenn die Testdurchführung für die Abklärung der Eignung für das Arbeitsverhältnis notwendig ist.[125] Für die Gültigkeit der Einwilligung muss diese von der betroffenen Person freiwillig erteilt werden. Das bedeutet insb., dass diese nicht aufgrund äusseren Drucks zustandekommen darf (EDÖB 2006, S. 7). Bei einer (z.B. hierarchisch bedingten) geschwächten Position des Betroffenen könnte die Einwilligung aufgrund von Druck gegeben werden; damit läge keine freiwillige und auch keine gültige Einwilligung vor.[126] Aufgrund der häufig bestehenden Machtasymmetrie in Bewerbungssituationen muss die Freiwilligkeit der Einwilligung sehr häufig bezweifelt werden, weil die Bewerberinnen und Bewerber regelmässig vor die Entscheidung gestellt würden, der Bearbeitung entweder einzuwilligen oder zu befürchten, aufgrund ihrer Nicht-Einwilligung schlechter abzuschneiden bzw. abgelehnt zu werden. Auf jeden Fall müsste die Freiwilligkeit verneint werden, wenn z.B. alle Arbeitgeber einer bestimmten Branche solche Abklärungen im Verlauf des Bewerbungsverfahrens routinemässig verlangen würden.

Als Alternative wäre es grundsätzlich denkbar, dass der Vertrauensarzt – mit der Einwilligung des Stellensuchenden – die hier besprochenen Technologien zur Diagnose einsetzt und so die Eignung für das Arbeitsverhältnis abklärt und anschliessend den künftigen Arbeitgeber über die Tauglichkeit des Stellenbewerbenden informiert (Wildhaber 2010).

Würde die Technologie während eines Bewerbungsgesprächs jedoch verdeckt eingesetzt, läge eine Persönlichkeitsverletzung durch den Arbeitgeber vor. Insb. bestünde ein Verstoss gegen das Gebot von Treu und Glauben.[127] Der betroffene Stellenbewerber könnte den Rechtsweg beschreiten; die Problematik besteht aber darin, dass er wohl kaum in Erfahrung bringen könnte, ob diese Technologie überhaupt eingesetzt wurde und ob eine allfällige Ablehnung damit zu tun hatte (zu genetischen Daten: Baumann 2012, Rz. 10).

3.7.5.3. Versicherungen

Ein weiteres denkbares Szenario wäre der Einsatz der oben genannten Technologien beim *Abschluss von Versicherungen*. Während im Bereich der obligatorischen Krankenversicherung eine Personalisierung der Prämien weitgehend ausgeschlossen ist, besteht diese Möglichkeit im Bereich der Privatversicherungen (Thouvenin 2019).[128] So existieren spezifische Entschädigungen in Form von Prämienvergünstigungen für eine bestimmte tägliche Schrittzahl oder für eine andere Form des gesunden Lebensstils (Thouvenin 2019). Eine Diskriminierung – z.B. die Verweigerung des Abschlusses des Vertrages oder weniger günstige Vertragsbedingungen aus sachfremden Gründen – ist als Persönlichkeitsverletzung zu qualifizieren. Aufgrund eines überwiegenden privaten Interesses kann diese jedoch gerechtfertigt werden. Das überwiegende Interesse wird im Falle von Versicherungen bejaht, die sachliche Gründe für die Individualisierung vorbringen können, was bei individua-

---

[125] Art. 328b und 362 OR.

[126] Urteil d. Eidgenöss. Datenschutzkommission vom 29. August 2003, VPB 2004 Nr. 68 S. 860, 874.

[127] Siehe auch Art.19 nDSG.

[128] Die Prämien der obligatorischen Krankenversicherung sind in Art. 61 ff. KVG und Art. 90c ff. KVV geregelt. Die Höhe der Prämie bestimmt sich nach Wohnort des Versicherten, seinem Alter und nach der Versicherungsform.

lisierten Angeboten aufgrund von Risikoprofilen der Fall sein dürfte. Versicherungen beziehen dabei Daten aus unterschiedlichen Quellen, um die Prämien zu individualisieren. Der zukünftige Versicherte hat zudem eine vorvertragliche Anzeigepflicht und muss sämtliche für die Beurteilung des Risikos relevanten Tatsachen mitteilen, sofern diese bekannt sind oder sein müssten.[129] Zusätzliche Daten werden aber auch bei Dritten erworben (Thouvenin 2019, S. 27–28). Denkbar wäre hier also eine Ableitung von Gesundheitsinformationen aus Stimm- oder Gesichtsdaten der an einem Versicherungsvertrag interessierten Person.

Die Erkennbarkeit der Datenbeschaffung ist nicht nur als Datenbearbeitungsgrundsatz im DSG vorgeschrieben, sondern auch im Versicherungsvertragsgesetz (VVG), welches eine gesetzlich geregelte Informationspflicht des Versicherers vorsieht. Der Versicherungsnehmer muss vor Vertragsabschluss verständlich über seine Identität, den wesentlichen Inhalt des Vertrags und weitere den Vertrag betreffende Aspekte informieren. Umgekehrt hat der Versicherer Informationen betreffend die Bearbeitung von Personendaten, Zweck und Art einer Datensammlung, mögliche weitere Empfänger und Aufbewahrung der Daten bekannt zu geben (Thouvenin 2019, S. 32).[130] Auch der Zweck der Datenbearbeitung hat erkennbar zu sein. Es stellt sich z.B. die Frage, ob Personen realisieren, dass ihr Stimmabdruck oder Gesichtsbild verwendet werden, um sie auf allfällige psychische oder physische Krankheiten zu untersuchen und anhand dessen den Vertrag zu individualisieren und die Prämien anzupassen. Der Versicherer muss den Betroffenen also mitteilen, wenn eine Datenbearbeitung zur Individualisierung des Angebots verwendet wird. Der Betroffene kann allenfalls mithilfe des Auskunftsrechts noch mehr Informationen anfordern (Thouvenin 2019, S. 33).

Gemäss dem Grundsatz der Zweckbindung sollen Daten nur für den angegebenen Zweck verwendet werden oder mit einem weiteren Zweck, der damit vereinbar ist. Hier stellt sich die Frage, ob diese Daten in Kombination mit anderen Daten zu neuen Resultaten führten, welche nicht mehr mit dem ursprünglichen Zweck vereinbar sind.

Der Verhältnismässigkeitsgrundsatz verlangt u.a., dass keine milderen Mittel zur Erreichung desselben Ziels vorhanden sind. Die Grundsätze der Verhältnismässigkeit und der Datenminimierung scheinen hier nicht eingehalten, da in Form von persönlichem Nachfragen einfachere und weniger invasive Mittel zur Abklärung der Gesundheit des Betroffenen vorliegen, denn auch die an Versicherungsverträgen Interessierten sind verpflichtet, wahrheitsgetreu Auskunft zu geben.

Der Grundsatz von Treu und Glauben greift insb. dann, wenn andere Grundsätze nicht zur Anwendung gelangen oder zu einem nachteiligen Ergebnis für die Betroffenen führen würden (Thouvenin 2019; Epiney und Nüesch 2015). Gemäss der Lehre ist es fraglich, ob die Individualisierung von Versicherungsverträgen insgesamt gegen den Grundsatz von Treu und Glauben verstösst (m.w.N. Thouvenin 2019). Eine Analyse der Stimme oder des Gesichtsbildes zur Erstellung einer Diagnose – von welcher der Betroffene u.U. gar nichts weiss – würde aber bedeutend weitergehen als eine blosse Individualisierung der Prämien und damit gegen diesen Grundsatz verstossen. Insb. bei einer psychischen Krankheit könnte es gut sein, dass diese bis anhin unbekannt (z.B. im Rahmen einer Depressionserkran-

[129] Art. 4 Abs. 1 VVG.

[130] Art. 3 Abs. 1 lit. g VVG.

kung) ist und es wäre verfehlt, wenn Betroffene mit dieser Diagnose und einer höheren Prämie alleingelassen würden.

Beim Grundsatz der Datenrichtigkeit stellt sich die Frage nach der inhaltlichen Richtigkeit der Daten. Zwar bieten erwähnte Diagnoseinstrumente unter optimalen Bedingungen eine hohe Wahrscheinlichkeit der richtigen Diagnosestellung, nicht zwingend jedoch auf Distanz (z.B. bei der Erstellung des Stimmabdruckes während eines Telefonats). Auch bei korrekter Anwendung besteht die Gefahr einer Einordnung von Personen in Gruppen (z.B. durch scheinbar erhöhte Krankheitsrisiken), in welche sie nicht gehören. Auch werden diese Technologien zurzeit nicht für sich alleinstehend, sondern im Rahmen von komplexen Diagnosen eingesetzt, was im Hinblick auf die Datenrichtigkeit weitere Fragen aufwirft.

Da für Zusatzversicherungen die Datenbearbeitungsvorschriften für Private massgebend sind, ist im Falle einer Persönlichkeitsverletzung das Vorliegen eines Rechtfertigungsgrundes zu prüfen. Die Individualisierung von Versicherungsverträgen und die dafür nötige Datenerhebung ist gesetzlich nicht vorgesehen (Thouvenin 2019, S. 36–38). Ein überwiegendes privates Interesse scheint fraglich. Für das Versicherungsunternehmen liegt im Hinblick auf die Erstellung der Versicherungsprämien ein Interesse vor; dies gilt auch für diejenigen, welche von der Datenbearbeitung (z.B. billigere Prämien) profitieren. Bezieht man das allgemeine Interesse der Betroffenen in die Interessenabwägung mit ein – gemäss ihrem informationellen Selbstbestimmungsrecht zu entscheiden, welche und insb. wie weitgehend ihre Informationen durch Dritte bearbeitet werden –, so scheint dies (v.a. hins. derjenigen, die durch die weitgehende Datenbearbeitung eine nachteilige Behandlung erfahren) höher als das Interesse der Versicherung bezüglich der Prämienanpassung (Thouvenin 2019, S. 39).

Allenfalls möglich erscheint somit einzig die Rechtfertigung der Persönlichkeitsverletzung mittels Einwilligung des Betroffenen. Da besonders schützenswerte Personendaten, nämlich Gesundheitsdaten, bearbeitet werden, muss die Einwilligung informiert, freiwillig und ausdrücklich erfolgen.[131] Informiert werden würde in diesem Fall bedeuten, dass der Betroffene über die Erhebung und die mögliche Diagnostizierung von Krankheiten anhand von Stimmabdrücken und Gesichtserkennungstechnologien und die daraus folgende Erstellung eines Risikoprofils und der entsprechenden Prämienanpassung explizit in Kenntnis gesetzt wird.[132] Das Erfordernis der Freiwilligkeit bedingt, dass für Betroffene auch eine gangbare Alternative zur Verfügung stünde. So dürfte ein Vertragsabschluss ohne die Verwendung der genannten Diagnosetools keine bedeutenden wirtschaftlichen Nachteile haben (EDÖB 2015, 2016; Thouvenin 2019),[133] dies insb., da die Datenbearbeitung im Blick auf den Zweck unverhältnismässig erscheint. Die Abklärung der Versicherung bezüglich vorbestehender Erkrankungen könnte sich nämlich vollumfänglich auf die Plicht der zu versichernden Person, alle relevanten Tatsachen vorzubringen, abstützen (m.w.H. Thouvenin 2019). Es könnte jedoch argumentiert werden – die Freiwilligkeit sei gegeben –, dass Betroffene auf andere Anbieter ausweichen könnten, welche ähnliche Angebote ohne Einsatz der fraglichen Technologien anbieten (Thouvenin 2019, S. 41; EDÖB 2015). Die Ausdrücklichkeit der

[131] Art. 6 Abs. 6 und 7 nDSG.

[132] Im Zusammenhang mit der Individualisierung von Versicherungsverträgen allg.: Thouvenin 2019.

[133] Welche davon ausgehen, dass wirtschaftliche Nachteile die Freiwilligkeit nicht ausschliessen.

Einwilligung verpflichtet zur aktiven Zustimmung des Betroffenen in die Datenbearbeitung (Maurer-Lambrou und Honsell 2014, S. 16–17). Ein ausbleibender Widerspruch nach dem Hinweis einer automatisierten Ansage, dass zum Vertragsabschluss eine Stimmanalyse via Telefon zur Erkennung für allfälliger Krankheiten durchgeführt werde, würde demnach nicht genügen.

3.7.5.4. Selbstdiagnose[134]

Applikationen im Gesundheitsbereich sind in zwei Kategorien einzuteilen: einerseits Apps, die dem Bundesgesetz über Arzneimittel und Medizinprodukte unterstehen, und anderseits Apps im Bereich Wellness/Fitness/Lifestyle, die keine medizinische Zweckbestimmung verfolgen und damit auch nicht spezifischen Regulierungen unterworfen sind (Klett 2017, S. 105; Gordon 2016; Isler 2019, S. 49; Muresan 2021; Klett und Verde 2016; Fuchs und Giovanettoni 2013).[135]

Das Bundesverwaltungsgericht prüft anhand von drei Kriterien, ob eine App den Begriff eines Medizinprodukts erfüllt (Isler 2019, S. 46–50; Leins-Zurmühle 2021, S. 139):[136] (1.) verfolgt die App eine medizinische Zweckbestimmung, (2.) erzeugt sie mittels Datenbearbeitung medizinische Angaben und (3.) sind diese Angaben auf eine bestimmte Person zugeschnitten.

1. Eine Software mit medizinischem Zweck liegt gemäss Art. 3 MepV vor, wenn diese bspw. zur Diagnose, Verhütung, Überwachung, Vorhersage, Prognose, Behandlung oder Linderung von Krankheiten eingesetzt werden kann.
2. Ebenfalls ist die Erzeugung oder Modifikation medizinischer Angaben zu evaluieren. Ausschlaggebend dabei ist, ob die App nicht bloss zur Datensammlung, Speicherung, Kommunikation oder Übertragung von Daten verwendet werden kann, sondern auch bspw. Berechnungen durchführt.
3. Die dritte Voraussetzung, der Generierung spezifischer Daten für nur einen bestimmten Nutzer, wird wohl regelmässig am meisten Schwierigkeiten bereiten. Gibt ein Nutzer Informationen über sich selbst ein und die App schlägt spezifische Handlungen vor, ist es ein Medizinprodukt (Isler 2019, S. 50; Leins-Zurmühle 2021).

Im Bereich des Medizinprodukterechts gilt das Prinzip der Selbstkontrolle durch den Produzenten, ob ein Medizinprodukt den regulatorischen Anforderungen entspricht. Bei gewissen Klassen von Medizinprodukten ist eine Prüfung durch eine akkreditierte Konformitätsbewertungsstelle oder das Erfüllen von anderweitigen Anforderungen nötig (Leins-Zurmühle

---

[134] Für eine vertiefte Auseinandersetzung mit den rechtlichen Aspekten der Selbstvermessung siehe Meidert et al. 2018, 177 ff.

[135] Bundesgesetz über Arzneimittel und Medizinprodukte (Heilmittelgesetz, HMG) vom 15. Dezember 2000, SR. 812.21; BVGer C-699/2016 vom 17. September 2018 E. 6.2.1; EuGH C-329/16 vom 7. Dezember 2017.

[136] BVGer C-699/2016 vom 17. September 2018 E. 5.3.

2021).[137] Als problematisch erachtet wird, dass Entwickler und Unternehmen, welche Apps entwickeln, mit den regulatorischen Anforderungen für Medizinprodukte wenig vertraut sind (Prieur 2017, S. 43; Isler 2019, S. 49). Für die Haftung eines fehlerhaften Produkts, welches zu einer Körperverletzung, zum Tod einer Person oder zu einem Sachschaden führt, ist das Produktehaftpflichtgesetz massgebend (m.w.H. Klett und Verde 2016; Lüchinger 2019; Landolt 2017, S. 100–102).

Der Verantwortliche müsste sich ebenfalls an die Datenbearbeitungsgrundsätze halten. Es handelt sich bei Nutzern in der Regel um bestimmbare Personen, da für das Herunterladen ein Konto benötigt wird. Insbesondere muss den betroffenen Nutzern klar sein, zu welchem Zweck die Daten bearbeitet, allenfalls an Dritte weitergegeben oder mit weiteren Daten verknüpft werden. Werden Daten an Dritte weitergegeben, muss die Einwilligung ausdrücklich sein (Prieur 2017, S. 71). Insbesondere ist auch die Datensicherheit (Verschlüsselung der Daten) zu beachten (Prieur 2017, S. 63). Auch sind im Blick auf das Verhältnismässigkeitsprinzip nur diejenigen Daten zu erheben, die für den bestimmten Zweck nötig sind, und nicht weitere Informationen zu erheben, um diese etwa weiterzuverkaufen (Prieur 2017, S. 77). Für Nutzer ist es nur schwer erkenntlich, ob Personendaten lokal oder bspw. auf einer Cloud (im Ausland) gespeichert werden. Ebenfalls kritisch einzuordnen ist, dass Nutzer durch das Akzeptieren der AGB möglicherweise dem Weiterverkauf ihrer Personendaten an Drittunternehmen zustimmen (Gordon 2016, S. 71). Auch ist im Einzelfall abzuklären, ob Gesundheitsdaten oder biometrische Daten, die eine Person eindeutig identifizieren, erhoben würden und damit besonders schützenswerte Daten vorliegen (Vokinger 2020, Rz. 7 f.).

3.7.6. Gesellschaftliche und ethische Herausforderungen

Hinsichtlich des Einsatzes von Stimm-, Sprach- und Gesichtserkennungstechnologien zur Erkennung physischer und psychischer Krankheiten stellen sich eine Reihe ethischer und gesellschaftlicher Herausforderungen.

Diese lassen sich grob in drei Ebenen unterteilen. Kurz- bis mittelfristig werden derartige Technologien voraussichtlich (1.) v.a. als Unterstützungstools für ärztliches Personal einerseits zum Einsatz kommen und (2.) für die Selbstdiagnose andererseits. Langfristig (3.) sind v.a. die Auswirkungen einer zunehmenden Perfektionierung derartiger Technologien auf das Gesundheitswesen und die individuelle Gesundheit zu erwarten. Im Mittelpunkt der folgenden Diskussion stehen überwiegend kurz- und mittelfristige Aspekte.

3.7.6.1. Selbstbestimmung – Nutzen oder Überforderung?

Einerseits kann die mittels Stimm-, Sprach- und Gesichtserkennungstechnologien personalisierte Medizin zur Erkennung von Krankheiten eine Verbesserung der Gesundheitsversorgung und einen Gewinn an Autonomie darstellen: Die Nutzung entsprechender Apps durch ärztliches Personal könnte zu einer Verbesserung der ärztlichen Diagnose- und Therapiemöglichkeiten führen. Therapien könnten besser auf die je individuellen gesundheitlichen

[137] Art. 23 f. MepV i.V.m. Art. 52 und Anhang IX MDR.

Bedingungen hin ausgerichtet und somit verbessert werden. Eine auf diese Weise verbesserte Diagnosestellung und Therapierung könnte auch zu einem Effizienzgewinn führen, in deren Folge Kapazitäten für andere erforderliche Tätigkeiten frei werden (Martinez-Martin 2019; Fiske et al. 2019). Zugleich könnten Interessierte durch die Nutzung entsprechender Gesundheitsapps selbst an der (Früh-)Erkennung von Krankheiten mitwirken. Ebenso könnten sie in die technologiegestützten Überwachungsprozesse ihrer Therapie eingebunden werden (Prainsack 2013, S. 30–31).

3.7.6.2. Algorithmische Diskriminierung und Validierung: Datenqualität und Vermeidung von Bias

Wie bei anderen Technologien, die mit automatisierten Datenbearbeitungen operieren, stellt sich auch bei der Erkennung von physischen und psychischen Krankheiten mittels Stimm-, Sprach- und Gesichtserkennungstechnologien das Problem der algorithmischen Diskriminierung. Sowohl der verwendete Algorithmus könnte einem Bias bzw. einer Fehlprogrammierung unterliegen als auch könnten die zugrunde liegenden **Trainingsdaten Verzerrungen** aufweisen. Die im Rahmen der vorliegenden Studie durchgeführten Recherchen deuten darauf hin, dass seitens der Hersteller der Vermeidung von Bias in Algorithmen im Medizinbereich grosse Bedeutung beigemessen wird (Martinez-Martin 2019, S. 2). Dies dürfte nicht zuletzt daran liegen, dass viele Anwendungen von Forschungsinstitutionen entwickelt werden (Universität St. Gallen 2019).

Die unzureichende **Validierung** derartiger Software könnte allerdings fatale Fehler nach sich ziehen, indem bspw. Korrelationen als Kausalitäten gedeutet werden (Ienca und Ignatiadis 2020, S. 83). Besonders problematisch könnte das Daten-Bias-Problem im Hinblick auf seltene Erkrankungen sein. Falls eine wachsende Zahl an Menschen an der Krankheitserkennung und behandlung mittels Stimm-, Sprach- und Gesichtserkennungstechnologien mitwirken sollte, könnten – wie in einigen Forschungsprojekten praktiziert – Erkenntnisse hins. seltener Erkrankungen erzielt werden, die aufgrund der geringen Patientenzahl und den ökonomischen Kosten auf klassischem medizinischen Wege nur schwer möglich wären. Andererseits werden beide Faktoren mit der Einführung der neuen Erkennungstechnologien nicht verschwinden. Vielmehr ist zu erwarten, dass ökonomische Faktoren auch weiterhin ausschlaggebend sein werden. Sobald der Markt für Krankheitserkennungstechnologien von privatwirtschaftlichen Anbietern dominiert wird (Frost & Sullivan 2021), wird sich diese Frage in verschärfter Weise stellen.

3.7.6.3. Kaum Kontrollen

Aufgrund der Schwierigkeiten bei der Entwicklung verlässlicher Anwendungen zur Erkennung von physischen und psychischen Krankheiten wird auf die besondere Bedeutung von **unabhängigen Validierungs- bzw. Zertifizierungsverfahren** verwiesen (You und Gui 2020). Angesichts der grossen Menge an unterschiedlichen Apps (Meyer-Lindenberg 2018, S. 865) setzt das für die Zertifizierung von Medizinprodukten zuständige Schweizerische Heilmittelinstitut Swissmedic auf Stichprobenkontrollen und ist ansonsten auf Meldungen hins. möglicher Verstösse auf andere Stellen angewiesen. Eine systematische Kontrolle

einer jeden App danach, ob es sich bei ihr um ein Medizinprodukt handelt, erfolgt also nicht. Interessierte Nutzerinnen und Nutzer sind weitestgehend auf Onlineratgeber angewiesen, um durch den unübersichtlichen Markt zu navigieren (Verbraucherzentrale.de 2020).

Der Algorithmus «DeepGestalt» wird in einer App zur Gendefekterkennung namens «Face2Gene» eingesetzt. Diese App wird laut Face2Gene (2021) in mehreren (Kinder-)Krankenhäusern weltweit eingesetzt. Konkrete Informationen über den Einsatz fehlen jedoch. Laut Peter Krawitz von der Universität Bonn sei die Software zurzeit bei etwa der Hälfte aller Humangenetiker bekannt und werde teilweise bereits im klinischen Alltag eingesetzt (MDR 2019). Eine Bewertung der Sicherheit solcher Systeme ist kaum möglich, da sehr vieles unbekannt ist und auch Forschungspublikationen wenig Informationen über verwendete Algorithmen preisgeben. Ebenso wenig ist deren Quellcode öffentlich, sodass unabhängige Institutionen diesen nicht auf Sicherheitslücken prüfen können. Unbekannt beim DeepGestalt-Algorithmus sind auch die verwendeten Trainingsdaten. Bekannt ist nur, dass 500.000 Gesichtsbilder von 10.000 verschiedenen Personen als Trainingsdaten herangezogen wurden. Diese stammen aus dem Jahr 2014 und wurden offenbar dem Internet entnommen (Ars Technica 2019). Eine Begutachtung der Prozeduren ist keine Anforderung an ein Medizinprodukt in der Schweiz (Swissmedic 2021).

3.7.6.4. Erosion des Solidaritätsprinzips und Ausgrenzung von Patienten durch ökonomische Rationalisierung

Maio (2012) bezeichnet die mögliche Ausgrenzung von Patienten durch Rationalisierung als eine mögliche Gefahr personalisierter Medizin. Demnach sei zu erwarten, dass eine Behandlung mit geringer Erfolgswahrscheinlichkeit in dem Masse schwieriger zu rechtfertigen sein dürfte, je genauer und repräsentativer die Daten werden, die mittels technologiegestützter Vorabuntersuchungen erhoben werden. In diesem Zusammenhang werde die Solidargemeinschaft vor die Herausforderung gestellt, neue Arrangements treffen zu müssen, bis zu welcher Wahrscheinlichkeit eine teure Therapie noch übernommen wird. An zusätzlicher Brisanz gewinnt dieses Argument, wenn berücksichtigt wird, dass zwar eine Besserung der Datenlage für verbreitete Krankheiten zu erwarten ist, dies für sehr seltene Erkrankungen jedoch möglicherweise nicht der Fall sein wird, weil es sich schlicht um sehr seltene Krankheiten handelt, zu denen nicht viele Daten vorliegen können. Problematisch könnte es dann sein, wenn ein zu grosses Vertrauen in datengestützte Krankheitserkennung deren Schwachstellen nicht mitberücksichtigt und so seltene Erkrankungen, die nicht mittels Software erkannt werden, der ökonomischen Rationalisierung anheimfallen (ebd.).

3.7.6.5. Abbau von (personellen) Kapazitäten

Die Hoffnung, dass die durch Technologieeinsatz erzielten Effizienzgewinne zu grösseren personellen Kapazitäten in medizinischen Einrichtungen und damit z.B. zu einer verbesserten Betreuungssituation führen, könnte sich in ihr Gegenteil verkehren. Sollte die Technologie zuverlässig funktionieren, könnten Effizienzgewinne auch als Argument zum Stellenabbau oder für anderweitige Einsparungen dienen – was durchaus erwünscht sein kann. Problematisch wäre es hingegen insb. in den Bereichen, in denen auch weiterhin eine

intensive menschliche Betreuung erforderlich ist oder wenn die durch Technologie erzielten Verbesserungen der Gesundheitsfürsorge durch den Abbau von Kapazitäten relativiert würden (Fiske et al. 2019).

3.7.6.6. Druck zur Mitwirkung

Auch die Rolle und die Verantwortung von Patientinnen und Patienten könnte sich im Zuge der Verbreitung app-gestützter Krankheitserkennungsmethoden verändern. Wenn die Möglichkeit der ständigen Erfassung krankheits- und behandlungsrelevanter Daten besteht, könnten Patienten dem impliziten Druck ausgesetzt werden, an der Früherkennung und Therapierung mitzuwirken. Besondere ethische Relevanz entfaltet dieser Druck im Hinblick auf das Recht auf Nichtwissen. So könnten Patienten, die keine dauerhafte Selbsttestung durchgeführt haben, im Falle einer ärztlichen Diagnose der Kritik ausgesetzt werden, sie hätten eine Früherkennung verhindert. So könnte die freiwillige Entscheidung zur Nutzung solcher Apps der zunächst moralischen Pflicht weichen, sie nutzen zu müssen. Zusammengeführt mit dem vorgenannten Argument der ökonomischen Rationalisierung, könnten individuelle «Versäumnisse» als Triebfeder einer entsolidarisierten Gesundheitsversorgung wirken, in der die wahrnehmbaren Leistungen vom Grad der eigenen Mitwirkung am Gesundheitserhalt abhängen und ein Nichtwissen-Wollen nicht mehr toleriert wird. Dieser ökonomische und soziale Druck würde zudem das zentrale (medizin-)ethische Prinzip der Selbstbestimmung aushöhlen (ebd. 18).

3.7.6.7. Individuelle Überforderung

Wenn die Möglichkeit individueller Kontrolle und der individuellen Mitwirkung an Therapien zur Pflicht wird, kann dies auch zu Überforderung führen. Weil es sich bei der Nutzung von Apps zur Krankheitserkennung um soziotechnische Systeme handelt, ist zwar davon auszugehen, dass die Nutzenden aller Voraussicht nach mit der Zeit ihre Nutzungskompetenzen ausbauen und sicherer in der Nutzung sein werden. Zugleich wird es aber auch immer Bevölkerungsgruppen, (sehr alte, gebrechliche, demente, obdachlose, an schwerwiegenden psychischen Erkrankungen leidende Menschen) geben, die nicht oder nur in sehr eingeschränktem Masse in der Lage sein werden, derartige technische Werkzeuge zu verwenden. Andere Bevölkerungsgruppen, die gesundheitlich in der Lage dazu wären, könnten hingegen aufgrund mangelnder finanzieller und zeitlicher Kapazitäten (z.B. Alleinerziehende oder sozial schwach situierte Personen) in geringerem Masse in der Lage sein, derartige Technologien zu verwenden (Prainsack 2013, S. 30–31).

3.7.6.8. Breite Einwilligung

Eine weitere Diskussion aus dem Kontext der individualisierten Medizin, die auf den Einsatz von Stimm-, Sprach- und Gesichtserkennungstechnologien übertragen werden kann, betrifft das Thema «breite Einwilligung». Angesichts der im Vorfeld oft unvorhersehbaren Potenziale, die sich durch die Zusammenführung und Bearbeitung unterschiedlicher Daten ergäben, stosse das bisherige Konzept der für spezifische Bearbeitungszwecke vorgesehenen

Einwilligung an Grenzen. Befürworter von breiten Einwilligungskonzepten sehen darin die Möglichkeit, auf möglichst effiziente Weise den grösstmöglichen Nutzen aus der Nutzung (personenbezogener) Daten ziehen zu können. Indem Betroffene in angemessener Weise über künftige Verwendungszwecke informiert und die Einhaltung dieser Zwecke durch Ethikkomitees überwacht werden, würde schliesslich die Gefahr der missbräuchlichen Nutzung minimiert. Schwierigkeiten, wie eine angemessene Information auch weniger gebildeter Patienten möglich und wie mit der Möglichkeit einer nicht erwünschten Nutzung der bereitgestellten Daten umzugehen wäre, könnten durch entsprechende Schulung des Fachpersonals adressiert werden (Barazzetti et al. 2021). Eine Gewährleistung dieser Aspekte wäre jedoch im Falle des selbstdiagnostischen Einsatzes deutlich schwieriger. Da solche Apps nicht nur von staatlich begleiteten Forschungsprojekten, sondern auch von privatwirtschaftlichen Anbietern entwickelt werden, die keiner Kontrolle unterliegen, und der Markteintritt von Apps ebenfalls keine systematischen Bewilligungsprozesse durchläuft, wäre es fraglich, ob und inwiefern eine missbräuchliche Nutzung bzw. Weitergabe von (personenbezogenen) Daten noch gewährleistet wäre.

3.7.6.9. Fehlinterpretation von Softwareergebnissen – Fatale Selbstdiagnostik

Eine weitere Herausforderung besteht in der Gefahr von Fehlinterpretationen der Softwareergebnisse. So stellt die korrekte Interpretation KI-basierter Ergebnisse bereits für das Fachpersonal eine neue Hürde dar (Fiske et al. 2019, S. 4–5; Amann et al. 2020). Sollten entsprechende Apps zunehmend auch für Laien zugänglich werden, potenziert sich die Gefahr fataler Selbstdiagnostik weiter (Ghassemi et al. 2021). Zur Mitigation des Risikos der Fehlinterpretation wird sowohl an der Verbesserung der Nachvollzieh- bzw. Erklärbarkeit von Softwareergebnissen geforscht als auch gefordert, dass in verstärkter Weise auf interne und externe Validierungen gesetzt werden sollte (You und Gui 2020). Wie oben erwähnt, werden voraussichtlich nicht alle Teile der Bevölkerung in gleichem Masse in der Lage sein, derartige Apps korrekt einzusetzen, sodass sich die Frage stellt, wie mit einem solchen absehbaren tiefgreifenderen Kompetenzproblem umzugehen ist. In diesem Zusammenhang wird auch ein generelles Verbot von Selbstdiagnose-Apps diskutiert (Schmidt 2019).

3.7.6.10. Verzicht auf genauere, aber invasivere Untersuchungen

Zudem könnte die Möglichkeit, dass durch Stimm- und Gesichtserkennungssoftware Diagnosen nicht invasiv gestellt werden können, ein entscheidendes Argument für die Einwilligung der Betroffenen sein. Während für manche Diagnosen u.U. sogar eine Operation nötig wäre, könnte durch diese Technologie ein grösserer medizinischer Eingriff verhindert werden. Kritisch zu bewerten ist dabei, dass viele Patientinnen und Patienten kaum medizinisches Fachwissen haben, aufgrund der geringen Invasivität eher in derartige Diagnosemethoden einwilligen, sodass medizinisch ggf. erforderliche invasive Untersuchungen nicht durchgeführt würden. Dieser Kritikpunkt liesse sich allerdings dadurch relativieren, dass die Betroffenen durch den Vertrauensarzt über alle relevanten Aspekte der verschiedenen Untersuchungsoptionen angemessen aufgeklärt werden.

3.7.6.11. Langfristperspektive: Chatbots, KI-Agenten und automatisierte Diagnosen

Mittel- bis langfristig wird erwartet, dass KI-basierte Gesundheitsanwendungen nicht nur im Einzelanwendungskontext zum Einsatz kommen werden, sondern sich zunehmend auch zu einem Teil integrierter KI-Agenten entwickeln werden. Derartige integrierte Anwendungen würden einerseits die vorgenannten Herausforderungen verstärken, andererseits auch neue Herausforderungen generieren. Etwa die Frage, ob, ggf. ab welchem Punkt und wie ein KI-Agent weitere Stellen über das als Suizidabsicht erkannte Verhalten einer Person benachrichtigen sollte (Fiske et al. 2019, S. 5). Eine weitere langfristige Herausforderung wird zudem darin liegen, einen sinnvollen gesellschaftlichen Umgang mit der zunehmenden Perfektionierung von softwarebasierten Diagnosen und der dadurch zunehmend weniger relevant werdenden Diagnosestellung durch menschliche Fachkräfte zu finden (Ienca und Ignatiadis 2020).

3.7.7. Zwischenfazit

Die Erkennung von physischen und psychischen Krankheiten mittels Stimm-, Sprach- und Gesichtserkennung hat in den vergangenen Jahren bedeutende Fortschritte erzielt. Inzwischen ist eine kaum überschaubare Vielfalt an Forschungsprojekten und von der Privatwirtschaft betriebenen Initiativen zur produktiven Nutzung der Technologien im medizinischen Kontext vorzufinden.

Die technischen Herausforderungen gleichen grundsätzlich den generellen Herausforderungen beim Einsatz der Technologien. In Testumgebungen werden teils sehr hohe Trefferraten erzielt, doch fehlt es auch in diesem Bereich an Untersuchungen, die derartige Angaben unter realen Bedingungen validieren. Die Frage, inwiefern die hohen Trefferraten auch in der Praxis erreicht werden können, bedarf somit weiterer Untersuchungen.

Zusätzlich besteht bei der Krankheitserkennung eine grosse konzeptionelle Herausforderung im Zusammenhang mit der Definition von Biomarkern, anhand derer die Erkennung einer Krankheit erfolgen soll. Je nach Krankheit stellt sich hier mehr oder weniger die Frage, ob und inwiefern die identifizierten Marker tatsächlich zur Erkennung einer Krankheit geeignet sind. Eine weitere Herausforderung dabei besteht in der Gewinnung von brauchbaren Trainingsdaten, wobei die Erschaffung synthetischer Daten mittels GANs nicht immer eine sinnvolle Alternative darstellen muss.

Aus rechtlicher Sicht ist es bei der Nutzung entsprechender Anwendungen seitens medizinischen Fachpersonals wichtig, gegenüber Patientinnen und Patienten Transparenz zu gewährleisten, um eine informierte Einwilligung zu ermöglichen, ohne die eine medizinische Behandlung grundsätzlich nicht möglich wäre.

Der Einsatz in Bewerbungskontexten ohne Wissen der Bewerber ist widerrechtlich und kann mit zivilrechtlichen Rechtsmitteln angefochten werden. Grundsätzlich möglich wäre er hingegen, wenn die Einwilligung des Betroffenen eingeholt wird. Allerdings muss die Freiwilligkeit der Einwilligung bezweifelt werden, weil die Bewerberinnen und Bewerber regelmässig vor die Entscheidung gestellt würden, der Bearbeitung entweder einzuwilligen oder

zu befürchten, aufgrund ihrer Nicht-Einwilligung schlechter abzuschneiden bzw. abgelehnt zu werden.

Eine Verwendung von Diagnosesoftware durch Krankenversicherer würde mit der Verletzung mehrerer Datenschutzgrundsätze einhergehen (Verhältnismässigkeit, Datenminimierung, Treu und Glauben). Möglich wäre sie nur mit der (ausdrücklichen) Einwilligung der Betroffenen, sofern ein Vertragsabschluss unter gleichen Bedingungen auch dann möglich wäre, wenn keine Einwilligung erfolgt.

Im Falle der stärkeren Verbreitung von Krankheitserkennungs-Apps fürs medizinische Fachpersonal oder zur Selbstdiagnose stellt sich zunächst insb. die Herausforderung der Gewährleistung ihrer technischen Zuverlässigkeit. Angesichts der grossen Zahl an Apps könnte die stichprobenartige Kontrolle der Swissmedic an Grenzen stossen.

Sollten die Apps Eingang in Diagnosestellungsprozesse finden, könnte sich zudem eine Reihe weiterer Herausforderungen stellen: etwa in Form der Erosion des Solidaritätsprinzips bzw. der Ausgrenzung von Patienten durch ökonomische Rationalisierung, des Abbaus von (personellen) Kapazitäten, die für die Gesundheitsfürsorge wichtig wären, eines Drucks zur Mitwirkung bei der Diagnosestellung, individuelle Überforderung bei der Nutzung der Apps, der Einholung einer breiten Einwilligung, der Gefahr der Fehlinterpretation von Softwareergebnissen und insb. fataler Selbstdiagnosen sowie des Verzichts auf genauere, aber invasivere Untersuchungen. Darüber hinaus würden sich mittel- bis langfristig weitere Herausforderungen stellen, sobald Krankheitserkennung in KI-Agenten oder Chatbots integriert wird und Diagnosestellung zunehmend nicht mehr durch Menschen erfolgt, sondern von automatisierter Software übernommen wird.

3.8. Emotionserkennung

Emotionserkennung wird in der Psychologie bereits seit Beginn des 20. Jahrhunderts untersucht (James 1913; Fuchs 2014). Seit den 1980er-Jahren wird das Thema auch in den Computerwissenschaften erforscht (Dewan et al. 2019; Cowie et al. 2001). Emotionserkennung ist ein Prozess, in dem aus audiovisuellen Signalen Anzeichen für emotionale Zustände abgeleitet werden (Fasel und Luettin 2003). Je nach Eingangssignal kann zwischen visueller und akustischer Emotionserkennung unterschieden werden. In der Literatur ist dies auch als Affective Computing bekannt (Bösel 2019; Picard 2000). Hier stellt das Kodierungsverfahren «Facial action coding system» (FACS) von Ekman (1978) den weltweit verbreiteten Standard zur Beschreibung von Gesichtsausdrücken dar.

Die möglichen Anwendungen der Emotionserkennung sind vielfältig: in medizinischen Anwendungen, bei der Müdigkeitserkennung von Autofahrern, in Computerspielen, bei Anrufen in Callcentern und bei der Entwicklung von sozialen Robotern (Li und Deng 2020, S. 1; Saxena et al. 2020, S. 54). Während der Marktwert für Emotionserkennung 2019 bei USD 17,1 Mrd. lag, wird eine jährliche Zuwachsarte von 18 % bis 2027 erwartet (MarketWatch (2021).[138] Während EU-Polizeibehörden die Emotionserkennung bei der Einreise am Flug-

[138] Zu den bekannteren Firmen zählen HireVue (2021), Affectiva (2021) oder die retorio GmbH (2021).

hafen zur Lügendetektion testeten (Hodgson 2019), werden Anwendungen im Handel zur Verhaltensanalyse genutzt. So wird die Technik insb. zur Analyse der Empfänglichkeit eine Person für bestimmte Werbebotschaften genutzt, indem ein Zufriedenheitswert berechnet wird. Diesen Wert nutzen Verkäufer, um ein gewünschtes (Kauf-)Verhalten hervorzurufen (sog. Neuromarketing) (Seng und Ang 2018; Hamelin et al. 2017; Bruhn et al. 2020). Anwendung finden die Systeme, trotz erheblicher Kritik, auch in Bewerbungsgesprächen (Wolfangel 2018b; Brennan 2020; Pluta 2019; Crawford et al. 2019; Richter 2019). Dort sollen Nervosität, Empathie und Zuverlässigkeit aus den Gesichtsausdrücken der Bewerbenden abgeleitet und somit die Entscheidungsfindung bei der Einstellungsentscheidung unterstützt werden (Hagerty und Albert 2021).

Obwohl es technisch einfach ist, Gesichtsbewegungen zu erkennen, ist deren Deutung unter Psychologen umstritten (Crawford 2021, S. 167), zumal es bereits an einer einheitlichen Definition des Emotionsbegriffs fehlt (Scherer 1996, S. 298). Barrett et al. (2019) kamen zu dem Schluss, dass es keine wissenschaftliche Unterstützung für die Annahme gibt, dass der emotionale Zustand einer Person leicht aus ihren Gesichtsbewegungen abgeleitet werden kann. Hagerty und Albert (2021) schlussfolgern, dass Emotionserkennung auf «wackeligen wissenschaftlichen Grundlagen» steht. Die Schwierigkeiten der Erkennung zeigen sich auch daran, dass selbst die Erkennungsleistung von Menschen bereits sehr unterschiedlich und abhängig von der jeweiligen «emotionale Kompetenz» (Scherer et al. 2018, S. 358), dem Alter und Geschlecht (Juen et al. 2012) und verschiedenen gesellschaftlichen und kulturellen Hintergründen ist (White 2017). Vielfach wird auch der Unterschied zwischen Emotionen und Gefühl nicht klar dargelegt (Scherer 2005, S. 695). So argumentierte James Williams bereits 1913, dass sich aus Gesichtsausdrücken oder Änderungen der Mimik nicht direkt auf Emotionen schliessen lässt (James 1913). Diese These wurde in den nachfolgenden Jahrzehnten mehrfach bestätigt (Cacioppo und Tassinary 1990; Wolfangel 2018b). Emotionen entstehen «aus dem Zusammenspiel körperlicher Reaktionen und im Gedächtnis gespeicherter Erfahrungen» (Barrett 2017) und müssen immer im Kontext der Situation bewertet werden (Scherer 1996, S. 307). Beispielsweise zeigten Choleriker immer negativere Emotionen. Gewisse Mimiken können auch doppeldeutig sein. So kann ein Stirnrunzeln sowohl Ärger und Wut als auch Überraschung zeigen (Wolfangel 2018b). Dementgegen gilt die Emotion «Scham» als vergleichsweise eindeutig erkennbar (Wolfangel 2018b).

Erschwerend für die Erkennung kommt hinzu, dass Emotionen meist nur ein paar Sekunden dauern und keinen klar abgrenzbaren Anfang (onset) und Ende (offset) aufzeigen (Hess und Kleck 1990, S. 370). Zudem sind verschiedene Arten von «Emotionen» (bspw. Vorlieben, Einstellungen, Gemütslage) unterschiedlich intensiv und dauerhaft (Scherer 2005, S. 704). Des Weiteren beschreibt der sog. Cross-Race-Effekt, dass Menschen aus unterschiedlichen Kulturkreisen mit Schwierigkeiten bei der gegenseitigen Interpretation ihrer Emotionen konfrontiert sind (Elfenbein und Ambady 2002; Scherer 1996, S. 317). V.a. bei der Emotion «Langeweile» gab es zwischen Kanadiern, Deutschen, Kolumbianern und Chinesen die grössten Unterschiede bei den wahrgenommenen physiologischen Merkmalen (Loderer et al. 2020, S. 1480).

Ein Bereich, der traditionell auf Emotionserkennung setzt, ist die Lügenerkennung. In einer Studie aus dem Jahr 1991 wurde untersucht, wie gut unterschiedliche Personengruppen (Richter, Psychiater, der amerikanische Secret Service, Studenten, N=509) Lügen von drit-

ten Personen erkennen können. In einem Video wurden zehn Personen gezeigt, die entweder logen oder die Wahrheit sagten. Es zeigte sich, dass lediglich die Agenten des Secret Service eine Lüge in etwas mehr als 50 % der Fälle korrekt erkannten (Ekman und O'Sullivan 1991). Da dies nur etwas mehr als Zufall ist, wollten Forschende aus verschiedenen Disziplinen die Lügen- und Emotionserkennung automatisieren. Daraus ist das Forschungsgebiet der «Automated Facial Expression Analysis» (AFEA) hervorgegangen. Neuere Methoden erreichen nun bereits eine Erkennungsrate von bis zu 99 % (Saxena et al. 2020, S. 53). Insb. die Kombination von «Partikelschwarmoptimierung», die biologisches Schwarmverhalten als Vorbild nutzt, mit der BES-Datenbank, die acht Emotionen von Sprachaufnahmen von zehn deutschen Sprechern beinhaltet, erreicht diese Höchstwerte (Saxena et al. 2020, S. 65). Wie in den vorangegangenen Anwendungsbereichen auch, geben die Autoren jedoch keine Informationen darüber, ob diese Raten auch ausserhalb von Laborbedingungen erreicht werden. Ekman et al. (2011) behaupten, dass unter Laborbedingungen eine Lüge durch einen Menschen anhand der Mimik zu 80 % und unter Hinzuziehung von Körperbewegungen und Stimme bis zu 90 % korrekt erkannt werden kann.

Durch den fortschreitenden Einsatz von Sprachassistenzsystemen, wie smarten Lautsprechern, stehen umfangreiche neue Trainingsdaten für KI-Systeme zur Verfügung (Schiller und McMahon 2019, S. 182; Fedotov et al. 2019). Daher rückt nun auch die Emotionserkennung mittels Sprache in den Fokus der Forschung. Die Stimme verrät vieles über die Emotionen einer Person – und dies laut Studien besser als Mimik (Bruhn et al. 2020, S. 625; Wolfangel 2018a). So zeigte Kraus (2017) in einer Studie mit fast 1800 Teilnehmern, dass Menschen Emotionen besser anhand der Stimme erkennen (MStimme=1,02) als durch die Kombination mit einem Gesichtsausdruck (MStimme/Gesicht=1,36, kleinere Werte spiegeln eine höhere Korrektheit wider). Als möglicher Grund wurde angeführt, dass es Menschen schwerer falle, innere Zustände nicht in der Stimme zum Ausdruck zu bringen. Gesichtsausdrücke liessen sich leichter manipulieren, da Menschen darauf sozial trainiert würden (Herrmann 2017).[139] Die Erklärung für die Verschlechterung der Erkennung bei der Kombination von Gesicht und Stimme erklärt er damit, dass zwei kognitiv anspruchsvolle Aufgaben gleichzeitig Menschen überfordern würden. Eine Studie aus dem Jahr 1993 zeigte bei der Emotionserkennung anhand von Gesichtern, dass die Beurteiler in rund 50 % der Fälle die richtige Emotion erkannten[140] (Scherer 1996, S. 319–320). In einem Experiment mit Opernsängern zeigten Scherer et al. (2017, S. 1), dass es robuste Stimmsignaturen für die Unterschiede von Traurigkeit und Zärtlichkeit sowie Wut, Freude und Stolz gibt. Zudem existierten «Hinweise dafür, dass stimmlich ausgedrückte Emotionen über die Grenzen von Sprache und Kultur hinweg erkennbar bleiben» (Scherer 1996, S. 320). Weiter zeigten Halbauer und Klarmann (2019), dass beim Einkauf mittels Sprache über Alexa die Stimmung der Kunden in dem Szenario mit einer Genauigkeit von 72,7 % korrekt erkannt wurde.

Doch Emotionserkennung kann auch unabhängig von Gesichtsmimik oder Stimmlage erfolgen. Cui et al. (2015, S. 1) nutzen Daten von Beschleunigungsmessgeräten, die an Armen

---

[139] Anders verhält es sich wiederum bei den Augenbewegungen, die ebenfalls viel über die inneren Zustände eines Menschen verraten können. Die Augen zählen zwar zum Gesicht, aber anders als bei der bewussten Manipulation von Gesichtsausdrücken sind sich viele Menschen über die Möglichkeiten der Interpretation ihrer Augenbewegungen nicht im Klaren (Hoppe et al. 2018).

[140] Das ist mehr als Zufall, da zehn Emotionen zur Auswahl standen. Zufallsdurchschnitt wären 10 %.

und Beinen der Teilnehmenden befestigt waren, um Emotionen aus dem Gang einer Person abzuleiten. Wut (85 %), Neutralität (78 %) und Fröhlichkeit (78 %) konnten dabei erkannt werden. Etwas bessere Erkennungsraten (91,3 %, 85,5 % und 88,5 %) mithilfe von Daten von smarten Armbändern erreichten Zhang et al. (2016, S. 1). Weil die bewusste Verfälschung der Gangart als schwierig gilt, wird den dabei erfassten Daten eine geringere Fehleranfälligkeit zugesprochen (Xu et al. 2020b, S. 1). Zudem können weitere Faktoren wie eye tracking, Bewegungen von Körperteilen, Parameter der Augen und generelle Gesten und Körperhaltungen ebenfalls zur Erkennung von Emotionen herangezogen werden (Dewan et al. 2019, S. 16–17). Auch aus der Interaktion von Nutzenden mit dem Smartphone, z.B. Touch-Verhalten, sind Rückschlüsse auf die Emotionen möglich (Simonazzi et al. 2021). Emotionserkennung kann somit auch ohne Wissen der Betroffenen erfolgen.

3.8.1. Technische Grundlagen und Möglichkeiten

Generell kann zwischen vier Methoden zur Emotionserkennung unterschieden werden, die von den jeweiligen Inputdaten abhängen (Saxena et al. 2020, S. 53): (1) Stimm- oder (2) Videomaterial, (3) Text oder (4) sog. multimodale Systeme, bei denen zusätzliche physiologische Faktoren (z.B. die Pulsrate) mit einbezogen werden (Li und Deng 2020, S. 1; Imani und Montazer 2019, S. 2; Scherer und Moors 2019, S. 727). Grundsätzlich können mit den Methoden aktuell sieben Emotionen erkannt werden: glücklich, neutral, Ekel, traurig, ängstlich, überrascht und wütend (Saxena et al. 2020, S. 53).

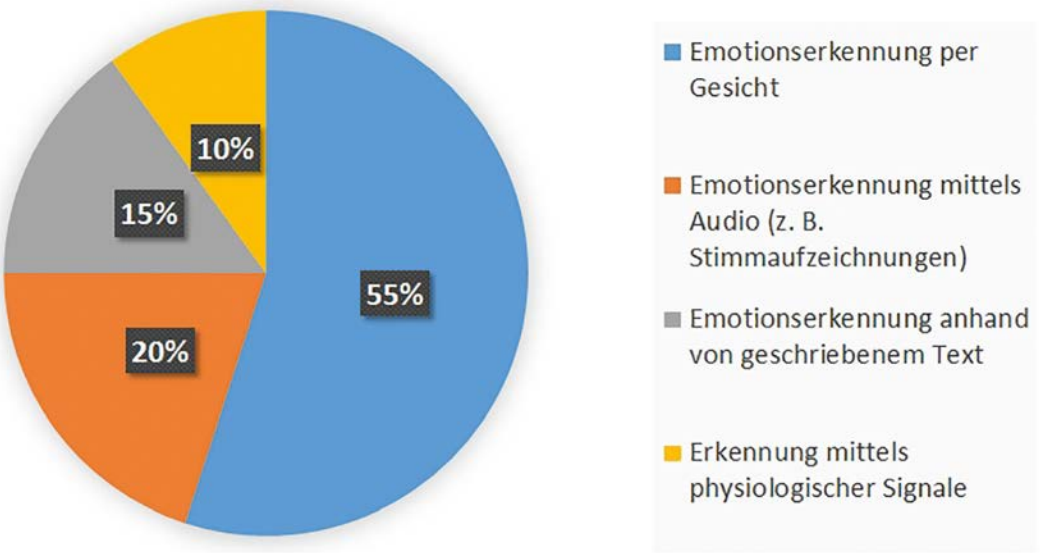

Abbildung 10: Verteilung der Anwendungen zur Emotionserkennung nach Art der Daten (Saxena et al. 2020, S. 73)

Gemeinsam haben alle Methoden, dass Trainingsdaten für die Algorithmen weder in optimaler Qualität (ähnliche Alters-, Geschlechts- und Kulturverteilung) noch in ausreichender Menge zur Verfügung stehen (Li und Deng 2020, S. 2). Dennoch gibt es Bestrebungen, die Datenbasis weiter auszubauen.

Im Bereich der visuellen Erkennung wurde in den 1990er-Jahren auf selbst erstellte Datensätze mit wenigen Einträgen gesetzt (Wu et al. 2012b, S. 393). Erst seit 2013 kamen erste Wettbewerbe unter Forschenden auf, wie «FER2013» (Goodfellow et al. 2013) und «Emotion Recognition in the Wild» (EmotiW) (Dhall et al. 2017), die eine gemeinsame Datenbasis nutzen. Seit einigen Jahren setzen Forschende zudem auf KI, um Datensätze zu generieren und zu annotieren (Li und Deng 2020, S. 1; Rouast et al. 2019). Verbesserungen der Daten werden stetig weitergeführt (Li und Deng 2020, S. 2).

3.8.1.1. Emotionserkennung anhand von Videomaterial

Die Technik der Erkennung von Emotionen anhand von Videomaterial hat in der Wissenschaft mehrere Begriffe, so insb. «facial expression recognition» (FER) und «Automated Facial Expression Analysis» (AFEA), die nur im Detail unterschiedlich, beim grundsätzlichen Vorgehen allerdings weitestgehend identisch sind. Die ersten Schritte beziehen sich stets darauf, ein Gesicht in Bild- oder Videomaterial zu erkennen (vgl. Kapitel 2.1.1). Daran schliessen sich drei weitere Schritte an: (1.) die Vorverarbeitung, (2.) das Erlernen von Eigenschaften sowie (3.) die Klassifikation der Eigenschaften. Während der Schritte 2 und 3 wird maschinelles Lernen mit künstlichen Netzen (Deep Networks) genutzt. Abbildung 11 zeigt diese Schritte.

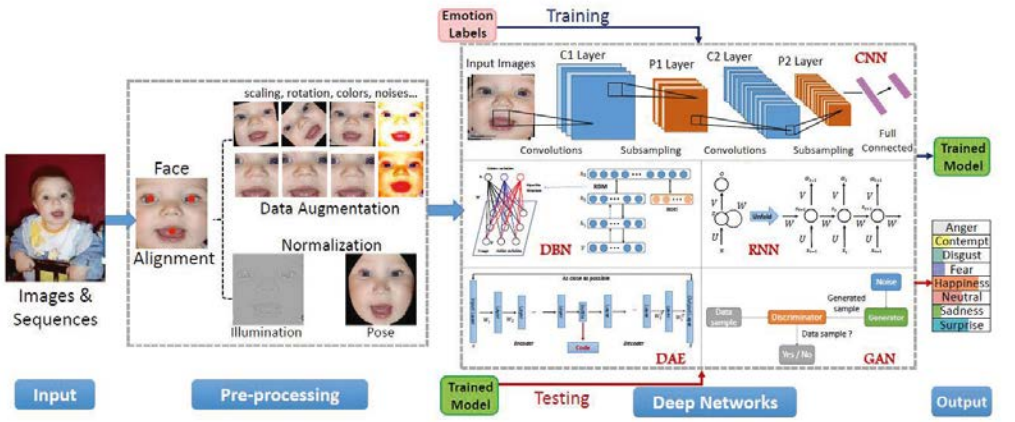

Abbildung 11: Schritte zur Erkennung von Emotionen anhang von Videomaterial (Li und Deng 2020, S. 4)

1. **Vorverarbeitung** (pre-processing): Hierbei wird das erkannte Gesicht gedreht, Helligkeiten und Kontraste werden angepasst und Hintergründe entfernt. Hierfür wird oft der sog. «Viola-Jones face detector»-Algorithmus verwendet (Viola und Jones 2001).
2. **Eigenschafts-Erlernen** (deep feature learning in deep networks): Hier wird versucht, Abstraktionen auf hoher Ebene durch hierarchische Architekturen mehrerer nicht linearer Transformationen und Darstellungen zu erfassen (Li und Deng 2020, S. 5). Hierfür wird eine Vielzahl von unterschiedlichen Methoden genutzt und verkettet. Vereinfacht ausgedrückt, läuft das Verfahren wie folgt ab: Sobald ein Gesicht und dessen Teile er-

kannt wurden, werden sog. «facial feature localization points» (FFLP) definiert und auf das Bild «gelegt» (vgl. Abbildung 12).

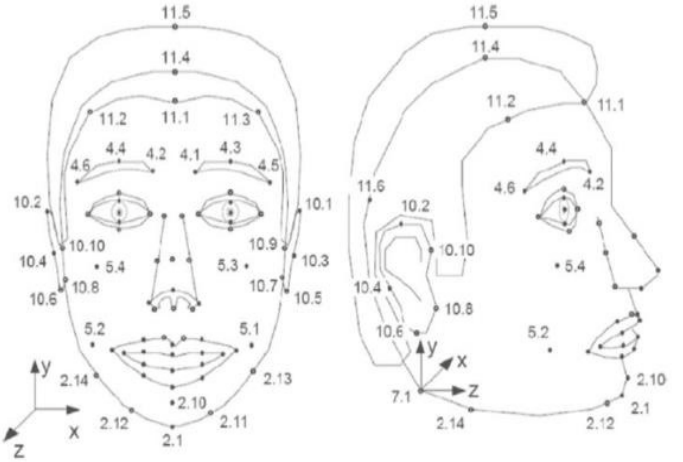

Abbildung 12: Facial feature localization points (Imani und Montazer 2019, S. 19)

Im nächsten Schritt werden die Abstände der Punkte zueinander gemessen. Bei Standbildern wird für die Analyse die sog. «deformation extraction» verwendet. Hierbei wird das Bild mit dem Gesicht in neutraler Mimik verglichen. Dafür müssen mehrere Bilder der Person vorliegen (Le et al. 2012). Die «motion extraction» kann bei Videoaufzeichnungen oder Bilderserien genutzt werden (Fasel und Luettin 2003). Hier werden die Abstände der Punkte über den Zeitverlauf analysiert (Le et al. 2012).

3. **Eigenschaftsklassifikation** (deep feature classification in deep networks): Im finalen Schritt wird versucht, das Bild einer Kategorie von Emotionen zuzuordnen. Diese kann dann mithilfe eines Notationssystems verglichen werden (Bruhn et al. 2020, S. 411). Hier stellt das Kodierungsverfahren «Facial action coding system» (FACS) von Ekman (1978) den weltweit verbreiteten Standard zur Beschreibung von Gesichtsausdrücken dar. In deren erster Version wurden sechs verschiedene Gesichtsausdrücke beschrieben: Wut, Ekel, Angst, Glück, Traurigkeit und Überraschung.

   Schlussendlich wird für jede der sechs Grundemotionen eine Wahrscheinlichkeit ermittelt. Im Fall der Abbildung 12 ist das dargestellte Kind mit höchster Wahrscheinlichkeit glücklich, mit etwas geringerer Wahrscheinlichkeit überrascht.

   Angesichts neuerer Erkenntnisse, dass Gesichtsausdrücke durchaus von der Kultur der Person abhängen (Jack et al. 2012), werden die Verfahren stetig verbessert und verändert. So stellt das «F-M FACS 3.0» eine Weiterentwicklung dar. Es bietet über 4000 Video- und Bildsequenzen in 4K-Auflösung und teils in 3D sowie Gesichtsbewegungen («Action Units») und 22 Zungenbewegungen (Freitas-Magalhães 2018).

Erschwert wird die Emotionserkennung durch eine Reihe von Hindernissen. Abgesehen von der bereits oben erwähnten konzeptionellen Diskussion über die Ableitung von Emotionen aus Mimik, sind die Algorithmen überfordert, wenn sich der Kopf während einer Mimik stark

dreht oder bewegt. In diesen Fällen können die Algorithmen keine Eigenschaften extrahieren (Dewan et al. 2019). Lee et al. (2019) analysierten mehrere Emotionserkennungssysteme und zeigten, dass die korrekte Erkennungsrate zwischen ca. 39 % und ca. 70 % lag. Wird jedoch der Nutzungskontext, die Art der Interaktion oder der Ort mit einbezogen, steigt die Erkennungsrate auf bis zu 75 % (Lee et al. 2019, S. 10147). CNN, die aus mehreren «Schichten» bestehen, zeigen bei der Emotionserkennung in Bildern die grösste Genauigkeit, selbst bei weniger Inputdaten (Mukhopadhyay et al. 2020, S. 109). Hier liegt die Fehlerrate bei 0,23%. Seit 2012 gibt es keinen Algorithmus, der besser abgeschnitten hat (Ciresan et al. 2012, S. 3644). Dieser Wert wurde jedoch für Gesichtserkennung und nicht explizit für Emotionserkennung erreicht. Auch kann ein sog. «Segmentation Error» vorkommen. In diesem Fall haben die Algorithmen falsche Punkte und Ausschnitte des Gesichts für die Analyse ausgewählt und erreichen daher schlechtere Erkennungsraten (Dewan et al. 2019, S. 16). Forscher der University of Southern California untersuchten in einem Experiment zudem, wie Nutzer auf Geldverluste reagieren. Die Emotionsdetektoren zeigten hier sehr schlechte Erkennungsraten. Eine quantitative Bewertung wurde zwar nicht vorgelegt, doch konnten die Forschenden zeigen, dass ein Lächeln sehr viele unterschiedliche Emotionen bedeuten kann: Die getesteten Systeme konnten ein aufgesetztes Lächeln weder von einem hämischen, ironischen oder unglücklichen Lächeln unterscheiden, noch von einem glücklichen Lächeln (Preto 2019). Zudem sind die untersuchten Algorithmen lediglich in der Lage, Grundemotionen zu erkennen. Bei der Erkennung von Mischformen oder komplexen Zusammenhängen hingegen versagen sie noch (Mukhopadhyay et al. 2020, S. 109).

3.8.1.2. Emotionserkennung anhand von Tonaufnahmen

Um Emotionen in Sprache erkennen zu können, sind gute Eingangsdaten mit möglichst wenig Nebengeräuschen wichtig (Imani und Montazer 2019, S. 16). Nach einer Vorverarbeitung (wie in Kapitel 2.1.3 beschrieben) folgen die Merkmalsextraktion und -klassifikation (Saxena et al. 2020, S. 60). Dies kann mithilfe sehr unterschiedlicher Vorgehen erfolgen, etwa mittels eines CNN (Mohammed et al. 2020). Weiter werden Discrete Wavelet Transform models, Anchor Models, Vector Space Modeling, Gaussian Mixture Models und Hybrid Models eingesetzt (Saxena et al. 2020, S. 60–61; Imani und Montazer 2019, S. 14).

Für die Erkennung des emotionalen Zustandes werden Stimmlage, Tempo, Intonation, prosodische Eigenschaften (z.B. Wort- und Satzakzente), physikalische Grössen und andere, nicht inhaltsrelevante Daten hinzugezogen (Saxena et al. 2020, S. 73). In der direkten Interaktion können auch Daten darüber, wie oft Person A einer anderen ins Wort fällt, einen Rückschluss auf Emotionen zulassen.

Unter Laborbedingungen erreichen moderne neuronale Netze, wie CNN, eine Genauigkeit von bis zu 99,47 % (Saxena et al. 2020, S. 61). Ältere Systeme bewegten sich zwischen 44 % und 94 % (Saxena et al. (2020, S. 61–62).

Unter Realbedingungen zeigen sich jedoch noch schwerwiegende Probleme bezüglich korrekter Erkennung. Deshalb werden bei manchen Anwendungen, z.B. in der Robotik, einfachere und weniger präzise Modelle zur Emotionserkennung verwendet. Sobald z.B. eine

Person ihre Mundwinkel nach oben zieht oder den Satz «ich bin fröhlich» spricht, wertet das System dies als «Freude» (Timms 2016, S. 706).

Andere Forschende entwickeln derweil privatheitsfreundliche Spracherkennungssysteme zur Interaktion mit IoT-Geräten (Aloufi et al. 2019). Hierzu klingt sich das System als Zwischenebene zwischen die Nutzer und Cloud-Dienste und «entemotionalisiert» die Sprachdaten, bevor diese zur Bearbeitung weitergeleitet werden.

3.8.2. Juristische Bewertung

Eine eindeutige rechtliche Bewertung von Emotionserkennungstechnologien lässt sich nur mit Blick auf konkrete Anwendungsfälle durchführen. Dennoch lassen sich vorliegend einige allgemeine rechtliche Einschätzungen vornehmen, die bei der Verwendung solcher Technologien zu berücksichtigen sind.

Bei der Emotionserkennung werden Stimm-, Sprach- und Gesichtsdaten von natürlichen Personen verwendet. Um besonders schützenswerte Personendaten i.S.v. Art. 5 lit. c nDSG handelt es sich dabei nur, wenn diese Daten (auch) zur eindeutigen Identifikation dieser Personen verwendet werden oder wenn auch Rückschlüsse auf die Gesundheit der betroffenen Person möglich werden, etwa wenn damit auch psychische Krankheiten diagnostiziert werden könnten.

Klarerweise fällt die Erkennung von Emotionen hingegen datenschutzrechtlich unter die Definition von «Profiling» (Art. 5 lit. f nDSG), denn ein solches liegt vor bei jeder Art der automatisierten Bearbeitung von Personendaten, die darin besteht, dass diese Daten verwendet werden, um bestimmte persönliche Aspekte oder Merkmale der Person zu bewerten oder vorherzusagen.[141] Dazu gehört nach der gesetzlichen Definition auch die Analyse oder Vorhersage des Verhaltens, was bei der Emotionserkennung u.E. erfüllt ist. U.U. kann auch ein Profiling mit hohem Risiko (Art. 5 lit. g nDSG) vorliegen, wenn nämlich das Profiling zu einer Verknüpfung von Daten führt, die eine Beurteilung wesentlicher Aspekte der Persönlichkeit einer natürlichen Person erlaubt. Somit sind die erhöhten datenschutzrechtlichen Anforderungen bei Profiling resp. Profiling mit hohem Risiko einzuhalten.

3.8.2.1. Die Datenschutzgrundsätze bei der Emotionserkennung

Die Einhaltung der Datenschutzgrundsätze hängt erheblich vom konkreten Anwendungsfall ab. An dieser Stelle werden deshalb lediglich allgemeine, einzelfallunabhängige Erwägungen angeführt. Der Grundsatz von Treu und Glauben umfasst die *Transparenz* der einzelnen Datenbearbeitungsschritte (Rosenthal 2020, Rz. 35).[142] Gemäss diesem Grundsatz müssen die Erhebung der Daten und deren Zweck für die betroffene Person ersichtlich sein. Wird eine Emotionserkennung ohne das Wissen der betroffenen Person eingesetzt, liegt ein schwerwiegender Verstoss gegen das Transparenzprinzip vor. Sowohl das Bearbeiten von

[141] So auch Bundesrat, Botschaft nDSG, BBl 2017 7022; siehe auch Art. 4 Abs. 4 DSGVO.

[142] Art. 6 Abs. 3 nDSG.

Daten sowie dessen Zweck (Erkennung von Emotionen) müssen der betroffenen Person zwingend mitgeteilt werden.

Der Grundsatz der *Zweckbindung* schreibt vor, dass Personendaten nur für einen bestimmten und für die Betroffenen erkennbaren Zweck beschafft und nur auf eine mit diesem Zweck vereinbare Art und Weise bearbeitet werden dürfen.[143] Diesem Grundsatz würde es widersprechen, wenn Daten, die zu einem anderen Zweck erhoben wurden (z.B. Sprachabdrücke bei der Bedienung eines smarten Lautsprechers oder bei telefonischen Kundendienstgesprächen), auch zwecks Emotionserkennung verwendet werden.

Der Grundsatz der *Datenrichtigkeit*[144] bezieht sich nicht nur auf die Inputdaten, sondern auch auf die aus diesen Daten abgeleiteten Schlussfolgerungen über die Emotionen einer Person. Angesichts der bestehenden wissenschaftlichen Ungenauigkeiten der Emotionserkennungstechnologie ist die Einhaltung dieses Grundsatzes infrage gestellt, wenngleich es auch hier auf den Einzelfall ankommen wird.

Selbstverständlich müssen auch die *sonstigen datenschutzrechtlichen Vorschriften* beachtet werden, namentlich die Gewährleistung der Betroffenenrechte (Recht auf Information, Auskunftsrecht, Recht auf Datenherausgabe oder -übertragung sowie Berichtigungsrecht), die Meldung von Verletzungen der Datensicherheit, die Vorschriften betreffend Datenübertragung ins Ausland sowie die Anforderungen an privatheitsfreundliche Technologie (privacy by design and by default).

3.8.2.2. Verwendung von Emotionserkennung durch Behörden

Wie oben erwähnt, gibt es Einsatzszenarien, bei denen die Emotionserkennung durch Behörden verwendet wird, bspw. in Form von Lügendetektoren bei der Einreise oder bei polizeilichen Vernehmungen. Auch die Aufmerksamkeitserkennung in der Schule (die im nächsten Kapitel näher behandelt wird) kann darunterfallen.

Die Verwendung von Emotionserkennungstechnologie stellt einen Eingriff in mehrere *Grundrechte* dar. So wird namentlich das Recht auf Privatsphäre (Art. 13 Abs. 1 BV) im Sinne eines «right to be let alone» (Warren und Brandeis 1890, S. 193) tangiert, zudem auch das Recht auf Schutz der persönlichen Daten (Art. 13 Abs. 2 BV), das Informationen mit bestimmbarem Bezug zu den psychischen Eigenschaften einer natürlichen Person schützt (Diggelmann 2015). Zur Frage, ob maschinelle Emotionserkennung durch staatliche Akteure auch in den Schutzbereich des Rechts auf persönliche Freiheit, insb. in seinem Teilgehalt des Rechts auf geistige Unversehrtheit (Art. 10 Abs. 2 BV) fällt, bestehen in der Rechtsprechung derzeit keine Anhaltspunkte. Die Emotionserkennung kann auch einen *chilling effect* auf das Recht auf Meinungsfreiheit (Art. 16 BV) im Sinne der «freedom of thought» haben, denn dieses Grundrecht umfasst «die Gesamtheit der Mitteilungen menschlichen Denkens», unabhängig davon, ob diese rational fassbar oder emotional geprägt sind, und damit auch subjektive Standpunkte sowie Empfindungen (Hertig 2015b). Die beständige

[143] Art. 6 Abs. 3 nDSG.

[144] Art. 6 Abs. 5 nDSG.

Möglichkeit der Analyse der Gefühle durch eine staatliche Behörde kann auf betroffene Personen eine abschreckende Wirkung haben. Schliesslich kann auch das Diskriminierungsverbot (Art. 8 Abs. 2 BV) tangiert sein, wenn Emotionserkennungstechnologien eingesetzt werden, die bspw. bei Personen mit dunklerer Hautfarbe konsistent weniger zuverlässig ist und ihnen Emotionen zuschreibt, die durch rassistische Stereotypen geprägt sind, etwa wenn schwarze Menschen häufiger als aggressiv oder wütend eingeschätzt werden und damit das Klischee der «angry black woman» reproduzieren.

Grundrechtseingriffe müssen eine gesetzliche Grundlage haben, sich auf ein öffentliches Interesse stützen und verhältnismässig sein. Die Frage nach der gesetzlichen Grundlage kann nur bei Vorliegen eines konkreten Anwendungsfalles beantwortet werden – immerhin ist hier darauf hinzuweisen, dass schon aus datenschutzrechtlichen Gründen eine Grundlage in einem Gesetz im formellen Sinn notwendig wäre, da ein Profiling stattfindet (Art. 34 Abs. 2 lit. b nDSG) und vermutlich auch ein schwerwiegender Eingriff in die Grundrechte vorliegt (Art. 34 Abs. 2 lit. c nDSG). Ein öffentliches Interesse wird sich in Form der Aufklärung von Straftaten oder der Verhinderung illegaler Einreisen wohl zumeist finden lassen. Jedoch wirft die Verhältnismässigkeit des Einsatzes dieser Technologie Fragen auf, sind doch gerade schon aufgrund der (noch) existierenden hohen Fehleranfälligkeit die Geeignetheit wie auch die Notwendigkeit zweifelhaft, und auch die Zumutbarkeit im Einzelfall angesichts des doch schweren Eingriffs in die Grundrechte der betroffenen Person wird in jedem Anwendungsfall sehr genau zu prüfen sein.

3.8.2.3. Verwendung von Emotionserkennung durch Private

Anwendungsgebiete der Emotionserkennung durch private Akteure liegen bspw. in den Bereichen Werbung und Marketing, aber auch im Arbeitskontext bei Bewerbungsgesprächen. Diese Datenbearbeitungen stellen in der Regel *Persönlichkeitsverletzungen* dar, da sie, wie oben dargelegt, entgegen den Datenbearbeitungsgrundsätzen vorgenommen werden. Es kann in der Regel auch nicht davon ausgegangen werden, dass die betroffenen Personen ihre (Gesichts, Sprach- oder Stimm-) Daten bereits dadurch schon allgemein zugänglich gemacht haben,[145] dass sie sich in der Öffentlichkeit aufhalten.[146] Als Rechtfertigung für die Persönlichkeitsverletzung kommen die Einwilligung, ein überwiegendes privates oder öffentliches Interesse oder eine gesetzliche Pflicht in Betracht.[147] Wird die Emotionserkennung ohne das Wissen der betroffenen Person durchgeführt, liegt von vornherein keine Einwilligung vor. Bei Bewerbungsgesprächen müsste zudem auch die Freiwilligkeit einer erteilten Einwilligung angesichts der Machtasymmetrie hinterfragt werden. Ob ein überwiegendes Interesse am Einsatz der Emotionserkennungstechnologie vorliegt, muss im jeweiligen Einzelfall analysiert werden, jedoch ist angesichts des sehr schwerwiegenden Eingriffs in die Privatsphäre ein solches Interesse nur zurückhaltend anzunehmen. In vielen

[145] Art. 30 Abs. 3 nDSG.

[146] Vgl. dazu die Stellungnahme des EDÖB 2020b, wonach nicht von einem Zugänglichmachen ausgegangen werden kann, wenn die Gesichtsdaten der Betroffenen unter Missachtung des Transparenzgrundsatzes mit ungefragt beschafften Daten abgeglichen werden sollen.

[147] Art. 31 Abs. 1 nDSG.

Konstellationen dürfte es deshalb an einer ausreichenden Rechtfertigung für diese Art der Datenbearbeitung fehlen.

3.8.2.4. Gesamtbewertung

Auch wenn eine endgültige Bewertung nur in Bezug auf konkrete Anwendungsfälle möglich ist, verdeutlichen die hier angestellten allgemeinen Betrachtungen, dass die Anwendung von Emotionserkennungstechnologien sowohl durch staatliche wie auch durch private Akteure in sehr vielen, wenn nicht den meisten, Fällen mit den verfassungs- und datenschutzrechtlichen Vorgaben nicht in Einklang zu bringen sein wird. Der EDSA und der Europäische Datenschutzbeauftragte (EDSB) haben denn auch in ihrer gemeinsamen Stellungnahme zum Vorschlag der EU-Kommission für eine Verordnung zur Regulierung künstlicher Intelligenz (Europäische Kommission 01.04.2021) die Ansicht vertreten, dass die biometrische Kategorisierung von Personen eine Verletzung der Menschenwürde darstellt und folglich Emotionserkennung durch künstliche Intelligenz als «highly undesirable» zu sehen sei und grundsätzlich verboten werden müsste (EDPB-EDPS 2021).

3.8.3. Gesellschaftliche und ethische Herausforderungen

Der Einsatz von Stimm-, Sprach- und Gesichtserkennungstechnologien zur Erkennung von Emotionen wirft eine Reihe von grundlegenden ethischen Herausforderungen auf, die sich unabhängig von konkreten Anwendungsfällen stellen. Sofern konkrete Anwendungen in den Mittelpunkt der Diskussion rücken, stellen sich zusätzlich weitere anwendungsfeldspezifische ethische Herausforderungen. Beispielhaft wird im Kapitel 3.9 auf einen solchen Einsatz von Emotionserkennungssoftware zum Zwecke der Aufmerksamkeitsanalyse bei Schülerinnen und Schülern eingegangen.

3.8.3.1. Intransparenz der (algorithmischen) Grundlagen der Emotionserkennung

Eine grundlegende Herausforderung der Emotionserkennung resultiert aus deren weitreichender Intransparenz. Während die konzeptionellen Grundlagen bekannt sind und einige wissenschaftliche Studien Erkennungsraten und mögliche Bias überprüfen (Wiggers 2020), herrscht grosse Intransparenz aufseiten privatwirtschaftlicher Anbieter von Emotionserkennung. Einige Anbieter unterziehen zwar ihre Produkte freiwilligen Audits (z.B. HireVue), zu vielen anderen Anbietern gibt es allerdings keine Informationen. So lässt sich aktuell schwer bewerten, mit welcher Zuverlässigkeit die auf dem Markt erhältlichen Anwendungen Emotionen erkennen können.

3.8.3.2. Normative Zementierung nicht validierter Annahmen

Angesichts der anhaltenden wissenschaftlichen Debatte rund um die Zuverlässigkeit der Emotionserkennung birgt deren Einsatz die Gefahr der normativen Zementierung eines möglicherweise nicht validen Verständnisses über Emotionen. Dies betrifft erstens die Sig-

nale (Gesichtsausdrücke, Stimmlage, Sprachinhalt), die gewissen Emotionen zugeordnet werden. Zweitens betrifft es die Interpretation der auf diese Weise zugeordneten Emotionen im Hinblick darauf, was diese über die Persönlichkeit, inneren Zustände oder (Handlungs-) Absichten verraten. Schliesslich wirken Zuordnung und weiter gehende Interpretation normativ auf sich selbst zurück: Wenn sich eine konzeptionelle Vorstellung über die «korrekte» Deutung eines Gesichtssignals gegenüber anderen Verständnissen, die bspw. kulturelle und kontextuelle Faktoren berücksichtigen, durchsetzt, könnte daraus eine normierende Wirkung auf die Gesellschaft ausgehen. Das mögliche Fehlresultat der Software wäre dann Sache des Individuums, das nicht die «richtigen» Signale ausgesendet hat, und kein konzeptionelles oder technologisches Versagen (Stark und Hoey 2020, S. 6–7).

3.8.3.3. Folgen eines möglichen Anpassungsdrucks

Zusammenhängend mit dem vorgenannten Punkt könnte Emotionserkennung auch zu einem Anpassungsdruck aufseiten Betroffener führen. Weil davon auszugehen ist, dass der Einsatz von Emotionserkennung transparent erfolgen muss (vgl. Kapitel 3.8.2), könnten Betroffene sich angesichts des Wissens um deren Einsatz dazu gedrängt fühlen, sich auf die je bestmögliche Weise präsentieren zu müssen, um die Erwartungen des jeweiligen Kontexts zu erfüllen. Auf ähnliche Weise, wie Abschreckungseffekte staatlicher Überwachung zu konformistischem Verhalten führen, könnten die von Emotionserkennung Betroffenen daran gehindert werden, sich auf freie Weise zu verhalten.

3.8.3.4. Zunehmendes Machtgefälle zwischen Individuum und Datenbearbeitern

Andererseits ist davon auszugehen, dass sich viele Betroffene zwar aufgrund der Transparenzvorgaben und der Betreiberpflicht zur Einholung der Einwilligung im Klaren darüber wären, dass Emotionserkennung eingesetzt wird. Allerdings würden sie vermutlich nur geringes Interesse an den Details des Einsatzes zeigen und bzw. oder unzureichendes Wissen für ein weiter gehendes Verständnis darüber mitbringen (Rossnagel et al. 2020, S. 20–23). Sollte die informierte Einwilligung auf diese Weise im Kontext der Emotionserkennung zur blossen Fiktion verkommen, würden Datenbearbeitern im Falle einer funktionierenden Emotionserkennung Möglichkeiten eröffnet, weitgehende Erkenntnisse über die Betroffenen zu sammeln, die nicht in ihrem eigenen Interesse sein müssen. Dies würde das Machtgefälle zwischen Unternehmen und Individuen weiter zugunsten Ersterer kippen.

Das Problem des Machtgefälles könnte sich auch dadurch verschärfen, dass Betroffene annehmen, sie könnten die Emotionserkennung überlisten. Im Falle einer unzuverlässigen Software wäre dies auch durchaus möglich (Zloteanu et al. 2018; Ma et al. 2021). So könnten Personen dem Einsatz von Emotionserkennung zustimmen, um mittels Vortäuschung Vorteile für sich selbst zu erzielen. Im Falle einer technisch weitgehend perfektionierten Emotionserkennung könnten sie sich allerdings in falscher Kontrolle wähnen, obwohl dem Betreiber der vorsätzliche Täuschungsversuch klar wäre und sich aus der Erkennung des Täuschungsversuchs ein umso grösserer Nachteil für die Betroffenen ergäbe.

Hierbei stellt sich ein ethisches Dilemma: Die Lüge ist einerseits ein sozial unerwünschtes Verhalten, das schädliche Effekte haben kann (Jordan 2014). Andererseits würde der unkontrollierte Einsatz technisch perfektionierter Emotionserkennung Betroffene womöglich auch in Kontexten durchschaubar machen, in denen aus guten Gründen keine tieferen Einblicke in ihre Gefühlswelt möglich sein sollten (Koops et al. 2016, S. 528–530). Angesichts der Alltäglichkeit des Täuschens und sich stetig verbessernder Algorithmen sollte die Frage, wo welche Grenzen zu ziehen sind, Gegenstand der öffentlichen Debatte sein (Weber 2010).

3.8.3.5. Emotionserkennung als Teil konvergierender soziotechnischer Systeme

Mittel- bis langfristig planen Anbieter von Emotionserkennung, ihre Anwendungen mit weiteren Digitaltechnologien zu verknüpfen. So soll bspw. die bereits heutzutage in einigen vernetzten Fahrzeugen eingesetzte Müdigkeitserkennung um weitere Funktionen erweitert werden. Die Überwachung der Emotionen von Fahrgästen soll bspw. dazu genutzt werden können, die Lautstärke der Musik automatisch an die erkannten Emotionen anzupassen: Die Lautstärke soll z.B. erhöht werden, wenn die Software erkennt, dass die Musik den Fahrgästen gefällt (McManus 2018). Derartige Entwicklungen in Richtung konvergierender soziotechnischer Systeme könnten sowohl verstärkend auf die vorgenannten Herausforderungen wirken als auch zu neuen Herausforderungen führen (Stahl 2021, S. 42–44).

3.8.4. Zwischenfazit

Die Emotionserkennung stellt eine der neueren Anwendungsmöglichkeiten von Stimm-, Sprach- und Gesichtserkennung dar und öffentliche wie private Akteure planen bereits deren Einsatz in vielen Bereichen.

Neben Schwierigkeiten bei der praktischen Erkennungsleistung, die weitestgehend den generellen technischen Herausforderungen der untersuchten Technologien gleichen, zeigen sich bei der Emotionserkennung darüber hinausgehende grundsätzliche Bedenken hinsichtlich der wissenschaftlichen Validität der Erkennungsergebnisse. Anwendungen der Emotionserkennung greifen zwar auf das Konzept Paul Ekmans zurück bzw. auf Weiterentwicklungen des Konzepts, zahlreiche Forschende weisen allerdings auf die konzeptionellen Mängel dieser Klassifizierung hin und stellen deren Gültigkeit grundsätzlich infrage. Dass Unternehmen und teils auch Behörden angesichts dieser Probleme den Einsatz von Emotionserkennung dennoch vorantreiben, ist kritisch zu bewerten. Dieses Problem potenziert sich weiter durch die auch in diesem Anwendungsfeld fehlende Evaluation von Anwendungen der Technologie seitens unabhängiger Stellen.

Auch in rechtlicher Hinsicht bringt Emotionserkennung zahlreiche Herausforderungen mit sich: Sofern sie ordentlich funktioniert, würde sie weitgehende Rückschlüsse über die Persönlichkeit oder Gesundheit einer Person ermöglichen. Wenn hingegen eine fehlerhafte Erkennung durchgeführt wird, könnte der Grundsatz der Datenrichtigkeit verletzt werden. Abhängig von der jeweiligen konkreten Anwendung ist davon auszugehen, dass die Anwendung von Emotionserkennungstechnologien sowohl durch staatliche wie auch durch

private Akteure in sehr vielen, wenn nicht den meisten, Fällen nicht mit den verfassungs- und datenschutzrechtlichen Vorgaben in Einklang zu bringen sein wird.

Darüber hinaus bringt Emotionserkennung auch Herausforderungen für die Selbstbestimmung mit sich: Diese könnte sowohl dann verletzt sein, wenn die Technologie fehlerhaft funktioniert, als auch im Falle einer fehlerfreien Funktion. Die Ungewissheit der Betroffenen, ob und inwiefern ihre Emotionen erkannt und wozu genau entsprechende Erkenntnisse verwendet werden, könnte sie in schwierige Abwägungssituationen bringen. Insb. in Kontexten, die einem strukturellen Machtgefälle unterliegen, könnte sich die Nutzung von Emotionserkennungstechnologien nachteilig auf Betroffene auswirken – statt den Betroffenen Selbstbestimmung zu ermöglichen, könnte die Einholung ihrer Einwilligung dieses Machtgefälle gar weiter zementieren, wenn keine sehr weitgehende Klarheit darüber herrscht, wie die Emotionserkennung funktioniert. Wenn zudem Emotionserkennung zukünftig Eingang in weitere datenbearbeitende Systeme und somit in alltägliche Routinen findet, könnte sowohl die Unübersichtlichkeit für die Betroffenen zunehmen als auch die Quantität und Qualität der gesammelten Daten: Inwiefern Betroffene und die Gesellschaft die Möglichkeit der allgegenwärtigen Transparenz ihrer Gefühle zulassen wollen, sollte Gegenstand der öffentlichen Debatte sein.

3.9. Aufmerksamkeitsanalyse in Schulen

Die automatische Aufmerksamkeitsanalyse von Schülerinnen und Schülern im Klassenraum wird bereits seit den 1980er-Jahren erforscht (Dewan et al. 2019, S. 1). In westlichen Ländern steht dabei laut der Befürworter die Erkennung von Lernbarrieren im Vordergrund, um den Lernerfolg zu steigern (D'Mello 2017, S. 115). Eingebettet ist diese Art der Überwachung von Schülern häufig in die laufenden Digitalisierungsbestrebungen von Schulen, die neben der Aufmerksamkeitsanalyse teils auch weitere Überwachungsmöglichkeiten (etwa die Analyse von Prüflingen während Onlineklausuren mittels Webcam zur Betrugsverhinderung) vorantreiben (Stapf et al. 2021b, S. 323). Technisch basiert die Aufmerksamkeitsanalyse stark auf der insb. videobasierten Emotionserkennung, die jedoch um weitere Aspekte, wie die Blickrichtung, erweitert wird.

V.a. Schulen in den USA, Grossbritannien, Australien und in China setzen Technologien zur Aufmerksamkeitsanalyse ein oder testen diese (Andrejevic und Selwyn 2020, S. 115; Kuhn 2019). So gaben US-Schulen in der Vergangenheit jährlich durchschnittlich 2,7 Milliarden Dollar für Überwachungs- und Sicherheitsprodukte zur Verhinderung von Gewalt, etwa in Form von Amokläufen, aus (Doffman 2018). Für Ausgaben in Emotionserkennungssysteme lassen sich zwar keine Daten finden. Jedoch kann zu einer bestehenden Videoüberwachungsinfrastruktur mit geringem zusätzlichem Aufwand eine Aufmerksamkeitsanalyse hinzugefügt werden. Mit umgerechnet rund 195.000 Schweizer Franken pro Schule ist die Erweiterung einer vorhandenen Kamerainfrastruktur um ein System zur Aufmerksamkeitsanalyse zudem relativ kostengünstig (Yujie 2019).

Im Bereich der Aufmerksamkeitsanalyse gibt es vielfältige Anwendungsfälle. Ein Grossteil der Anwendungen zielt darauf ab, den Lernerfolg der Schüler im Klassenzimmer zu steigern (Andrejevic und Selwyn 2020; D'Mello 2017) und die Lernstrategie und Inhalte anzupassen

(Yang et al. 2018, S. 2). Üblicherweise wird damit argumentiert, dass Lehrkräfte im Unterricht nicht gleichzeitig alle Schülerinnen und Schüler im Blick haben können (Bouhlal et al. 2020) und mittels automatisierten Feedbacks an die Lehrkräfte deren Arbeit vereinfacht und somit die Qualität des Unterrichts verbessert werden könne (Timms 2016, S. 710).

In China werden solche Systeme auch dazu genutzt, Lehrkräfte zu überwachen (Gong et al. 2020). Als konkretes Beispiel kann die «Niulanshan First Secondary School» in Peking genannt werden. Dort wird seit 2017 ein Emotionserkennungssystem der Firma Hanwang Technology (ausserhalb von China als «Hanvon» bekannt) eingesetzt, über deren Einsatz die Schüler anfangs nicht informiert waren (Yujie 2019) und auch die Zustimmung der Eltern nicht erforderlich war (Kretschmer 2021). Die Systeme sind in jedem Klassenzimmer installiert und machen sekündlich ein Foto der Schüler. Zudem werden jeden Monat neue Gesichtsbilder der Schüler hinzugefügt, um physiologische Veränderungen direkt zu erfassen (ARTICLE 19 2021, S. 30). Die Kameras sind klein und ohne Vorwissen kaum identifizierbar. Das System zeigt dann bspw., wie lange die Schüler den Blick auf die Tafel gerichtet hatten, und errechnet auf dieser Basis einen Aufmerksamkeits-Score, den Lehrkräfte und Eltern einsehen können. Insgesamt können fünf Verhaltensweisen erkannt werden: *zuhörend*, *Fragen beantwortend*, *schreibend*, *mit anderen interagierend*, *schlafend*. Auch Fragen wie «Redet meine Tochter zu viel mit ihrer Tischnachbarin? Sollten wir sie trennen?» sollen durch die Auswertung beantwortet werden können (Yujie 2019). Zudem können in Gruppendiskussionen die Antworten der Schüler gefilmt und ausgewertet werden, um auch dort zu berechnen, wie aufmerksam sie in der Diskussion waren.

Ein ähnliches System namens «4Little Trees» wird in Hongkong entwickelt. Es ermittelt und bewertet die Emotionen der Kinder, während sie Hausaufgaben erledigen. Dieses System soll den Grad der Motivation erkennen und daraus mögliche Noten vorhersagen können (Crawford 2021, S. 167).

Grundsätzlich kann die Emotionserkennung zur Ermittlung des Aufmerksamkeitslevels auf allen Plattformen zum Einsatz kommen, bei denen Bilddaten von Schülern und Studierenden entstehen, auch in Onlinelernplattformen, wie dem seit Beginn der Coronapandemie von Millionen Schülern und Studierenden genutzten Moodle (Guillén-Gámez und García-Magariño 2014, S. 173; Magdin et al. 2016, S. 66). Ein weiterer Anwendungsfall sind Lernroboter (Imbernón Cuadrado et al. 2019). Hier gilt es als besonders wichtig, dass die Lernroboter die Emotionen der Kinder erkennen und deuten können, um die Interaktion mit diesen ansprechender zu gestalten (Nikopoulou et al. 2018, S. 104). Auch jenseits des engeren Lernkontexts wird Emotionserkennung an Schulen genutzt, etwa zur Erkennung der von Schülern ausgehenden Gewalt anhand der Analyse ihrer Sprache (Han et al. 2018; Andrejevic und Selwyn 2020, S. 120) oder zur Erkennung von Mobbingopfern (Iliou und Paschalidis 2011, S. 18; Gao und Ye 2019). Ein bedeutsamer Teil[148] der Forschung in diesem Bereich beschäftigt sich mit der Erkennung von psychischen Auffälligkeiten bei den Kindern, wie Autismus und ADHS (Mehmood und Lee 2017). Schliesslich war während der Coronapandemie vermehrt darüber diskutiert worden, wie durch den Einsatz von Video-

[148] Basierend auf einer quantitativen Literatursuche in Scopus im Mai 2021 mit Suchstring «TITLE-ABS-KEY ((emotion OR attention OR affec-tive AND state) AND (recognition OR detec-tion) AND (school OR student OR pupil OR child))» fielen fast 60 % der Treffer auf den Bereich der Medizin und Psychologie.

überwachung Schulöffnungen umgesetzt werden könnten (Jargon 2020). Neu entwickelte Algorithmen sollten es ermöglichen, die Einhaltung der Abstandsregeln und der Maskenpflicht zu überwachen (Heilweil 2020).

3.9.1. Technische Grundlagen und Möglichkeiten

Emotionserkennung kann aus technischer Sicht in drei verschiedene Methoden unterteilt werden: manuell, halb automatisch und vollautomatisch (Dewan et al. 2019, S. 3). Wie in Abbildung 13 zu sehen, ist die Erkennung mittels Gesichtsausdrücken und Stimme ein vollautomatischer Prozess, der keinerlei Beteiligung der Schüler erfordert und damit vollkommen verdeckt erfolgen kann.

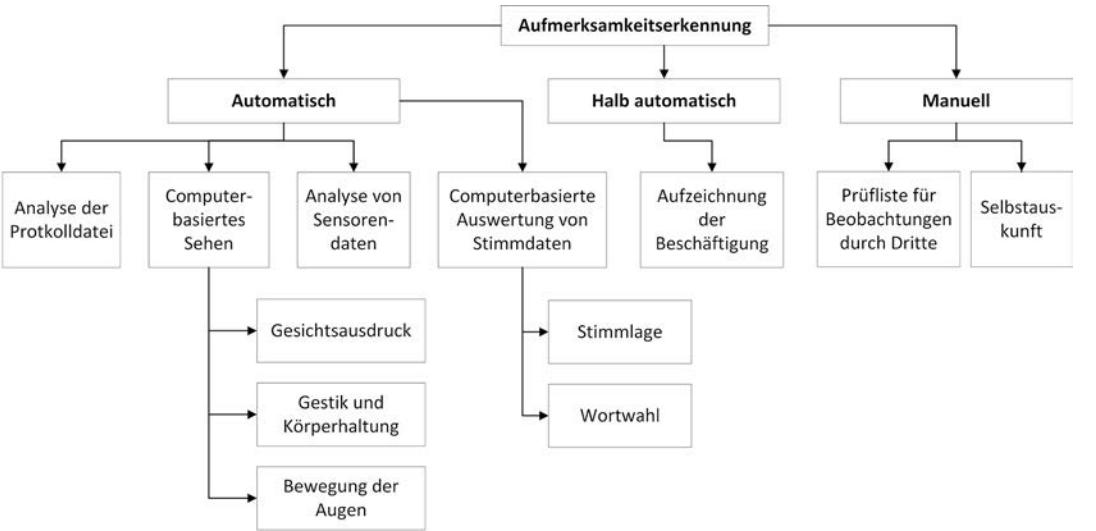

Abbildung 13: Taxonomie von Aufmerksamkeitserkennungssystemen (eigene Abbildung, Anlehnung an Dewan et al. 2019, S. 3)

Schulen sind aus mehreren Gründen technisch prädestiniert für den Einsatz von Gesichtserkennung: So werden viele Schulen in den USA, UK, Australien und China bereits heute videoüberwacht. Zudem ist es an vielen Schulen Tradition, Gesichtsfotos ihrer Schüler für Ausweise und Jahrbücher zu nutzen. Daraus lässt sich mit relativ wenig Aufwand eine Datenbank mit Namen und Gesichtern erstellen. Darüber hinaus ist die praktische Umsetzung der Technologie in Schulen als Institutionen mit relativ homogenen demografischen Merkmalen (Alter, Geschlecht und Ethnien) einfacher als in anderen öffentlichen Einrichtungen wie Krankenhäusern bzw. im öffentlichen Raum im Allgemeinen (Andrejevic und Selwyn 2020, S. 120). Dennoch gibt es auch hier die üblichen Probleme der Falscherkennung.

Grundsätzlich werden für die Aufmerksamkeitsanalyse die gleichen drei Bearbeitungsschritte wie bei der Emotionserkennung (Kapitel 3.8.1) durchgeführt. Als relevante biometrische Merkmale werden Augenbewegungen (eye tracking), Bewegungen von Körperteilen, Augenpartien, Gesten und Bewegungen genutzt (Dewan et al. 2019, S. 16–17). Als besonders aussagekräftig gilt dabei das Anheben der Augenbrauen, das Zukneifen der Augenlider so-

wie die Grübchenbildung neben dem Mund (Grafsgaard et al. 2013). Zudem soll die Erfassung von sog. Mikro-Emotionen, die weniger als eine halbe Sekunde dauern, Hinweise darauf geben können, was Schüler zu diesem Zeitpunkt denken (Liaw et al. 2014; Liaw et al. 2021). Da die Aufmerksamkeitsanalyse bislang v.a. in China eingesetzt wird, gibt es keine unabhängigen Evaluationen ihrer technischen Zuverlässigkeit.

3.9.2. Juristische Bewertung

Bislang wird, soweit ersichtlich, in Schweizer Schulen keine Aufmerksamkeitserkennungstechnologie verwendet. Einige Schulen haben zwar Videoüberwachungssysteme installiert, dies jedoch nach aktuellem Kenntnisstand nur in den Aussenbereichen (Pausenplätze, Veloständer etc.), nicht jedoch in den Klassenzimmern.[149] Die vorliegende rechtliche Analyse bezieht sich deshalb auf hypothetische Anwendungsfälle. Das Schulwesen ist in der Schweiz Sache der Kantone. Entsprechend sind – sofern es sich um öffentliche Schulen handelt – die kantonalen Datenschutzgesetze anwendbar, während Privatschulen unter das Datenschutzgesetz des Bundes fallen. Da hier hypothetische Anwendungsfälle behandelt werden und sich zudem die kantonalen Regeln zur Datenbearbeitung nur unwesentlich von denjenigen des Bundes unterscheiden, wird die vorliegende Analyse aus Einheitlichkeitsgründen anhand des Datenschutzgesetzes des Bundes vorgenommen, wobei, falls notwendig, auf kantonale Besonderheiten hingewiesen wird.

Bei der Aufmerksamkeitsanalyse werden Stimm-, Sprach- und Gesichtsdaten von Schülerinnen und Schülern verwendet. Um *besonders schützenswerte Personendaten* i.S.v. Art. 5 lit. c nDSG handelt es sich, wenn diese Daten (auch) Aufschluss über die Gesundheit der Betroffenen geben (Ziff. 2), etwa wenn damit auch psychische Auffälligkeiten erkannt werden sollen oder wenn sie zur eindeutigen Identifikation der Kinder verwendet werden (Ziff. 4). Da zur Aufmerksamkeitsanalyse notwendigerweise auch eine Zuordnung der Daten zu einzelnen Schülern notwendig ist und dies jeweils anhand des Gesichtsbildes oder des Sprachabdrucks erfolgt, liegt hier in jedem Fall eine eindeutige Identifizierung i.S.v. Art. 5 lit. c Ziff. 4 nDSG vor, es werden also bereits aus diesem Grund besonders schützenswerte Personendaten bearbeitet.

Zudem fällt die Aufmerksamkeitsanalyse in Schulen datenschutzrechtlich auch unter die Definition von «*Profiling*» (Art. 5 lit. f nDSG), da damit das Verhalten der betroffenen Schüler analysiert wird. Ggf. liegt sogar ein Profiling mit hohem Risiko (Art. 5 lit. g nDSG) vor, wenn das Profiling zu einer Verknüpfung von Daten führt, die eine Beurteilung wesentlicher Aspekte der Persönlichkeit der Schülerinnen und Schüler erlaubt. Somit sind die erhöhten datenschutzrechtlichen Anforderungen bei Profiling resp. Profiling mit hohem Risiko einzuhalten.

[149] Vgl. bspw. SRF vom 1. Februar 2019, «Kontrolle vs. Verantwortung – Schulhäuser rüsten bei der Videoüberwachung auf».

3.9.2.1. Die Datenschutzgrundsätze bei der Aufmerksamkeitsanalyse

Der Grundsatz von Treu und Glauben umfasst die *Transparenz* der einzelnen Datenbearbeitungsschritte (Rosenthal 2020).[150] Gemäss diesem Grundsatz muss die Erhebung der Daten und deren Zweck für die betroffenen Schülerinnen und Schüler sowie deren Eltern ersichtlich sein. Würde die Aufmerksamkeitsanalyse ohne das Wissen der Schülerinnen und Schüler und ihrer Eltern eingesetzt, läge ein schwerwiegender Verstoss gegen das Transparenzprinzip vor. Gleiches gilt, wenn ohne das Wissen der Lehrkräfte auch deren Fähigkeiten oder Verhalten analysiert würden. Die Standorte von Videokameras und Mikrofonen müssen signalisiert werden.

Ob der Grundsatz der *Verhältnismässigkeit* bei der Bearbeitung von Personendaten eingehalten ist, ist im Einzelfall zu prüfen (Epiney und Nüesch 2015).[151] Die Datenbearbeitung muss für die Erreichung des verfolgten Zwecks geeignet, erforderlich und zumutbar sein. Die Geeignetheit der automatisierten Aufmerksamkeitsanalyse ist im Lichte der (noch) existierenden hohen Fehleranfälligkeit fraglich (dazu schon oben zur Emotionserkennung, 3.8.2). Es stellt sich auch die Frage, ob nicht andere, mildere Mittel möglich wären. Weiterhin erscheint die «traditionelle» Aufmerksamkeitsüberwachung durch die Lehrpersonen im Klassenzimmer als geeignetes milderes Mittel. Auf jeden Fall ist die Zumutbarkeit dieser Form der Datenbearbeitung stark infrage zu stellen, da es sich um einen schweren Eingriff in die Privatsphäre handelt, welcher bei Minderjährigen auch noch einmal strenger zu beurteilen ist als bei Erwachsenen.

Der Grundsatz der *Datenrichtigkeit*[152] bezieht sich nicht nur auf die Inputdaten, sondern auch auf die aus diesen Daten abgeleiteten Schlussfolgerungen über die Emotionen oder Aufmerksamkeit der Schülerinnen und Schüler. Wie bereits im Kontext der Emotionserkennung angemerkt, muss die Einhaltung dieses Grundsatzes angesichts der bestehenden wissenschaftlichen Ungenauigkeiten der Emotionserkennungstechnologien infrage gestellt werden, wenngleich es auch hier auf den Einzelfall ankommen wird.

Selbstverständlich müssen auch die sonstigen datenschutzrechtlichen Vorschriften beachtet werden, namentlich die Gewährleistung der Betroffenenrechte (Recht auf Information, Auskunftsrecht, Recht auf Datenherausgabe oder -übertragung sowie Berichtigungsrecht), die Meldung von Verletzungen der Datensicherheit, die Vorschriften betreffend Datenübertragung ins Ausland sowie die Anforderungen an privatheitsfreundliche Technologie (privacy by design and by default).

3.9.2.2. Aufmerksamkeitserkennung in öffentlichen Schulen

Die Verwendung von Aufmerksamkeitserkennungstechnologien stellt einen Eingriff in mehrere Grundrechte dar. Dabei sind die gleichen Grundrechte betroffen wie bereits bei der Emotionserkennung (3.8.2) ausgeführt, also das Recht auf Privatsphäre (Art. 13 Abs. 1 BV),

[150] Art. 6 Abs. 3 nDSG.

[151] BGE 138 II 345 E. 9.2.

[152] Art. 6 Abs. 5 nDSG.

das Recht auf Schutz der persönlichen Daten (Art. 13 Abs. 2 BV), das Recht auf Meinungsfreiheit (Art. 16 BV) sowie das Diskriminierungsverbot (Art. 8 Abs. 2 BV).

Hinzu kommt, dass es sich bei den Betroffenen in aller Regel um Minderjährige handelt, sodass auch das *Recht von Kindern und Jugendlichen auf Schutz* (Art. 11 BV) tangiert wird. Gemäss dieser Garantie haben Kinder und Jugendliche Anspruch auf besonderen Schutz ihrer Unversehrtheit und auf Förderung ihrer Entwicklung. Mit Art. 11 BV erhält die vorrangige Beachtung des Kindeswohls (Art. 3 Kinderrechtskonvention) Verfassungsrang.[153] Art. 11 BV stellt nach herrschender Lehre kein eigenständiges Grundrecht dar, sondern erhöht den Schutzbereich der anderen Grundrechte, wenn es sich bei den Betroffenen um Minderjährige handelt (Müller und Schefer 2008, S. 806; Biaggini 2017), insb. wenn diese Rechte für die Persönlichkeitsentwicklung von Kindern und Jugendlichen wichtig sind (Müller und Schefer 2008, S. 807; Reusser und Lüscher 2014, N 14). Daneben wirkt sich Art. 11 BV auch auf die Rechtfertigung eines Grundrechtseingriffs aus: Die gesetzliche Grundlage für einen Eingriff muss bei Minderjährigen erhöhten Anforderungen genügen und die spezifischen Schutzbedürfnisse von Kindern und Jugendlichen sind bei der Abwägung im Rahmen der Verhältnismässigkeit zu berücksichtigen (Müller und Schefer 2008, S. 806).

Grundrechtseingriffe müssen eine gesetzliche Grundlage haben, sich auf ein öffentliches Interesse stützen und verhältnismässig sein. Die Verwendung einer Aufmerksamkeitsanalysetechnologie in Schulen bedürfte somit einer (kantonalen) *gesetzlichen Grundlage*. Allein schon aus datenschutzrechtlichen Gründen wäre eine Grundlage in einem Gesetz im formellen Sinn notwendig, da besonders schützenswerte Personendaten bearbeitet werden und ein Profiling stattfindet[154] und möglicherweise auch ein schwerwiegender Eingriff in die Grundrechte stattfindet[155], wobei Letzteres im Einzelfall spezifisch geprüft werden müsste. Allenfalls ergeben sich weitere Vorgaben aus dem kantonalen Recht, so sehen z.B. einige Kantone besondere Voraussetzungen für die Verwendung von Videoüberwachungstechnologien vor.[156]

Das *öffentliche Interesse* kann in der Verbesserung des Lernfortschritts der Schülerinnen und Schüler gesehen werden, in gewissen Fällen (abhängig vom Einzelfall) dürften auch weitere Interessen wie etwa Gewaltprävention vorliegen. Jedoch wirft die Verhältnismässigkeit des Einsatzes dieser Technologie Fragen auf (dazu bereits oben); im Vergleich zum Einsatz bei Erwachsenen ist die Verhältnismässigkeit, und insb. die Zumutbarkeit, bei Minderjährigen aufgrund von Art. 11 BV strenger zu beurteilen. Dabei ist auch zu beachten, dass das Wissen um die konstante Überwachung nicht nur den Lernfortschritt stören kann und damit gerade den intendierten Effekt verfehlen könnte und zudem voraussichtlich zu einer Verhaltensanpassung der Schülerinnen und Schüler führen würde, die mit ihrem Recht

---

[153] BGE 132 III 359 E. 4.4.2; 126 II 377 E. 5d.

[154] Art. 34 Abs. 2 lit. a und b nDSG.

[155] Art. 34 Abs. 2 lit. c nDSG.

[156] Gemäss § 17 des basel-städtischen Gesetzes über die Information und den Datenschutz (Informations- und Datenschutzgesetz, IDG) bspw. darf Videoüberwachung, bei welcher Personen identifiziert werden können, nur zum Schutz von Personen und Sachen vor strafbaren Handlungen bzw. zur Verfolgung solcher strafbarer Handlungen eingesetzt werden. Der Einsatz zum Zweck der Aufmerksamkeitserkennung in der Schule wäre damit von vornherein rechtswidrig.

auf Entwicklung (Art. 6 Abs. 2 KRK) in Konflikt geraten könnte. Angesichts der weiterhin bestehenden geringen Zuverlässigkeit der Technologie dürften sich alternative Massnahmen, insb. didaktischer, personeller oder baulicher Art, als wesentlich geeigneter zur Erreichung des intendierten Ziels erweisen.

Aufmerksamkeitserkennung in Privatschulen

Die Aufmerksamkeitsanalyse, wie sie hier beschrieben ist, stellt in der Regel eine *Persönlichkeitsverletzung* der betroffenen Schülerinnen und Schüler dar. Es kann auch nicht davon ausgegangen werden, dass diese ihre Daten bereits dadurch schon allgemein zugänglich gemacht haben,[157] dass sie sich im Schulzimmer aufhalten. Als Rechtfertigung für die Persönlichkeitsverletzung kommen die Einwilligung, ein überwiegendes privates oder öffentliches Interesse oder eine gesetzliche Pflicht in Betracht.[158]

Wird die Aufmerksamkeitsanalyse ohne das Wissen der betroffenen Schülerinnen und Schüler und/oder Lehrpersonen durchgeführt, liegt von vornherein keine Einwilligung vor. Die Einwilligung muss von jeder/m einzelnen Schülerin und Schüler erteilt werden, und dies nach angemessener Information und freiwillig.[159] Sie muss zudem ausdrücklich erfolgen.[160] Da eine Einwilligung nur von einer urteilsfähigen Person erteilt werden kann, stellt sich die Frage, unter welchen Bedingungen Schülerinnen und Schüler urteilsfähig sind bzw. wann eine Einwilligung der Erziehungsberechtigten notwendig wäre. Urteilsfähig im Sinne des Zivilgesetzbuches ist, wer nicht wegen seines Kindesalters oder infolge von «geistiger Behinderung, psychischer Störung, Rausch oder ähnlicher Zustände» ausserstande ist, vernunftgemäss zu handeln.[161] Kinder und Jugendliche sind somit urteilsfähig, wenn sie sich einen eigenen Willen bilden und gemäss diesem Willen handeln können. Das Datenschutzgesetz kennt, anders als die DSGVO,[162] keine feste Altersgrenze. Schülerinnen und Schüler der Volksschule sind in der Regel zwischen 6 und 16 Jahre alt. In dieser Zeitspanne variiert die Fähigkeit, einen eigenen Willen zu bilden und diesem Willen entsprechend zu handeln, erheblich, auch unter Gleichaltrigen. Da es auch für Erwachsene schwierig ist, die Konsequenzen einer Datenbearbeitung abzuschätzen, ist dies für Kinder umso schwieriger. Darum kann ihnen bezüglich der Bearbeitung von besonders schützenswerten Personendaten sowie bei einem Profiling höchstens gegen Ende der Schulzeit Urteilsfähigkeit zugesprochen werden. Davor sollte auch eine Einwilligung der Erziehungsberechtigten vorliegen (Erziehungsdirektion des Kantons Bern Amt für Kindergarten, Volksschule und Beratung, S. 13; educa.ch 2009, S. 32–33).

In diesem Kontext ist auf eine Entscheidung der Schwedischen Datenschutzbehörde (DPA) hinzuweisen, welche im Jahr 2019 eine DSGVO-Busse gegen eine Schule verhängte, welche im Rahmen eines Pilotprojekts Gesichtserkennungssoftware zum Zweck der Anwesen-

---

[157] Art. 30 Abs. 3 nDSG.

[158] Art. 31 Abs. 1 nDSG.

[159] Art. 6 Abs. 6 nDSG.

[160] Art. 6 Abs. 7 lit. a und b nDSG.

[161] Art. 16 ZGB.

[162] Art. 8 Abs. 1 DSGVO sieht das sechzehnte Lebensjahr als Altersgrenze vor. Dieser Artikel bezieht sich allerdings auf «Dienste der Informationsgesellschaft» und nicht auf Datenbearbeitungen im Bildungskontext.

heitskontrolle der Schülerinnen und Schüler eingesetzt hatte. Zwar hatten die Eltern der Schüler in die Datenbearbeitung eingewilligt, jedoch konstatierte die DPA ein deutliches Ungleichgewicht zwischen der Schulbehörde und den betroffenen Personen aufgrund des Abhängigkeitsverhältnisses der betroffenen Schüler zur Schule, sodass zwar eine Einwilligung vorliege, diese aber nicht freiwillig erteilt worden sei. Der Einsatz der Gesichtserkennungssoftware zur Anwesenheitskontrolle wurde darüber hinaus auch als unverhältnismässig beurteilt, da genügend andere mildere Mittel zur Verfügung stünden, um die An- bzw. Abwesenheit einzelner Schüler einer Klasse zu kontrollieren, ohne derart stark in deren Rechte und Freiheiten einzugreifen.[163] In diesem Verfahren wurde die Gesichtserkennung zwar zu einem anderen Zweck eingesetzt (Anwesenheitskontrolle) als dem hier untersuchten (Aufmerksamkeitsanalyse), jedoch lassen sich die Überlegungen *mutatis mutandis* auch auf den hier einschlägigen Fall übertragen, zumal die Aufmerksamkeitsanalyse einen noch schwereren Eingriff in die Privatsphäre als die Anwesenheitskontrolle darstellt.

Praktische Probleme dürften sich ergeben, wenn aus einer Schulklasse nicht alle Kinder resp. Eltern eingewilligt haben. In diesen Fällen müsste mit technischen Mitteln sichergestellt werden, dass die Daten der nicht einwilligenden Schülerinnen und Schüler nicht bearbeitet werden.

In gewissen Fällen, z.B. der Erkennung von Mobbingopfern oder der Prävention von Gewalttaten, liesse sich auch erwägen, ob ein überwiegendes Interesse am Einsatz der Aufmerksamkeitserkennungstechnologie vorliegt, jedoch ist angesichts des sehr schwerwiegenden Eingriffs in die Privatsphäre ein solches Interesse nur zurückhaltend anzunehmen. In vielen Konstellationen dürfte es deshalb an einer ausreichenden Rechtfertigung für diese Art der Datenbearbeitung fehlen.

Gesamtbewertung

Wie auch schon bei der Emotionserkennung allgemein, lässt sich für die Aufmerksamkeitserkennung an Schulen festhalten, dass die Anwendung dieser Technologie sowohl durch staatliche wie auch durch private Akteure in sehr vielen, wenn nicht den meisten, Fällen nicht mit den verfassungs- und datenschutzrechtlichen Vorgaben in Einklang zu bringen sein wird, zumal sich im hier untersuchten Fall die Vorgaben dadurch, dass die Betroffenen minderjährig sind, noch erhöhen.

3.9.3. Gesellschaftliche und ethische Herausforderungen

Die demokratischen Grundrechte und Freiheiten sollen nicht nur Erwachsenen, sondern auch Kindern eine offene Zukunft ermöglichen. Viele Staaten weltweit, darunter auch die Schweiz, haben sich dazu verpflichtet, Kinder- und Jugendrechte umzusetzen, wozu auch der Schutz der kindlichen Privatsphäre zählt (BSV 2022). Demnach bedürfen Kinder geschützter Räume, um ihre Persönlichkeit und ihre Selbstbestimmung erproben und erlenen zu können, um so zu selbstbestimmten Erwachsenen zu werden. In diesem Sinne kommt

[163] Beschluss vom 20.8.2019, Medienmitteilung abrufbar unter https://www.imy.se/om-oss/arkiv/nyhetsarkiv/sanktionsavgift-for-ansiktsigenkanning-i-skola/ (zuletzt besucht am 22.7.2021); eine deutsche Zusammenfassung findet sich unter Reuter 2019b.

auch dem Schutz der Privatheit von Kindern eine wesentliche Bedeutung für Demokratie und Rechtsstaatlichkeit zu (Stapf et al. 2021a, S. 352–353). In westlichen Demokratien gilt die Schule als ein Ort, an dem Schülerinnen und Schüler eine eigene Identität entwickeln sollen. Für viele Schülerinnen und Schüler gehört ein «Verstecken» zur Art und Weise, wie sie Schule «durchführen» (Gordon et al. 2000).

Ethische Herausforderungen ergeben sich insb. im Spannungsfeld zwischen elterlicher und staatlicher Schutz- und Fürsorgepflicht einerseits und der kindlichen Selbstbestimmung andererseits (Stapf et al. 2021a, S. 351–352). Denn der Staat und seine Bildungsinstitutionen sollen die kindliche Entwicklung zu Selbstbestimmung fördern, indem Lernziele gesetzt und ihre Erreichung überprüft wird. Dies erfolgt durch die unmittelbare Aufsicht der Lehrkräfte, der nachgelagerten Aufsicht durch Eltern sowie der übergeordneten Kontrolle durch weitere Organe und Massnahmen, etwa Kindes- und Erwachsenenschutzbehörden (KESB) oder Bildungsmonitoring. Nach diesem Verständnis sind Leistungsüberprüfung und Entwicklung kindlicher Selbstbestimmung kein Widerspruch, sondern Letzteres wird durch Ersteres unterstützt (Herzog 2015). Der Einsatz neuer Technologien, wie der Emotionserkennung zum Zwecke der Aufmerksamkeitsanalyse in Schulen, könnte sich in dieses Gefüge einreihen und so die verbesserte Leistungsüberprüfung von Schülern gewährleisten und zugleich Kapazitäten aufseiten der Lehrkräfte frei machen, die sie in eine intensivere Unterrichtsbetreuung investieren könnten.[164]

Demgegenüber steht die Kritik, dass ein Übermass an Überwachung entgegen des anvisierten Ziels der Steigerung der Lernfähigkeit diese stören könne, wie schon in der rechtlichen Diskussion angesprochen (vgl. 3.9.2). Insgesamt lassen sich viele der bereits in den vorangegangenen Kapiteln diskutierten Herausforderungen auf den Einsatz von Emotionserkennung zum Zwecke der Aufmerksamkeitsanalyse in Schulen übertragen, so etwa die Diskussionen zur Zuverlässigkeit und Genauigkeit der Technologie und der Gefahr von Abschreckungseffekten und damit einhergehenden Verhaltensanpassungen, die im Folgenden kurz auf den Kontext der Aufmerksamkeitsanalyse in Schulen übertragen werden.

Im Hinblick auf die Zuverlässigkeit bestehen zunächst v.a. grundsätzliche Bedenken hinsichtlich der Frage, ob und inwiefern die Analyse von Gesichtsausdrücken Rückschlüsse über die individuelle Aufmerksamkeit zulässt (Dewan et al. 2019, S. 16). Diskriminierungen wären aber auch dann zu erwarten, wenn die grundsätzlichen Probleme gelöst wären, aber die genutzten Trainingsdaten und Algorithmen ein Bias aufweisen (Andrejevic und Selwyn 2020, S. 123). Infolge dieser technisch-konzeptionellen Herausforderungen könnten Schüler, deren Verhalten bzw. Gesichtsausdrücke etwa aus kulturellen, gesundheitlichen oder sonstigen individuellen Gründen nicht dem rechnerisch erwarteten Standard entsprechen, Benachteiligungen erfahren. So könnte ein Schüler, der aus dem Fenster schaut, durchaus über eine Problemstellung nachdenken. Fraglich wäre, ob ein Aufmerksamkeitsanalysesystem dies als Ablenkung deuten würde, da der Blick nicht auf die Tafel gerichtet war (Lewis und Diamond 2015).

---

[164] Anders stellt sich dies freilich in China dar, wo v.a. der Gedanke der Leistungssteigerung durch Disziplinierung im Vordergrund stehe (Yujie 2019). In der repräsentativen Bevölkerungsumfrage, die im Rahmen der vorliegende Studie durchgeführt wurde, befürworteten zwar nur 13 % der Befragten den Einsatz der Aufmerksamkeitsanalyse in der Schweiz. Jedoch sahen 69 % dieser Befürworter in ihr eine Möglichkeit zur besseren Disziplinierung der Schülerinnen und Schüler.

Besondere Sprengkraft gegenüber anderen Anwendungsformen der Gesichtserkennung besitzt die Aufmerksamkeitsanalyse, weil ihr Einsatz in stärkerem Masse zu automatisierten Schlüssen verleiten könnte. Anders als bspw. Gesichtserkennung zu Identifikationszwecken, die im Regelfall eine Überprüfung der Softwareergebnisse durch die Kontrolle der Identität der Person erforderlich macht, würde ein Echtzeiteinsatz der Aufmerksamkeitsanalyse einen dauerhaften Informationsfluss und Entscheidungsdruck verursachen. Wenn es bei der Aufmerksamkeitsanalyse etwa darum ginge, dass die Lehrkraft im Unterrichtsalltag in Echtzeit einen Aufmerksamkeitsscore für jede Schülerin und jeden Schüler angezeigt bekäme, könnte die Lehrkraft dazu geneigt sein, diesem Score zu vertrauen, ohne ständig selbst nachzuprüfen, ob die Einstufung korrekt ist. So würde aus einem Unterstützungswerkzeug für Lehrkräfte möglicherweise ein automatisiertes Entscheidungssystem. Langfristig könnte ein Interesse an der langfristigen Speicherung der erfassten Daten und an ihrer Zusammenführung mit zusätzlichen Verhaltensdaten und Datenbanken entstehen, was zu weiteren Herausforderungen führen würde. Auf diese Weise wäre nicht nur das Recht von Kindern auf eine offene Zukunft gefährdet, zusätzlich würden sehr weitgehende Profile über die betroffenen Kinder erstellbar, die zu vielfältigen Nutzungszwecken einladen könnten, etwa der Weitergabe der Daten an andere Stellen und dem Übergang zu einer zunehmend vollständigen automatisierten Leistungsbewertung von Schülerinnen und Schülern, womöglich auf Basis prädiktiver Analytik (Andrejevic und Selwyn 2020, S. 121; Stark 2019, S. 53).

Schülerinnen und Schüler könnten aufgrund des Einsatzes der Aufmerksamkeitsanalyse unter Konformitätsdruck geraten und ihr Verhalten ändern. Bis zu einem gewissen Grad wäre dieses Resultat auch durchaus erwünscht. Schliesslich würde im Falle eines Einsatzes gerade die Steigerung der Aufmerksamkeit beabsichtigt. Im Falle einer technisch nicht weitgehend zuverlässigen Aufmerksamkeitsanalyse könnten jedoch unabsehbare Konsequenzen entstehen. Wenn sich bspw. herausstellen sollte, dass bestimmte Gesichtsausdrücke bessere Leistungen versprechen, könnten Schüler diese schlicht nachahmen, nur um von der Software nicht markiert zu werden. Ob auf diese Weise ein höheres Mass an Aufmerksamkeit für die Lerninhalte erzielt würde, ist fraglich. Eine technisch weitgehend zuverlässige Aufmerksamkeitsanalyse könnte hingegen unintendierte Konsequenzen mit sich bringen. Wenn schon kleinste Unaufmerksamkeiten erfasst würden, könnte sich ein Angstzustand des dauerhaft Überwachtwerdens entfalten. Schüler, die sich bloss aus Angst vor Konsequenzen regelkonform verhalten, könnten in ihrer freien Identitätsentwicklung gestört werden (Gordon et al. 2000). Denkbar ist aber auch, dass sich solche Schüler, die unzureichendes Wissen über die Folgen der stattfindenden Datenbearbeitungen und über Verhaltens- und Vermeidungsmöglichkeiten besitzen (Rossnagel und Richter 2017), stattdessen unbesorgt verhalten und Nachteile gegenüber bewusst agierenden Schülern erleiden. Wie die Debatten zum Digital bzw. Privacy Divide zeigen, könnten hierbei soziodemografische Faktoren ausschlaggebend sein und in struktureller Diskriminierung resultieren (Alhazmi et al. 2022) .

Kritisch eingewendet wird auch, dass Massnahmen unter dem Vorwand der Steigerung der Lernfähigkeit eingeführt würden, etwa das Bildungsmonitoring, obwohl sie vielmehr auf die ökonomische Durchrationalisierung der Bildungsvermittlung abzielten (Herzog 2015). Ausgehend von diesem Argument wäre es denkbar, dass der Einsatz von Aufmerksamkeitsanalysen in Schulen nicht zu mehr freiwerdenden Kapazitäten aufseiten der Lehrkräfte und einer verbesserten Betreuungssituation führen würde, sondern bspw. zu einer Vergrös-

serung der Klassen, weil eine Lehrkraft mittels Technologieunterstützung in der Lage wäre, den Unterricht effizienter zu gestalten.

Je nach Ausgestaltung des entsprechenden Unterrichts könnte bei der Einführung von Aufmerksamkeitserkennung das Ziel der Steigerung des Lernerfolgs auch dadurch beeinträchtigt werden, indem zu stark auf die Technologie fokussiert wird und Kernelemente erfolgreichen Unterrichts, etwa Dialog, Austausch und Möglichkeiten für eigene Fragen und Experimente, in den Hintergrund rücken (Saltman (2016, S. 62–63).

Weitere Kritik widmet sich der Missbrauchsanfälligkeit der Technologie: Durch Gesichtserkennung würden autoritäre Tendenzen in Schulen gestärkt und neue Räume der Ausnutzung autoritärer Macht eröffnet, z.B. durch Lehrkräfte, Schulleiter oder Behörden und Eltern. Auch wenn ein rechtskonformer Einsatz derartige Spielräume vermutlich stark reduzieren würde, bliebe ein Restrisiko missbräuchlicher Nutzung (Hartzog 2018).

3.9.4. Zwischenfazit

Das für die Schweiz hypothetische Szenario der Emotionserkennung zum Zwecke der Aufmerksamkeitsanalyse in Schulen brächte zahlreiche Herausforderungen mit sich. Im Bereich der technischen Funktionsfähigkeit stellen sich dieselben Herausforderungen hins. Trefferraten wie bei der Stimm-, Sprach- und Gesichtserkennung im Allgemeinen. Ausserdem ergeben sich auch hier die bei der Emotionserkennung diskutierten Bedenken hinsichtlich der wissenschaftlichen Validität der korrekten Interpretation von Emotionen mittels v.a. der Analyse von Gesichtsausdrücken.

Auch im Bereich der rechtlichen Einschätzung ist stark zu bezweifeln, dass die Aufmerksamkeitsanalyse an Schulen sich je mit den verfassungs- und datenschutzrechtlichen Vorgaben in Einklang bringen lässt.

Im Bereich der ethischen und gesellschaftlichen Herausforderungen ergeben sich ebenfalls Überlappungen zu vorangegangenen Diskussionen, insb. im Hinblick auf die Gefahr von Abschreckungseffekten und damit einhergehenden Verhaltensanpassungen. Derartige mögliche Auswirkungen entfalten jedoch bei Kindern und Jugendlichen eine besondere Brisanz, da sie deren Entwicklung zu selbstbestimmten Individuen beeinträchtigen können.

3.10. Jedermann-Identifikation

Von den in der vorliegenden Studie untersuchten Anwendungsfällen, stellt die Jedermann-Identifikation das nach heutigem Stand fiktivste Anwendungsszenario dar. Die grundsätzliche Idee hinter diesem Anwendungsfall verweist auf die Möglichkeit der Nutzung von Gesichtserkennungstechnologien durch jeden beliebigen Menschen, ohne dass diese Personen beim Einsatz der Technologie bemerkt werden. Zunehmend kleiner werdende Hardware im Bereich sowohl der Kamera- als auch der Prozessortechnik und hocheffiziente Gesichtserkennungsalgorithmen bergen das Potenzial für die Identifikation von Menschen in der Öffentlichkeit, sozusagen «im Vorbeigehen». Im Rahmen dieses Anwendungsfalles

wollen wir nun die Personenidentifikation mittels Einsatzes von Gesichtserkennungssoftware auf einer Datenbrille näher untersuchen.

Datenbrillen sind v.a. in zwei Nutzungskategorien einzuordnen: «Virtual Reality», bei der ein Nutzender *durch die Ausblendung der realen Welt in ein computererzeugtes Abbild einer realen oder künstlichen Welt eintaucht* einerseits, und «Augmented Reality», bei der *in Echtzeit Objekte oder Inhalte in die reale Umgebung eingeblendet bzw. überblendet werden, um die reale Welt für die Technologienutzenden gezielt anzureichern* (Vogel et al. 2020, S. 45). Für den vorliegenden Untersuchungsfall interessiert v.a. der Einsatz einer Datenbrille im Rahmen von Augmented Reality, bei der es möglich ist, Zusatzinformationen zur Umwelt einzublenden, etwa den Namen oder das Onlineprofil einer Person.

Am Thema Augmented Reality wird bereits seit Ende der 1960er-Jahre wissenschaftlich geforscht (Vogel et al. 2020, S. 20), grössere öffentliche Aufmerksamkeit zog das Thema allerdings eher als Beiwerk zunächst im Laufe der 1990er-Jahre auf sich, als insb. das Thema Virtual Reality Eingang in die mediale Populärkultur fand (Jantschewski 2019). Nachdem Hard- und Software den gesellschaftlichen Erwartungen nicht ansatzweise standhalten konnten und das Thema zunächst von der Bildfläche verschwand, gab erst die Mitteilung Googles, an einer Datenbrille zu arbeiten, dem Thema neuen Auftrieb (Velazco 2012). Währenddessen arbeiteten andere Hersteller ebenfalls an der Entwicklung moderner Datenbrillen, so etwa Canon (Bergen 2013) oder Brilliantservice. Der Prototyp der Datenbrille des Herstellers Brilliantservice konnte zudem bereits zur Vorstellung auf dem Mobile World Congress Anfang 2013 die eingebaute rudimentäre Gesichtserkennungstechnologie demonstrieren (Piltch 2013).

Schon kurz nachdem Google im Februar 2013 erste Modelle der Datenbrille Google Glass an ausgewählte Tester und Google I/O-Softwareentwickler ausgeliefert hatte, begann eine intensive Debatte über die ethischen Folgen der Nutzung. Zunächst stand dabei die Videoaufnahmefunktionalität der Brille im Vordergrund, die es ermöglichen würde, jederzeit versteckte Audio-, Bild- und Videoaufnahmen der Umgebung zu erstellen (Arthur 2013b), sodass eine Debatte rund um die Aushandlung einer neuen sozialen Etiquette bei Nutzung von Datenbrillen entbrannte (Warman 2013). Bald wurden zudem IT-Sicherheitsmängel bekannt, die eine vollständige Kaperung und damit Fernsteuerung des Geräts seitens des Angreifers ermöglichten – jener White Hat-Hacker, der die Schwachstelle öffentlich gemacht hatte, wies zudem darauf hin, dass er nur zehn Minuten für den Angriff benötigt hatte (Arthur 2013a). Schliesslich stellten verschiedene Entwickler, obwohl Google bewusst keine Gesichtserkennungsfunktion implementiert hatte, Drittanbieter-Apps zur Gesichtserkennung vor (NeatoCode Techniques 2013). Obwohl das Google-Glass-Entwicklerteam wohl lange vor diesen Vorfällen über die Datenschutzprobleme diskutiert und nach Lösungen gesucht hatte, blieb der Konzern zunächst untätig (Bilton 2012). Erst die im Zuge der Debatte rund um die Implementierung des Gesichtserkennungsfeatures und um die heimlichen Aufnahmemöglichkeiten entbrannte Kritik veranlasste Google Anfang Juni 2013 schliesslich dazu, die Gesichtserkennung bei Google Glass explizit zu verbieten und fortan Apps, welche die entsprechende Funktionalität beinhalten, auf dem App-Markt nicht freizugeben. Darüber hinaus wurde auch das Abschalten des Displays der Brille während einer Aufnahme verboten, sodass Gefilmte in unmittelbarer Nähe bei ausreichend guten Lichtbedingungen stets erkennen können sollten, dass sie gefilmt werden (Donath 2013).

Trotzdem forschten Entwicklerinnen und Entwickler auch weiterhin an Gesichtserkennungsanwendungen (Fraunhofer IIS 2014), während Hacker unterdessen die weitreichenden Eingriffsmöglichkeiten in die Datenbrille demonstrierten (Christiaan008 2014). Beispielsweise erschufen die Entwickler der App NameTag eine Datenbank mit zwei Millionen Bildern und kündigten an, in Zukunft standardmässig – also ohne Einwilligung der Betroffenen, diese sollten lediglich die Möglichkeit des Opt-Out erhalten – auch auf Daten aus sozialen Netzwerken wie Facebook oder aus Onlinepartnerbörsen zuzugreifen. Derweil hofften die Entwickler angesichts des von Google ausgesprochenen Gesichtserkennungsverbots, dass entweder Google einlenken und die Funktionalität wieder erlauben werde oder früher oder später andere Datenbrillenentwickler ihre Gesichtserkennungstechnologie einsetzen bzw. erlauben würden (Tremmel 2014).

Die anhaltende Kritik an Googles Datenbrille kulminierte indes zunächst weltweit in der Aufstellung von Verbotsschildern in Restaurants, Kinos oder Krankenhäusern, die das Tragen von Datenbrillen untersagten (Beuth 2013). Für Personen, die die soziale Etiquette (also das Abnehmen der Brille in sozialen Situationen) nicht einzuhalten bereit waren, entwickelte sich gar das Schimpfwort «Glasshole» (Lobo 2013). Schliesslich verschob Google den für 2014 vorgesehenen Verkaufsstart hinaus und weitete lediglich die Teilnehmerzahl für die Open-Beta aus. Doch dabei zeigte sich, dass die von Google ersehnte Kundschaft ausblieb (Wohlsen 2014). Ende des Jahres wurde ausserdem bekannt, dass sich zwischenzeitlich viele App-Entwickler die Entwicklung von Glass-Apps eingestellt hatten, weil sich bei ihnen die Ansicht durchgesetzt hätte, dass die Brille eher für einen spezialisierten Einsatz und weniger für den Massenmarkt geeignet wäre (Donath 2014). Daraufhin beschloss Google Anfang 2015, den Vertrieb der Datenbrille an Privatnutzer vollständig einzustellen, und kündigte an, sich fortan stärker auf spezialisierte B2B-Bereiche zu fokussieren (dpa 2015). Eine unter Schweizer Internetnutzenden repräsentative Umfrage aus dem Jahr 2016 verdeutlichte das fehlende Interesse aufseiten von Privatpersonen: Während 29 % der Befragten ihr konkretes Desinteresse und 19 % Interesse am Kauf einer Datenbrille äusserten, wussten 53 % nicht, worum es sich bei Datenbrillen handelt. Auf die Rückfrage hins. der Gründe für das Desinteresse gaben einige Befragte Datenschutz- oder sonstige Bedenken (zusammen rund 11 %) an, deutlich mehr Personen (zusammen rund 35 %) verwiesen aber auf den mangelnden persönlichen Nutzen, den sie in der Nutzung einer Datenbrille erkannten (IAB Switzerland 2016).

Seither ist es in der öffentlichen Debatte rund um Datenbrillen zwar stiller geworden, doch im Rahmen von Forschungsprojekten wurden produktive Nutzungen vorangetrieben. Mittlerweile werden Datenbrillen in unterschiedlichen Bereichen eingesetzt, etwa im Medizinbereich, im Bauwesen, im Bereich der Produktion, Logistik und Werbung sowie seitens Polizisten (Vogel et al. 2020, S. 24; Thomas et al. 2020, S. 14). Zwar gibt es inzwischen durchaus eine Reihe von Datenbrillen für den Konsumentenmarkt. Diese sind jedoch meist noch sehr deutlich als Datenbrille erkennbar.[165] Eine Ausnahme bildet die Datenbrille «Focals» des Herstellers North aus dem Jahr 2019. Diese verfügt zwar nicht über die in anderen Datenbrillen ansonsten üblicherweise integrierte Kamera, lässt sich optisch aber kaum von herkömmlichen Brillen unterscheiden, während die anderen gängigen AR-Funktionalitäten

[165] Siehe z.B. die Auflistung in: Rupareliya 2021.

enthalten sind. Dennoch war auch diese Datenbrille kein Erfolg und das Start-up wurde letztlich von Google übernommen (Herbig 2020).

Fortschritte in der Technikminiaturisierung lassen aber den Schluss zu, dass (auch mit Kameras ausgestattete) Datenbrillen sich künftig nicht oder kaum mehr von herkömmlichen Brillen unterscheiden lassen werden. Zuletzt zog im Jahr 2021 eine Kooperation zwischen Ray Ban und Facebook Aufmerksamkeit auf sich, bei der eine Augmented-Realitiy-Sonnenbrille mit integrierter und schwer erkennbarer Kamera entwickelt wurde, die kaum von herkömmlichen Sonnenbrillen zu unterscheiden ist. Die in der Schweiz zurzeit noch nicht erhältliche Brille enthält keine Gesichtserkennung; über ein entsprechendes Nachfolgemodell mit Gesichtserkennung wird aber bereits spekuliert. Während der Filmaufnahmen schaltet sich an der Brille ein kleines Licht ein. Gemäss einem Selbstversuch einer Reporterin in New York hätten die Passanten jedoch kaum bemerkt, dass sie gefilmt wurden (Fulterer 2021). Samsung, Google und Sony haben zudem bereits Patente auf Kontaktlinsen mit Mikrokameras angemeldet (Peters 2018). Daneben wird seit einigen Jahren an Augmented-Reality-Kontaktlinsen geforscht (Kreppmeier 2020).

Während der Durchbruch im Bereich von Datenbrillen bislang ausgeblieben ist, haben sich Gesichtserkennungstechnologien zwischenzeitlich stark weiterentwickelt. Wie einfach es ist, selbst eine Überwachungsmaschine zu bauen, demonstrierte das Schweizer Fernsehen Anfang 2020: Die Journalisten konnten in kurzer Zeit Kandidierende der eidgenössischen Wahlen auf Fotos identifizieren, die von anderen Nutzerinnen auf Instagram gepostet worden waren (SRF 2020). Ein wichtiger Ausgangspunkt für das im Rahmen dieses Anwendungsfalles diskutierte Szenario der Jedermann-Identifikation ist allerdings die russische App FindFace.[166] Entwickelt wurde die App vom erst im Jahre 2015 gegründeten russischen Unternehmen NTechLab. FindFace erlaubte es, Fotos der Gesichter beliebiger Menschen, etwa ein ungesehen in der Öffentlichkeit geschossenes Porträtfoto, mit der Profilbilddatenbank der russischen Social-Media-Plattform vKontakte abzugleichen und sie so zu identifizieren. Die App war Anfang 2016 in Russland populär geworden, woraufhin auch westliche Medien über das Thema berichteten. Problematisiert wurde insb. das mit der App sich bietende Potenzial der unbemerkten Identifikation fremder Menschen und damit auch der Aufhebung der individuellen Anonymität im öffentlichen Raum. Alexander Kabakov, Mitgründer und Mitentwickler der App, zeigte sich indes von der Kritik unbeeindruckt und vertrat die Ansicht, dass FindFace Dating revolutionieren könne: «If you see someone you like, you can photograph them, find their identity, and then send them a friend request.» (Walker 2016). Bald darauf wurde bekannt, dass die App zu Doxing-Zwecken gegen russische Pornodarstellerinnen verwendet wurde. Nachdem die Frauen auf vKontakte identifiziert wurden, veröffentlichten die Täter das Material und spammten die Familien und Kontakte der Personen auf vKontakte mit ihren Funden (Rothrock 2016). Zudem erfolgte die in Abschnitt 3.4 erwähnte Identifikation von oppositionellen Demonstranten mittels FindFace (Hans 2017).

Nachdem vergleichbare Software in westlichen Ländern zunächst keine Anwendung fand, demonstrierte insb. der Skandal rund um Clearview AI, dass auch in Demokratien unbemerkt Gesichter mit den Datenbanken von Facebook, Instagram und Youtube abgeglichen werden

[166] Da die Google-Bildersuche einfache Mustererkennung (Farben, Kleidung, Hintergründe usw.) und keine biometrische Identifikation durchführt, bleibt diese in der folgenden Analyse unbeachtet (Google Inc. 2022b).

(Beuth 2021). Im Juni 2020 wurde schliesslich bekannt, dass die polnische Firma PimEyes eine ähnliche Gesichtserkennungsfunktionalität bietet, wie zuvor FindFace (Gershgorn 2020). So war es möglich, auf der Webseite von PimEyes beliebige Fotos hochzuladen und diese mit 900.000 Millionen (Stand: April 2020) Fotos von Gesichtern abzugleichen. Als Ergebnis wurde zwar nicht unmittelbar der Name der gesuchten Person ausgegeben, doch die Anzeige des Fundorts der passenden Fotos ermöglichte in vielen Fällen Rückschlüsse über Namen, Beruf, Wohnort usw. Die Gründer von PimEyes gaben an, ihre Fotodatenbank auf Basis von Web-Scraping öffentlich zugänglichen Bildmaterials, das Scraping explizit erlaubt, aufgebaut zu haben, etwa Tumblr oder Nachrichtenseiten und Blogs. Tests demonstrierten jedoch, dass auch Bildmaterial aus Instagram, YouTube, TikTok, Twitter und vKontakte in den Suchergebnissen enthalten war (Laufer und Meineck 2020). PimEyes teilte diesbezüglich mit, dass sie sich selbst nicht in der Verantwortung sehen, wenn geschütztes Bildmaterial aufseiten Dritter ungeschützt zum Scraping freigegeben werde. Eine Einwilligung der Betroffenen wurde hingegen gar nicht eingeholt (Wakefield 2020).

Als problematisch an dem Dienst wurde auch die Ermutigung zum Hochladen der Fotos fremder Menschen und zum Stalking bewertet. Die Webseite bewarb ihren Dienst bspw. mit der Suche nach Fotos prominenter Menschen wie Ophra Winfrey oder Johnny Depp. Zudem betrieben die Gründer von PimEyes einen weiteren Dienst namens *Catfished*, der förmlich zum Stalking aufrief: «Verwenden Sie die Gesichtserkennung, um mehr über Ihren Schwarm herauszufinden!» Infolge massiver Kritik wurde die Firmenpolitik schliesslich wesentlich abgeändert: Inzwischen präsentiert sich PimEyes als Dienst zum Schutz der eigenen Privatheit. An mehreren Stellen wird darauf hingewiesen, dass der Dienst zur Suche nach Fotos von sich selbst verwendet werden soll, um insb. die missbräuchliche Nutzung eigener Fotografien (etwa im Rahmen von Fake-Profilen oder Revenge Porn) im Netz zu entdecken. Zudem wurden die Verweise auf Rabatte bei Massenabfragen, zum Hochladen der Fotos fremder Personen und zur Eingrenzung der Suche auf «Adult Sites» entfernt sowie die Webseite *Catfished* vom Netz genommen (Laufer und Meineck 2020). Nichtsdestotrotz ist die Webseite weiterhin online und es lassen sich auch weiterhin beliebige Fotos und damit auch Fotos fremder Menschen hochladen und analysieren.

Schon heute ist es also möglich, die Gesichter fremder Menschen ungesehen mit Smartphones und anderen kamerafähigen Geräten abzufotografieren und mittels Gesichtserkennungstechnologie nach der Identität dieser Personen zu suchen. Einschränkungen existieren zwar noch, denn die mit Kameras ausgestatteten Datenbrillen sind noch sehr gut als solche erkennbar und die Gesichtserkennungsdienste verfügen noch nicht über die Gesichter aller Menschen. Doch der Punkt, an dem die Technikminiaturisierung einen Grad erreicht haben wird, bei dem die Fotografen ungesehen Fotos beliebiger Personen schiessen und diese identifizieren können werden, ist nicht mehr fern.

3.10.1. Technische Grundlagen und Möglichkeiten

Da derzeit keine smarten Brillen auf dem Markt verfügbar sind, beschäftigt sich dieses technische Kapitel mit der Funktionalität der Plattform PimEyes.

Auf dieser Plattform können Endnutzer ohne Anmeldung entweder mittels Kamera ein Bild von sich selbst machen und hochladen oder ein beliebiges Bild von deren Festplatte wählen. Dabei scheint die Aufmachung der Webseite so, als solle Nutzenden damit mehr Transparenz im Hinblick darauf geschaffen werden, auf welchen Webseiten ein Bild von ihnen auftaucht. Zudem bewirbt das Unternehmen die Suchfunktion als ein Werkzeug zur Durchsetzung des Urheberrechts (PimEyes 2021). Laut Webseite kann ein Gesicht auch auf Fotos in unterschiedlicher Grösse und Auflösung, mit anderen Hintergründen, in Gruppenbildern und auch bei einem anderen Haarschnitt erkannt werden (PimEyes 2021). Laut eigener Aussage besteht die Datenbank von PimEyes im Jahr 2020 aus 900 Millionen Gesichtern, wobei 1 Terrabyte an Daten jeden Tag analysiert wird und insgesamt 150 Millionen Webseiten durchsucht werden (Internet Archive 2020). Immerhin scheinen die Anbieter eine Möglichkeit zur Löschung von Bildern des eigenen Gesichts aus deren Datenbank anzubieten.[167] Wie effektiv dies ist, bleibt jedoch unklar, da es keine unabhängigen Untersuchungen gibt.

Zudem bietet PimEyes ein Premium-Abo. Dieses ermöglicht es u.a., E-Mail-Benachrichtigungen zu erhalten, sobald ein Gesicht auf einer neuen Webseite auftaucht, eine Vielzahl von Bildern automatisiert hochzuladen (z.B. 100 Millionen pro Monat) und mittels einer Programmierschnittstelle (API) die Funktionalität von PrimEye in eigene Anwendungen einzubauen.

Diese Funktionalität überrascht nicht, denn es gibt bereits Strafverfolgungsbehörden, die den Dienst in eigene Anwendungen integrieren. Z.B. ist sie seit 2018 in die Software «Paliscope» eines schwedischen Unternehmens integriert und hilft Polizeien, Daten aus verschiedenen Quellen im Internet zusammenzufinden (Laufer und Meineck 2020; Paliscope 2021). Mit einer neugegründeten Firma «Faceware AI» wollen die Macher hinter PimEyes zusätzlich spezifische Dienste für Strafverfolger anbieten.

In der Literatur finden sich indes keinerlei Informationen über den verwendeten Algorithmus, Erkennungsraten oder Falsch-Positive-Raten. Im Blog von PimEyes findet sich zwar eine Seite darüber, wie die Bildersuche funktioniert, jedoch sind hier nur allgemeine und oberflächliche Informationen zu finden und keine genauen Details zum Einsatz in der PimEyes-Anwendung.[168] Um dennoch etwas mehr über die Plattform zu erfahren, haben wir einen kurzen eigenen Test mit einem Bild des schweizerischen Präsidenten Guy Parmelin im Juni 2021 gemacht. Das Ergebnis bestätigt die Funktion und die Vielzahl an gefundenen Bildern innerhalb weniger Sekunden. Unser hochgeladenes Foto zeigt sich in den ersten vier Reihen als 1:1-Fund. Danach folgten Fotos mit unterschiedlichem Hintergrund, anderen Posen und Gesichtsausdrücken und ein Link zur jeweiligen Fundseite, der gegen Gebühr sichtbar wird.

Es darf jedoch nicht vergessen werden, dass auch die Rückwärtssuche von Google-Bilder solche Ergebnisse liefern kann, wenngleich die Erkennungsleistung (bewusst) schlechter ist und auch auf Objekte und nicht nur Gesichter abzielt (Bitirim et al. 2020).

---

[167] https://pimeyes.com/en/faq/remove-from-database.

[168] https://pimeyes.com/en/blog/how-does-image-search-work.

3.10.2. Juristische Bewertung

Unter dem Thema «Jedermann-Identifikation» werden verschiedene Anwendungsfälle diskutiert, die rechtlich jeweils unterschiedlich zu bewerten sind. Im Folgenden geht es nicht um die Analyse der Rechtskonformität einer konkreten Anwendung, sondern um eine generelle rechtliche Einschätzung der Entwicklung, wonach Identifikation mittels Gesichtserkennungstechnologien für jedermann zugänglich wird, sei es mittels einer «Datenbrille», einer App oder Ähnlichem. Im Folgenden werden deshalb bloss die generellen rechtlichen Überlegungen, die diesen Anwendungen gemein sind, dargestellt.

Bei der Jedermann-Identifikation werden per definitionem biometrische Daten zur eindeutigen Identifizierung natürlicher Personen verwendet.[169] Es werden somit *besonders schützenswerte Personendaten* i.S.v. Art. 5 lit. c Ziff. 4 nDSG bearbeitet.

Verantwortlich für die Datenbearbeitung sind einerseits die Nutzenden, denn sie entscheiden über das Wann und Wo der Datenbearbeitung und deren Zweck (Identifizierung) und deren konkrete Mittel (aufgenommene Gesichtsbilder). Gemäss Art. 5 lit. j nDSG kann – genau wie gemäss der DSGVO[170] – die Verantwortlichkeit für die Datenbearbeitung auch geteilt sein, was hier der Fall ist, denn zusätzlich zu den Nutzenden entscheidet der Betreiber oder Hersteller der Software oder der Applikation über die Gesamtfunktionalität einschliesslich der Daten und der Häufigkeit ihrer Erfassung sowie ihrer Weitergabe an andere (Art. 29 WP 2014, S. 11). Diese geteilte Verantwortlichkeit wirft im Einzelfall komplexe Abgrenzungs- und Zuständigkeitsfragen auf, etwa die Frage, wer für die Datensicherheit oder die Wahrung der Betroffenenrechte zuständig ist oder wer in welchem Umfang für Datenschutzverstösse haftet.[171]

Nicht immer befinden sich die Verantwortlichen für die Datenbearbeitung in der Schweiz oder in der Europäischen Union, die meisten Betreiber oder Hersteller entsprechender Technologien sind im Ausland niedergelassen. Das nDSG sieht, ähnlich wie die DSGVO,[172] vor, dass sich der *räumliche Geltungsbereich* des Gesetzes auf alle Sachverhalte erstreckt, die sich in der Schweiz auswirken, selbst wenn sie im Ausland veranlasst werden.[173] Somit ist das nDSG auch auf Jedermann-Identifikationen vollumfänglich anwendbar, wenn diese durch Personen in der Schweiz durchgeführt werden oder wenn sich die Betroffenen in der Schweiz befinden.

Im hier untersuchten Fall handelt es sich um Datenbearbeitungen durch Private. Die oben beschriebenen Technologien wie z.B. Datenbrillen können auch durch staatliche Akteure genutzt werden, etwa durch Polizeikräfte[174]; dieser Fall wird vorliegend aber nicht untersucht. Private benötigen für persönlichkeitsverletzende Datenbearbeitungen eine *Recht-*

---

[169] Siehe auch (im Kontext der DSGVO) Schindler und Hornung (2021, 515–517, 520). Mit eindeutig ist nämlich nicht «eineindeutig» gemeint; eine gewisse Fehlerrate schliesst das Vorliegen von Biometrie nicht aus.

[170] Art. 4 Ziff. 8 und Art. 26 DSGVO.

[171] Siehe dazu auch EuGH, Rs. C-210/16 (*Wirtschaftsakademie*), 5.6.2018, ECLI:EU:C:2018:388.

[172] Art. 3 Abs. 2 DSGVO.

[173] Art. 3 Abs. 1 nDSG.

[174] Siehe die Beispiele für polizeiliche Nutzung von Datenbrillen in EDPS 2019, S. 6.

fertigung, diese kann durch Einwilligung, ein überwiegendes privates oder öffentliches Interesse oder durch gesetzliche Vorschrift vorliegen.[175] Die Jedermann-Identifikation, wie hier beschrieben, dürfte grundsätzlich eine Persönlichkeitsverletzung gem. Art. 28 ZGB darstellen, da diese auch Verletzungen des Rechts auf Anonymität umfasst.[176] Da besonders schützenswerte Personendaten bearbeitet werden, muss die Einwilligung der betroffenen Personen ausdrücklich erfolgen.[177] Eine solche ausdrückliche Einwilligung wird aber kaum je vorliegen und von den Nutzenden der Jedermann-Identifikation ja gerade nicht gesucht, geht es doch in der Regel um eine heimliche Identifizierung der betroffenen Person. Wer ins Blickfeld einer Datenbrille gerät oder von einer Identifizierungs-App identifiziert wird, hat normalerweise keine Möglichkeit, eine Einwilligung zu erteilen und ordnungsgemäss über die Datenbearbeitung informiert zu werden. Nach Art. 30 Abs. 3 nDSG liegt in der Regel keine Persönlichkeitsverletzung vor, wenn die betroffenen Personen ihre Daten allgemein zugänglich gemacht und eine Bearbeitung nicht ausdrücklich untersagt haben. Diese Regel kann aber keine Anwendung finden, wenn die Gesichtsdaten der Betroffenen unter Missachtung des Transparenzgrundsatzes mit ungefragt beschafften Daten (z.B. aus einem sozialen Netzwerk oder einer Internetplattform) abgeglichen werden sollen (EDÖB 2020b). Es ist bei der Jedermann-Identifikation auch kein überwiegendes privates oder öffentliches Interesse ersichtlich, das diesen sehr schweren Eingriff in die Privatsphäre der Betroffenen rechtfertigen könnte.

Die *Datenschutzgrundsätze* müssen vollumfänglich beachtet werden. Im Kontext der Jedermann-Identifikation stellen sich insb. Fragen zum Grundsatz von Treu und Glauben sowie zum daraus abgeleiteten Transparenzgebot.[178] Es liegt auf der Hand, dass das Aufbauen einer biometrischen Datenbank (bspw. mittels Web-Scraping in sozialen Medien und Internetplattformen) zwecks Abgleichens mit von Nutzern hochgeladenen Gesichtsbildern in schwerer Weise gegen den *Transparenzgrundsatz* verstösst. Diese Bearbeitung geht wesentlich über das bei Suchmaschinen Übliche hinaus und die von der ungefragten Verwendung ihrer Gesichtsbilder Betroffenen können eine derartige Zweckentfremdung nicht voraussehen (EDÖB 2020b). In diesen Fällen liegt somit auch ein Verstoss gegen den Grundsatz der *Zweckbindung* vor. Werden die Daten von Dritten gegen den Willen von sozialen Plattformen beschafft, stellt dies grundsätzlich einen unkontrollierten Abfluss von Personendaten (Data Breach) und damit eine Verletzung der *Datensicherheit* dar, für den nebst den Datenbearbeitern (also den Softwarebetreibern und ggf. auch den Nutzenden der Jedermann-Identifikation) auch die Betreiber der sozialen Netzwerke gegenüber ihren Nutzern einzustehen haben (EDÖB 2020b). Auch die *Verhältnismässigkeit* dieser Datenbearbeitung ist stark infrage gestellt. Für die korrekt oder fälschlicherweise identifizierten Personen können die Datenbearbeitungen ferner erhebliche Folgen nach sich ziehen, deren Vermeidung regelmässig höher zu gewichten sein wird als das Interesse an der Nutzung solcher App (Bühlmann und Schüepp 2020). Eine *Datenminimierung* bspw. mittels Anonymisierung der

---

[175] Art. 31 Abs. 1 nDSG.

[176] Eine Persönlichkeitsverletzung kann nämlich auch durch die Aufdeckung der Anonymität erfolgen (vgl. Urteil des Kantonsgerichts St. Gallen v. 23.6.1983, publiziert in SJZ 1985, 161 ff.; Striptease-Tänzerin).

[177] Art. 6 Abs. 7 lit. a nDSG.

[178] Art. 6 Abs. 2, 3 nDSG.

Daten liegt – anders als bei Applikationen wie Google Street View – gerade nicht im Sinne der Jedermann-Identifikation und ist somit hier keine Option.

Abschliessend zu dieser kurzen rechtlichen Einschätzung kann festgehalten werden, dass die ungefragte Beschaffung von Gesichtsdaten zwecks Weiterbearbeitung unter Einsatz automatischer Gesichtserkennungstechnologien nur schwerlich im Einklang mit den datenschutzrechtlichen Grundsätzen erfolgen kann. Da die für private Datenbearbeitungen notwendige Einwilligung gerade nicht im Sinne der Jedermann-Identifikation ist und andere Rechtfertigungsgründe nicht ersichtlich sind, ist eine solche Verwendung der Gesichtserkennungstechnologie in der Regel rechtswidrig.

3.10.3. Gesellschaftliche und ethische Herausforderungen

Die Möglichkeit, Menschen in beliebigen Situationen zu identifizieren und ggf. ohne ihr Wissen weitreichende Informationen über ihr Privatleben abzurufen, birgt eine Reihe von sowohl eher evidenten als auch hypothetischen Problemen bzw. Herausforderungen mit sich. Zwingend hypothetisch sind einige Herausforderungen v.a. deshalb, weil die im Rahmen dieses Kapitels beschriebene Jedermann-Identifikation ein fiktives Szenario darstellt.

Zunächst soll allerdings noch auf mögliche erwünschte Nutzungsweisen von Datenbrillen mit Gesichtserkennungsfunktion eingegangen werden. Hier ist insb. an die Möglichkeit der bewussten Nutzung nach Einwilligung aller Betroffenen zu denken. Dies könnten z.B. Veranstaltungen sein, bei denen das Kennenlernen neuer Menschen unter Abrufung von freiwillig geteilten Hintergrundinformationen im Zentrum steht: ein Geschäftstreffen, bei dem Namen und Zusatzinformationen zu den jeweiligen Partnern eingeblendet werden; Konferenzen, auf denen die Teilnehmenden Namen, institutionelle Zugehörigkeit, fachliche Interessen und Schwerpunkte usw. teilen; bis hin zu Dating-Partys, auf denen Singles ihre Interessen und Hobbys teilen, um mögliche Partner kennenzulernen. Sofern die Nutzung von Datenbrillen in derartigen Kontexten transparent wäre und Betroffene sich der Nutzung bzw. Datenpreisgabe auch verweigern[179] **könnten, sind hier zahlrei**che Nutzungspotenziale denkbar. Datenbrillen mit integrierter Videofunktion könnten aber auch illegitimes staatliches Handeln aufdecken helfen, indem etwa Polizeigewalt gefilmt und direkt ins Internet gestreamt wird (Vincent 2019). Andere wünschenswerte Nutzungsweisen von Datenbrillen haben weniger mit der Video- bzw. Gesichtserkennungsfunktion als vielmehr mit den sonstigen Eigenschaften derartiger Geräte zu tun: Einblendung von Zusatzinformationen mittels AR, etwa zur Steigerung der Arbeitseffektivität oder Unterst**ützung bei alltäglichen Dingen, für Unterhaltungszwecke wie Videospiele**, aber auch das Betrachten von Videos und Fotos. Insgesamt lässt sich feststellen, dass Datenbrillen v.a. unabhängig von einer möglichen Gesichtserkennungsfunktion eine Reihe von Nutzungszwecken aufweisen.

[179] Wobei es darauf ankäme, ob ein solches Verweigern ein technisch sichergestelltes Nicht-erkannt-werden bedeuten oder zu sozialer Ausgrenzung und damit zu neuen Problemen führen würde, weil Verweigerer dem entsprechenden Kontext fernbleiben müssten.

3.10.3.1. Vollständige Erosion der Anonymität in der Öffentlichkeit

Die zentrale Gefahr besteht darin, dass Anonymität in der Öffentlichkeit durch die Jedermann-Identifikation sukzessive abhandenkommen könnte. Aufgrund der Allgegenwart von Videoüberwachung und Smartphones mit Kameras ist das unerkannte Bewegen durch den öffentlichen Raum bereits heute weitgehend eingeschränkt. Wie das Beispiel der vom SRF selbst gebauten Überwachungsmaschine verdeutlichte, können schon heute im Internet verfügbare Bilder nach Gesichtern durchsucht werden (Grossenbacher und Michel 2020). Wer war wann auf welcher Veranstaltung oder Demonstration? Wer hat sich mal mit wem unterhalten? Selbst ohne Gesichtserkennung in Echtzeit würde dieses Problem verstärkt, weil die Verbreitung von Datenbrillen mit Aufnahmefunktion (Warman 2013) sehr wahrscheinlich zu einer weiteren Zunahme der Aufzeichnung und Übertragung des öffentlichen und (semi-) privaten Lebens führen würde. Dadurch stünde auch mehr Bildmaterial zur Verfügung, das nachträglich zur Gesichtserkennung verwendet werden könnte. Im nächsten Schritt wäre es dann denkbar, dass illegale Datenhändler Profile anlegen, in denen die Interessen, besuchten Orte usw. von Menschen gebündelt und zum Kauf angeboten werden.

Mittels Datenbrillen mit integrierter Gesichtserkennungsfunktion könnte die Deanonymisierung hingegen in Echtzeit «im Vorbeigehen» erfolgen. Menschen würden sich dann, noch mehr als heute schon, Gedanken über ihr Verhalten in der Öffentlichkeit machen, darüber, wo sie sich aufhalten, was sie dort wem sagen. Eine Folge wären Abschreckungseffekte hins. der Wahrnehmung von Bürgerrechten. Demonstrierende müssten etwa befürchten, dass Gegendemonstranten sie deanonymisieren und bei ihren Arbeitgebern melden. Je nach politischer Gesinnung und der Radikalität einer Demonstration könnte eine solche Form der Deanonymisierung zwar legitim erscheinen (man denke an Personen, die Gewalttaten verüben, die dann dem Arbeitgeber gemeldet werden). Es sollte jedoch bedacht werden, dass auf diese Weise verschiedene, zuvor getrennte Lebensbereiche noch mehr als heute schon miteinander verquickt würden und die Kontrolle über das eigene Selbstbild, das der Öffentlichkeit zur Schau getragen wird, noch mehr entgleiten würde. Auch hier liesse sich einwenden, dass die Jedermann-Identifikation ein Hinter-die-Maske-blicken erlauben könnte, die je nach Situation sicherlich auch wünschenswert wäre. Welche genauen Folgen die Erosion der öffentlichen Anonymität hätte, ist zwar schwer abzuschätzen. Klar ist jedenfalls: Was in der Öffentlichkeit noch als privat gelten darf, würde zunehmend von Dritten bzw. der Gemeinschaft bestimmt und nicht mehr von den Individuen selbst, wie es in westlichen Demokratien über viele Jahrzehnte üblich war (Rössler 2001). Tradierte soziale Praktiken würden erodieren und was an deren Stelle tritt, müsste neu ausgehandelt werden.

3.10.3.2. Belästigungs- und Stalking-Problem

Die Erosion der Anonymität kann insb. zu einer Zunahme und veränderten Qualität von Stalking führen. Wenn die Gesichtserkennung auf Datenbrillen mit personenbezogenen Daten aus sozialen Netzwerken und anderen Teilen des Internets verknüpft würde, könnten mittels Jedermann-Identifikation auch Wohnort, häufig aufgesuchte Lokale, Hobbys usw. herausgefunden werden. Ein Stalker oder andere Personen mit Belästigungsabsicht könnten so aus einer Masse an Personen potenzielle Opfer unter Hinzuziehung derartiger Zusatzinformationen auswählen und hätten mehr Möglichkeiten des Zugangs (Walker 2016).

3.10.3.3. Staatliche und private Dauerüberwachung

Schon heute können Ermittler mittels eines sog. Staatstrojaners unbemerkt Zugriff auf Digitalgeräte erhalten und diese aus der Ferne kontrollieren oder Daten abgreifen. Im Falle der verbreiteten Nutzung von Datenbrillen würde das Arsenal staatlicher Überwachungsmöglichkeiten um die Möglichkeit der Fernkontrolle von Datenbrillen um ein weiteres mächtiges Überwachungsinstrument erweitert werden (Kire 2020). Vermutlich wäre zwar ein rechtskonformer Einsatz einer derartigen Überwachung möglich, doch würden sich dann dieselben Herausforderungen, etwa hins. der Gefahr einer staatlichen Dauerüberwachung, stellen, wie sie bereits unter Abschnitt 3.4 diskutiert wurden.

Zugleich könnte der staatliche Zugriff auf Datenbrillen zu Überwachungszwecken auch einer missbräuchlichen Nutzung durch Kriminelle Vorschub leisten. Denn Zugriff auf die Geräte erhalten Ermittler, indem sie vorhandene Sicherheitslücken ausnutzen. Informationen über derartige Sicherheitslücken werden wiederum zumeist auf Grau- und Schwarzmärkten zum Kauf angeboten. Der staatlich-behördliche Rückgriff auf derartige Sicherheitslücken birgt allerdings zugleich die Gefahr, dass auch andere Personen oder Gruppen diese für illegale Zwecke missbrauchen **können**. Denn das Melden einer Sicherheitslücke würde zwar die IT-Sicherheit der Nutzerinnen und Nutzer erhöhen, doch ist es konträr zum Überwachungsinteresse der Behörden, die ihre Überwachungstätigkeit gerade nicht fortsetzen könnten, wenn die Lücke geschlossen würde. Daher wird staatlichen Überwachungsbehörden seitens der digitalen Zivilgesellschaft der Vorwurf gemacht, sie würden eine Gefährdung der IT-Sicherheit aller in Kauf nehmen bzw. forcieren (Kire 2020). Ausgenutzt werden derartige **Lücken** allerdings auch von Cyberkriminellen (Ablon und Andy Bogart 2017). So könnte bspw. ein unberechtigter Fernzugriff auf die (Gesichtserkennungs-)Funktionalität von Datenbrillen möglich werden. Dadurch könnte sowohl die Privatheit des Trägers als auch der Personen in dessen Sichtfeld verletzt werden. Weil Datenbrillen in beliebigen Situationen und Kontexten getragen werden können, ergeben sich hier ebenso viele Möglichkeiten der missbräuchlichen Nutzung, ob in privaten, beruflichen oder öffentlichen Kontexten.

3.10.3.4. Schwierige Entziehbarkeit

Anders als bei den anderen Anwendungsgebieten dieser Studie birgt das Szenario der Jedermann-Identifikation die enorme Gefahr, sich der Überwachung nicht entziehen zu können. Während öffentliche Plätze, Stadien oder Diagnose-Apps unter Inkaufnahme der entsprechenden Konsequenzen gemieden werden **könnten,** wäre eine Vermeidung der Begegnung mit Datenbrillenträgern weitaus schwieriger. Sofern das Tragen einer Datenbrille in der Öffentlichkeit nicht vollständig verboten und stattdessen etwa ausschliesslich der Einsatz einer Gesichtserkennungssoftware auf Datenbrillen verboten würde, könnten sich Menschen im Sichtfeld einer Datenbrille nie wirklich sicher sein, ob sie nicht doch aufgezeichnet bzw. identifiziert werden, weil der Träger illegalerweise doch eine entsprechende Software nutzt. Wollten sich Betroffene Sicherheit verschaffen, müssten sie auf Selbstdatenschutzpraktiken setzen, etwa in Form von Maskierung oder Gesichtsverschleierung – was jedoch für viele Menschen nicht infrage kommen dürfte (vgl. auch die Ergebnisse zur Bevölkerungsumfrage unter Abschnitt 5.2.8). Wie die «Glasshole»-Diskussionen im Rahmen von

Google Glass gezeigt haben (Lobo 2013), wären wohl v.a. soziale Konflikte über das Tragen von Datenbrillen in der Öffentlichkeit zu erwarten.

3.10.3.5. Wissens- und Machtgefälle in sozialen Situationen

Ein heute noch sehr fiktives, aber zukünftig durchaus mögliches Szenario betrifft das Wissens- und Machtgefälle zwischen den Nutzern fortgeschrittener AR-Kapazitäten von Datenbrillen und dem Rest der Gesellschaft. Mit der fortschreitenden Entwicklung von künstlicher Intelligenz und der weiteren Technologieminiaturisierung könnten Datenbrillen in die Lage versetzt werden, den Trägern in Echtzeit kontextspezifische Informationen einzublenden. Das romantische erste Kennenlernen, bei dem vermeintlich zufällig ein gemeinsames Interessengebiet Thema war, könnte dann Ergebnis einer per Gesichtserkennung erfolgten Identifizierung und der Durchleuchtung des sozialen Lebens sein. Wenn mittels Spracherkennung automatisch alles, was das Gegenüber spricht, aufgezeichnet wird, könnte die eingesetzte Software darauf ausgerichtet werden, das Gespräch mit Zusatzinformationen zu versorgen. Wenn bspw. eine Frage gestellt wurde, könnten passende Antworten aus dem Internet abgerufen und vorgeschlagen werden. Seltene oder Fachbegriffe könnten erläutert werden bis hin zu der Möglichkeit, mittels eines personalisierten KI-Coaches die idealen Antworten und Reaktionen errechnet zu bekommen, um soziale Situationen zu den eigenen Gunsten zu manipulieren (TheCGBros 2012). Eingewendet werden könnte dagegen, dass sich das Problem vermutlich in dem Masse verringern würde, in dem diese Möglichkeiten möglichst vielen Menschen zugänglich gemacht würden, sodass das Wissensgefälle nicht einseitig ausfiele. Ausgehend von heutigen Entwicklungstendenzen wäre aber vermutlich am ehesten eine Art technisches Wettrüsten zu erwarten: Diejenigen, die sich die neuesten Datenbrillen mit den grössten Rechenkapazitäten und die teuersten KI-Coaching-Apps leisten könnten, wären im Vorteil gegenüber jenen, die es sich nicht leisten könnten oder die im Unklaren über das Manipulationspotenzial wären.

3.10.4. Zwischenfazit

Wie die Ausführungen zur Jedermann-Identifikation gezeigt haben, handelt es sich dabei um eine sowohl rechtlich als auch ethisch hochproblematische Möglichkeit der Nutzung von Gesichtserkennungstechnologien.

Sofern das Tragen von Datenbrillen in der Öffentlichkeit erlaubt bliebe, wäre nicht auszuschliessen, dass Gesichtserkennungstechnologie zum Zwecke der Jedermann-Identifikation zum Einsatz kommt. In diesem Fall wäre es fraglich, wie der Schutz der Rechte der Bevölkerung gewährleistet werden könnte. Das Tragen von gesichtserkennungshemmendem Make-up oder Maskierung käme nur für einen kleinen Bevölkerungsteil infrage (vgl. Kapitel 5.2.2 und 5.2.4) und würde – sofern sie denn zuverlässig funktioniert – auch nur diese Personen vor der Erkennung bewahren. Problematisch wäre ein solches Setzen auf Selbstschutzmassnahmen, weil es die staatliche Schutzverantwortung auf den Einzelnen abwälzen würde. Entscheidungstragende müssten aber auch den Schutz der breiten Bevölkerung im Blick behalten.

Gesetzliche Regulierung in Form bspw. eines Verbots des Tragens von Datenbrillen in der Öffentlichkeit oder der Verknüpfung von Datenbrillen mit Gesichtserkennungstechnologie könnte deren Verbreitung zwar massiv eindämmen, jedoch würden auf diese Weise auch erwünschte Nutzungsweisen erstickt. Gleichzeitig könnten Datenbrillen, sofern sie für legale Zwecke erlaubt und käuflich erwerbbar wären, auch für den Zweck der Jedermann-Identifikation umgerüstet werden. Dieses Dilemma weisst auf einen Regulierungsbedarf hin, deren Details weiterer sorgfältiger Klärung bedürfen.

Eine Massnahme der gesellschaftlichen Selbstregulierung wäre, dass im Ergebnis eines gesellschaftlichen Diskurses neue soziale Normen etabliert werden, die das Tragen von Datenbrillen sozial normieren. So könnten zukünftig z.B. Hinweisschilder darauf aufmerksam machen, dass das Tragen an gewissen Orten unerwünscht ist oder – umgekehrt – dass eben damit zu rechnen ist.

Jegliche staatliche Regulierung oder gesellschaftliche Selbstregulierung hins. des Tragens würde allerdings dann an ihre Grenzen stossen, sobald Datenbrillen-Technologie einen derartigen Miniaturisierungsgrad erreicht, dass Datenbrillen mit blossem Auge nicht mehr von herkömmlichen Brillen zu unterscheiden wären.

4. Die Perspektive von Bürgerinnen und Bürgern in Fokusgruppen

Zur qualitativen Untersuchung der gesellschaftlichen Wahrnehmung von Stimm, Sprach- und Gesichtserkennung führte die Stiftung für Technologiefolgenabschätzung TA-SWISS gemeinsam mit dem Projektteam Fokusgruppen durch.

Im Vordergrund standen dabei Fragen nach einem besseren Verständnis dessen, welche Aspekte der besprochenen Stimm-, Sprach- und Gesichtserkennungsanwendungen aus welchen Gründen als erwünscht oder unerwünscht gelten und wie die Schweizer Politik und Gesellschaft mit diesen Fragen umgehen sollte.

Im ersten Abschnitt (4.1) werden Ziele und Ablauf der Fokusgruppen vorgestellt. Im Anschluss folgen die Ergebnisse der Diskussionen der Anwendungsgebiete (4.2). Gemeinsame und querliegende Vor- und Nachteile sowie Empfehlungen werden in Abschnitt 4.3 zusammengefasst und darauf basierende Schlussfolgerungen (4.4) gezogen.

4.1. Ziele und Ablauf der Fokusgruppen

Die Fokusgruppen fanden am 20. Oktober 2021 in Bern statt. Die Rekrutierung der insgesamt 15 Teilnehmenden (Tabelle 6) erfolgte durch die TA-SWISS-Geschäftsstelle. Die Bewerbung der Veranstaltung erfolgte zum einen über das TA-SWISS-Netzwerk und zum anderen über Aushänge im lokalen Detailhandel. Dabei wurde auch kurz über Inhalt und Zielsetzung der Fokusgruppen informiert.

Tabelle 6: Soziodemografische Merkmale der Fokusgruppen-Teilnehmenden

	Weiblich	**Männlich**	**Total**
Alter:			
26–35	1	2	**3**
36–45	–	1	**1**
46–55	3	2	**5**
56–65	–	1	**1**
66+	1	1	**2**
k. A.		3	**3**
Sprache der Fokusgruppe:			
Deutsch	3	7	**10**
Französisch	2	3	**5**

	Weiblich	Männlich	Total
Teilnahme:			
Halbtägig	3	6	**9**
Ganztägig	2	4	**6**
Total	**5**	**10**	**15**

Die Veranstaltung bestand aus einem Vormittags- und einem Nachmittagsprogramm. Die Teilnehmenden konnten sich entweder an einem Halbtag oder ganztags beteiligen. Die Fokusgruppen wurden sowohl auf Deutsch als auch auf Französisch angeboten.

Nachdem die Teilnehmenden der Fokusgruppen im Plenum begrüsst worden waren, führte die Projektgruppe im Laufe des Vor- und Nachmittags insgesamt fünf Fokusgruppen durch, wobei in jeder Gruppe jeweils zwei Themen diskutiert wurden. Somit fanden über den Tag verteilt insgesamt zehn Fokusgruppen-Diskussionen statt.

Jede Diskussion zu einem Anwendungsgebiet dauerte zwischen 70–80 Minuten. Sechs Diskussionen wurden auf Deutsch und vier auf Französisch durchgeführt. An jeder Fokusgruppe nahmen drei bis fünf Personen teil. Daneben beteiligten sich zwei bis vier weitere Personen aus dem Projektteam und der TA-SWISS-Geschäftsstelle zwecks Moderation und Dokumentation der Diskussionen (Tabelle 7).

Die überwiegende Mehrheit der Teilnehmenden hatte einen akademischen Hintergrund (Fachhochschule oder Universität), eine Person verfügte über einen Berufsschulabschluss. Das konkrete Vorwissen zu den diskutierten Themen wurde zwar nicht abgefragt, doch es stellte sich bei Projektteam und TA-SWISS-Geschäftsstelle der Eindruck ein, dass die Teilnehmenden mehrheitlich gut informiert über das Themengebiet der Stimm-, Sprach- und Gesichtserkennung waren.

Tabelle 7: Übersicht über die Zusammensetzung und die Themen der einzelnen Fokusgruppen

Gruppe	Teilnehmende	Sprache	Diskussionsthema
1	5	Deutsch	Stadionüberwachung
			Erkennung psychischer Krankheiten
2	3	Französisch	Smarte Lautsprecher
			Polizeiliche Überwachung
1	4	Deutsch	Smarte Lautsprecher
			Jedermann-Identifikation
2	5	Deutsch	Polizeiliche Überwachung
			Authentifizierung via Stimme
3	5	Französisch	Emotionserkennung/Aufmerksamkeitsanalyse
			Erkennung physischer Krankheiten

Die Struktur aller zehn Fokusgruppen-Diskussionen war identisch: Als Erstes wurden den Teilnehmenden die grundlegenden Rahmenbedingungen des jeweiligen Anwendungsgebiets seitens der Moderation in Form eines kurzen fachlichen Inputs nähergebracht. Alle Fach-Inputs hatten zum Ziel, die Funktionsweise der Stimm-, Sprach- und Gesichtserkennung im jeweiligen Anwendungskontext zu beschreiben und die damit verbundenen möglichen Vor- und Nachteile in möglichst ausgewogener Form anzudiskutieren. Die Fach-Inputs waren zudem bebildert und bei zwei der Inputs wurden auch Videos gezeigt, um ein besseres Verständnis zu ermöglichen. Im Anschluss leitete die Moderation zur Gruppendiskussion über, indem die Teilnehmenden gefragt wurden, was sie über das Anwendungsgebiet denken und wie ihrer Meinung nach mit dem Thema umgegangen werden sollte.

Die Zielsetzung der Fokusgruppen war es, die mit dem Anwendungsgebiet verbundenen Chancen bzw. Hoffnungen und Risiken bzw. Ängste der Teilnehmenden in Erfahrung zu bringen. Von besonderem Interesse war dabei, wie die mögliche Abwägung zwischen positiven und negativen Aspekten erfolgte und welche Aspekte aus welchen Gründen überwogen. Des Weiteren ging es darum, herauszufinden, welchen Umgang mit Stimm-, Sprach- und Gesichtserkennungstechnologien sich die Teilnehmenden wünschten. Sofern die Teilnehmenden nicht bereits von alleine die im Fokus der Projektgruppe stehenden Punkte diskutierten, wurde die Diskussion zu den entsprechenden Fragen seitens der Moderation angestossen.

Die Diskussionen wurden zum einen seitens zusätzlicher Personen aus dem Projektteam sowie der TA-SWISS-Geschäftsstelle vor Ort protokolliert. Zum anderen wurde nach Einholung der Einwilligung der Teilnehmenden eine Sprachaufzeichnung der jeweiligen Diskussionen erstellt, auf die nötigenfalls bei der Ergebnisdarstellung zurückgegriffen und die im Anschluss gelöscht wurden.

4.2. Diskussion der Anwendungsgebiete

Im Vordergrund der Ergebnisdarstellung der Fokusgruppen stehen die hervorstechendsten Punkte aus den Diskussionen im Hinblick auf Vor- und Nachteile sowie Empfehlungen.

4.2.1. Smarte Lautsprecher

Tabelle 8: Zusammenstellung der Fokusgruppe zu «smarte Lautsprecher»

Thema	Sprache	Teilnehmende
Smarte Lautsprecher	Deutsch	4
	Französisch	3

4.2.1.1. Kurzzusammenfassung

Die französischsprachige Diskussionsgruppe bestand aus drei Teilnehmenden, wovon zwei Mitglieder einen smarten Lautsprecher privat nutzten. Die Gruppe war sich einig, dass die rote Linie betreffend das Gefühl überwacht zu werden und des damit einhergehenden nötigen Schutzes der Privatsphäre überschritten wurde. Vorteile der smarten Lautsprecher wurden in der Vereinfachung des Alltags und dem dadurch entstehenden Zeitgewinn gesehen, auch der Spass und das Interesse an Technologien spielte bei den Nutzern eine Rolle. Ein Gesprächsteilnehmer würde diese Geräte nur im Falle absoluter Notwendigkeit verwenden (etwa als Babyfon). Alle Teilnehmenden fanden, dass die Datenerhebung und die Datennutzung durch die Anbieter transparenter gemacht werden sollten und die bestehenden gesetzlichen Vorschriften verschärft bzw. konsequenter durchgesetzt werden sollten. Die Wichtigkeit regelmässiger Sicherheitsupdates wurde ebenfalls hervorgehoben.

Zwei der vier Personen aus der deutschsprachigen Diskussionsgruppe hatten bereits einen smarten Lautsprecher genutzt. Bei einer Teilnehmerin wurde ein Google Home durch ein Familienmitglied aufgestellt, sie selbst sah jedoch keinerlei Nutzen darin und würde ein solches Gerät nie selbst anschaffen. Die zweite Teilnehmerin hatte hingegen aus Interesse einen Nest-Lautsprecher (Google) verwendet, die Nutzung jedoch sofort eingestellt, nachdem sie den Datenaustausch aus technischer Sicht analysiert hatte.

Grundsätzlich standen die Teilnehmer der deutschsprachigen Diskussionsgruppe der Nutzung von smarten Lautsprechern ebenfalls eher kritisch gegenüber. Drei der vier Personen sahen keinen konkreten Nutzen eines solchen Geräts. Das Thema Datenschutz kam direkt in der ersten Aussage zum Anwendungsfall zur Sprache und zog sich durch die ganze weitere Diskussion. Die Gefahr der Dauerüberwachung und Intransparenz der Datenerhebung und Datenauswertung stellten Leitthemen der Befragten dar. Dies spiegelt sich auch in der Vielzahl der geäusserten Nachteile wider. Vorteile wurden erst auf explizite Nachfrage geäussert. Jegliche Vorteile wurden mit dem Argument infrage gestellt, dass für einen relativ geringen Nutzen sehr viele Daten erhoben würden. Trotz der erheblichen Kritik an smarten Lautsprechern tendierten die Teilnehmenden aber eher dazu, die Geräte stärker zu regulieren und Missbrauch zu unterbinden statt den (in Zukunft möglichweise zunehmenden) Nutzen zu verhindern.

4.2.1.2. Nachteile

Ein grosser Kritikpunkt in beiden Diskussionsgruppen war die fehlende Transparenz. Laut den Teilnehmenden sei es völlig unklar, welche (Meta-)Daten zusätzlich erhoben und wie diese ausgewertet werden.

Die Sorge des Dauermithörens wurde in beiden Gruppen als kritisch empfunden. Die Bedienung eines smarten Lautsprechers per Stimme sei sehr «niederschwellig» und «bequem». Im eigenen Zuhause verhalte man sich eher spontan und denke selten über das Gesprochene nach, sodass in unbewusster Weise mehr Daten preisgegeben würden. Auch würden die Geräte umso nützlicher, je mehr andere Geräte damit verbunden und damit mehr Daten erhoben würden (*smart home*), was wiederum zu einer verstärkten Nutzung und Gewöh-

nungseffekten führe. So sei es auch besonders kritisch, wenn weitere Aufnahmen von z.B. Kindern oder anderen Dritten als «Beifang» erhoben und ausgewertet werden. Metaphorisch wurde hier der smarte Lautsprecher in der deutschsprachigen Gruppe als ein «Staubsauger» bezeichnet, der «alle Daten aufsaugt». Geäussert wurden auch Befürchtungen um die Datensicherheit bzw. Angst vor Hackerangriffen (Einbruch unter Ausnutzung von Smart-Home-Schwachstellen, Passwortdiebstahl).

Ein Teilnehmer der französischsprachigen Gruppe äusserte sich dahin gehend, dass er zwar von dieser Technologie fasziniert sei, sie jedoch auch ein grosses Risiko darstelle. Die privaten Anbieter hätten sehr viel Geld und ihre Geschäfte seien nur schwer kontrollierbar. Die Schweiz könne diese nicht alleine regulieren. Weiter wisse man nicht, auf welchen Servern die entsprechenden Daten gespeichert werden. Ein anderer Teilnehmer präzisierte, es sei insb. die Frage, ob die entsprechenden (Datenschutz-)Gesetze auch umgesetzt und im Falle der Übertretung abschreckende Strafen ausgesprochen würden.

Die Teilnehmenden der französischsprachigen Gruppen waren sich einig, dass es wichtig sei, dass jede Person selbst die Kontrolle über die Daten habe und diese auch löschen lassen könne. Zwar bestünde ein gesetzlicher Zugang, aber es sei fraglich, ob man alle Daten erhalte und was man damit machen könne.

In ähnlicher Weise wurde in der deutschsprachigen Diskussion betont, dass Hersteller global agierten und die Datenbearbeitung nicht in Europa oder in der Schweiz stattfinde, was zusätzliches Misstrauen schaffe. Weiter sei unklar, nach welchen Regeln das System arbeitet und warum es trotz gleicher Fragestellung zu unterschiedlichen Ergebnissen kommt. In diesem Kontext wurde auch das Profiling angesprochen. Konsens bestand darüber, dass hierfür keine rein personenbezogenen Daten mehr erhoben werden müssen, sondern mittels anonymisierter Daten eine Gruppenzugehörigkeit festgestellt werden kann. Dadurch, dass solche anonymisierten Daten nicht unter das Datenschutzrecht fallen, werde dessen Schutzfunktion untergraben. Im weiteren Verlauf der deutschsprachigen Diskussion wurden zunehmend grundlegende Aspekte des Datenschutzes besprochen, insb. dass der Begriff eines «personenbezogenen Datums» nicht endlich definiert werden und eine Anonymisierung in der Praxis wirkungslos sein könne, wenn auch Datenbearbeitungen, die auf anonymisierten personenbezogenen Daten beruhen, unmittelbare Folgen für Individuen entfalten könnten.

Auf die Frage, welche konkreten und individuellen Missbrauchspotenziale die Teilnehmenden sehen, antworteten diese, dass man nicht von unmittelbarem (z.B. körperlichem) Schaden sprechen könne, sondern von Manipulation. Als Beispiel wurde eingebracht, dass sich die Suchergebnisse unterscheiden würden, wenn eine männliche oder eine weibliche Stimme die (gleiche) Frage stellt. Diese Probleme seien aber weniger relevant, wenn es sich nur um Befehle zur Steuerung des Smart Home handele. Sehr wohl relevant wurde indes die Befürchtung bewertet, dass durch Datenaggregation und Austausch mit anderen Dienstleistern negative Konsequenzen folgen könnten. «Die individuelle Datenspur wird mit ausgewertet und dadurch könnte ich später Nachteile erleiden.» Bei der Menge der erhobenen Daten trauten sich die Teilnehmenden auch nicht zu, einen Überblick über ihre Daten zu behalten. Deshalb liessen sich einmal erhobene Daten nie wieder komplett löschen. Weitere Bedenken betrafen personalisierte Preise auf Basis zuvor erzielter Erkenntnisse aus Stimmdaten sowie Wettbewerbsverzerrungen und Manipulation des Handelns, da ein

smarter Lautsprecher oft nur ein einziges Ergebnis anzeige. Als besonders unangenehm empfanden die Teilnehmenden es, falls für solche Analysen Emotionen aus den Stimm- und Sprachdaten herausgelesen werden.

Auch in der französischsprachigen Gruppe wurde die Frage einer möglichen Diskriminierung durch smarte Lautsprecher angesprochen. So wurde angenommen, dass Alexa vielleicht mehr Schwierigkeiten habe, Frauen oder Personen mit Beeinträchtigungen beim Sprechen zu erkennen, weil derartige Software häufig auf die Stimmen gesunder Männer zugeschnitten sei. Zudem sahen die Teilnehmenden Gefahren des Missbrauchs durch die privaten Anbieter. Demnach könnten diese versuchen, die Meinung der Nutzer zu manipulieren (so zeige Google bei der Recherche je nach Nutzer ja auch andere Ergebnisse). Weitere zukünftige Skandale – wie Cambridge Analytica (politisches Micro-Targeting auf sozialen Onlinenetzwerken zur Wahlmanipulation) – wären für eine Teilnehmerin auch ein möglicher Grund, die weitere Nutzung smarter Lautsprecher einzustellen.

In der deutschsprachigen Gruppe wurde zudem argumentiert, dass die «Stimme mehr als nur Inhalt» sei und das persönlichste Merkmal, das ein Mensch hat und sich daraus viel über das Individuum ableiten liesse, wie z.B. der Gesundheitszustand. Die in der Gegenwart erhobenen Daten könnten in der Zukunft nachteilig für die Person sein, z.B. wenn es darum geht, eine Versicherung abzuschliessen. Eine Befragte äusserte sich ganz konkret dazu, dass die Aufzeichnung einer solch grossen Menge und Qualität an Daten nicht gerechtfertigt sei, nur um das Licht einzuschalten.

Die in der deutschsprachigen Diskussion gestellte Frage, welche Art von Daten oder Datentypen die Teilnehmenden als besonders kritisch erachten, wurde von einer Person damit beantwortet, dass es per se keine schlechten oder guten Daten gäbe, weil es darauf ankomme, was aus den Daten gemacht wird. Kritisch bewerteten alle Teilnehmenden, dass niemand das Potenzial der erhobenen Daten kenne und auch nicht, wie diese aktuell oder in Zukunft ausgewertet werden könnten. Beispielsweise könnten zwei Datenpunkte alleine nicht relevant sein. Durch eine Zusammenführung könnten jedoch sehr aussagekräftige Schlüsse aus denselben Daten gezogen werden.

Weiter wurde in beiden Gruppen aufgeworfen, dass der allmähliche gesellschaftliche Übergang zur Sprachsteuerung bei nachfolgenden Generationen eine zunehmende Abhängigkeit von sprachsteuerungsbasierten Systemen schaffen und damit zu einer Dealphabetisierung führen könnte.

Weitere negative Aspekte, die diskutiert wurden, waren die Datensicherheit (Hacker könnten versuchen, Passwörter zu Bankkonten zu erhalten oder einzubrechen, wenn niemand zu Hause ist), oder der Verkauf von Daten an Dritte. Einen negativen Effekt habe die Nutzung der Geräte auch auf die Umwelt, da die Server viel Energie benötigten und damit bei einer vermehrten Nutzung auch der Energiebedarf ansteigen würde.

Darüber hinaus wurde in der französischsprachigen Gruppe auch diskutiert, dass für gewisse Personengruppen der Umgang mit den Geräten schwierig sein könnte, da sie etwa nicht wüssten, wie sie mit dem Gerät sprechen sollten (insb. ältere Menschen). Zudem wurden Probleme beim Erkennen von Akzenten und Dialekten thematisiert. Damit sie ihre Nutzer

besser verständen, könnte man zwar die eigene Stimme speichern. Das sei aber im Hinblick auf den Datenschutz bedenklich.

4.2.1.3. Vorteile

In beiden Gruppen stand im Hinblick auf die Vorteile smarter Lautsprecher der Komfortgewinn im Vordergrund. Ein Teilnehmer der französischsprachigen Diskussion teilte mit, dass er Alexa aus Spass und Freude an der Technologie verwende. Eine andere Nutzerin aus derselben Gruppe führte zudem an, dass smarte Lautsprecher sehr praktisch seien, weil man während des Kochens die Musik leiser machen oder den Fernseher einstellen könne. Die Frage nach konkreten wünschenswerten aktuellen Anwendungsfällen beantworteten die Teilnehmenden der deutschsprachigen Gruppe mit der Steuerung eines Smart Homes oder der Möglichkeit, Simultanübersetzungen, Gruppentelefonkonferenzen oder automatisierte Protokollerstellung in Geschäftsbesprechungen durchführen zu können. Gleichzeitig merkte eine Befragte an, dass sie sich keine Anwendung vorstellen könne, welche nicht schon jetzt mittels Smartphone möglich sei. Anders als in der französischsprachigen Gruppe war die Betonung der Vorteile in der deutschsprachigen Gruppe etwas zurückhaltender. Selbst als Vorteile geäussert wurden, erwähnten die Befragten direkt daraus resultierende negative Konsequenzen. Trotzdem waren sich die Teilnehmenden dieser Gruppe einig, dass die Bedienung eines smarten Lautsprechers per Stimme sehr bequem sei und die Zukunft darstelle. Den Teilnehmenden war auch bewusst, dass sich Vorteile durch künftige Anwendungen entwickeln könnten, die man sich gegenwärtig nicht vorstellen könne.

Als grosser Vorteil der smarten Lautsprecher wurde von einer französischsprachigen Teilnehmerin deren Nutzung für ein unabhängiges Leben beeinträchtigter Personen aufgeführt, insb. für sehbeeinträchtigte Personen oder allgemein als Hilfestellung bei kleineren Tätigkeiten, wie etwa das Wechseln des Fernsehkanals. Die Selbstbestimmung der Betroffenen könnte dadurch stark erhöht werden. Die Gruppe befand den Einsatz der Lautsprecher auch als positiv, wenn diese als Unterstützung für ältere Personen genutzt würden. So könnten diese bei Bedarf, z.B. einem Sturz, den Notruf mithilfe des Lautsprechers wählen. In der deutschsprachigen Gruppe wurde die Möglichkeit des Hilferufs zwar ebenfalls genannt, allerdings wurde auch hier darauf verwiesen, dass es bereits andere Lösungen (z.B. ein Armband) gäbe, welche die gleiche Funktionalität erfüllten.

In der französischsprachigen Gruppe wurde zudem angesprochen, dass einsame Menschen durch die Interaktion mit einem Sprachassistenten möglicherweise das Gefühl erhalten könnten, weniger alleine zu sein und sich die Nutzung somit positiv auf die Psyche auswirkt.

4.2.1.4. Empfehlungen

Die Teilnehmenden beider Gruppen sprachen sich dafür aus, dass die Einsatzzwecke der smarten Lautsprecher deutlich spezifischer bestimmt sein sollten. Die Nutzung der erhobenen Daten für illegitime Zusatzzwecke sollte gesetzlich verboten sein. Als Beispiel hierfür wurde in der deutschsprachigen Gruppe das Einschalten des Lichts mittels Sprachbefehl

genannt. Wenn die dabei erhobenen Stimm- und Sprachdaten für andere Zwecke als das Einschalten des Lichts genutzt würden, würde eine rote Linie überschritten, was verboten sein müsse. In Verbindung mit dieser Empfehlung wurde auch dafür votiert, dass generell deutlich weniger Daten erhoben werden, damit die Gefahr der missbräuchlichen Nutzung dieser Daten minimiert wird. Insb., so die Teilnehmenden der französischsprachigen Gruppe, müsste sichergestellt werden, dass diese Geräte nicht zu politischen Zwecken bzw. zur Manipulation der Bevölkerungsmeinung verwendet werden. Generell solle ein Opt-In zur Anwendung kommen und Opt-Outs sollten gespeichert werden.

Aus dem Unverständnis, das in der deutschsprachigen Gruppe dem Umstand entgegengebracht wurde, dass für einfache Sprachbefehle, wie das Einschalten des Lichts, Daten auf Server übertragen werden, folgte die Empfehlung, dass möglichst viele Daten lokal auf dem jeweiligen Gerät selbst bearbeitet werden sollten (Edge Computing). Wünschenswert sei es auch, heimische bzw. europäische Anbieter im Entwickeln von datenschutzkonformen Konkurrenzprodukten seitens der Politik zu unterstützen.

Die Teilnehmenden der deutschsprachigen Diskussion sprachen sich auch dagegen aus, dass persönliche Profile erstellt werden, damit z.B. keine personalisierte dynamische Preisanpassung für Produkte erfolgen kann. Doch zwei Teilnehmende merkten an, dass heutzutage v.a. Gruppenprofile auf Basis anonymisierter personenbezogener Daten erstellt werden, die aus dem Datenschutzrecht herausfallen, indem nicht personenbezogene Datenpunkte zueinander in Bezug gesetzt werden (Personen, die X kaufen, kauften auch Y und Z). Daraufhin wurde darüber diskutiert, ob nicht auch diese Daten als personenbezogen anzusehen seien, weil sie doch personen- bzw. gruppenbezogene Effekte hätten. Der Vorschlag, den Anwendungsbereich des Datenschutzrechts entsprechend auszweiten, erhielt weitestgehend Zustimmung, doch eine Person sprach sich auch dagegen aus, weil sie befürchtete, dass dann keine Differenzierung mehr zwischen personenbezogenen und nicht personenbezogenen Daten möglich wäre.

In der französischsprachigen Gruppe wurde ausserdem der Umsetzung der geltenden Gesetze (etwa auch im Hinblick auf die Einhaltung von Löschpflichten) eine zentrale Bedeutung zugeschrieben und dafür votiert, dass bei Gesetzesverstössen privater Unternehmen hohe Geldstrafen vorgesehen werden sollten, welche für das Unternehmen deutlich spürbar und abschreckend wären.

Ein zweiter zentraler Punkt in beiden Gruppendiskussionen betraf die Schaffung von voller Transparenz über die Art, Qualität und Anzahl der Daten sowie deren Übertragung und Speicherung. Die Teilnehmenden der deutschsprachigen Diskussion sprachen hier besonders detaillierte Empfehlungen aus und befürworteten die Schaffung einer Art von «Beipackzettel für smarte Lautsprecher», in welchem diese Informationen verständlich beschrieben sind. Gleichzeitig sahen sie das Problem, dass dieser «Beipackzettel» insb. für Einblicke in die Datenbearbeitung wichtig sei und nicht rein für die Datenerhebung im Gerät. Daher wurde auch ein Ex-post-Datenreport für erforderlich erachtet, der in regelmässigen Abständen an die Nutzenden versendet wird und transparent und leicht verständlich zusammenfasst, welche Daten von welchem Anbieter erhoben und an welche anderen Stellen weitergereicht wurden. Neben der Schaffung von Transparenz, solle ein solcher Report die Nutzung der Betroffenenrechte vereinfachen, indem bspw. Löschungswünsche unmittelbar an die jeweiligen Stellen kommuniziert werden können. Dazu wurde es auch als erforderlich angese-

hen, dass der Bericht nicht bloss die Angaben der Hersteller des smarten Lautsprechers beinhaltet, sondern auch Informationen zu jeder auf dem Gerät genutzten Anwendung. Im Zusammenhang mit Transparenz wurden auch mehr individuelle Wahl- bzw. Einstellungsmöglichkeiten gefordert. Weil es nicht ersichtlich sei, weshalb zwei Nutzenden zwei unterschiedliche Ergebnisse angezeigt werden, müsse Transparenz über derartige Ergebnisse hergestellt und den Nutzenden die Möglichkeit eröffnet werden, hier eine bewusste Auswahl zu treffen.

In beiden Gruppen wurde eine zu grosse Machtanhäufung bei den Anbietern für smarte Lautsprecher gesehen. Es dürfe keinen Anbieter geben, der «auf seiner Plattform alles kanalisiert». Dementsprechend wurde in der französischsprachigen Diskussion vorgeschlagen, dass die Software smarter Lautsprecher *open source* sein sollte. In der deutschsprachigen Diskussion wurde in ähnlicher Weise eine Entkopplung von Software und Hardware vorgeschlagen, wonach die Hersteller der Geräte offene Schnittstellen anbieten sollten, an denen andere Anbieter (z.B. mit eigenen App-Stores) andocken könnten. So könnten spezielle smarte Lautsprecher z.B. für ältere Menschen entwickelt werden, welche nur den wichtigsten Funktionsumfang haben (z.B. Hilferuf nach Sturz) und die bspw. vom Schweizerischen Roten Kreuz entwickelt werden. In diesem Zusammenhang wurde in beiden Gruppen auch problematisiert, dass eine solche Öffnung mit hoher Wahrscheinlichkeit Probleme im Hinblick auf die Datensicherheit mit sich bringen werde. Ein weiterer aus der französischsprachigen Gruppe stammender Vorschlag sah in diesem Zusammenhang vor, dass eine «trusted platform» (z.B. Datentreuhänder) als Filter zwischen Nutzende und Diensteanbieter geschaltet wird, die personenbezogene oder emotionsbezogene Daten entfernt und dann neutrale Daten an die externen Diensteanbieter zur Bearbeitung des Sprachbefehls weitergibt. Wenn es mehrere Datentreuhänder gäbe, könnten zudem auch unterschiedliche Datenpreisgabemodelle unterstützt werden, sodass sich Nutzende zwischen Voreinstellungen für einen freizügigeren oder vorsichtigeren Umgang mit ihren Daten entscheiden können.

In der französischsprachigen Gruppe wurde auch über die bessere Information der Nutzenden über die Notwendigkeit regelmässiger Sicherheitsupdates gesprochen. Problematisch dabei sei, dass die Anbieter vieler billiger Geräte die Software nicht lange genug pflegten. Nach Ablauf des Support-Zeitraums gebe es dann keine Updates und die Geräte würden deutlich anfälliger für Angriffe.

In der französischsprachigen Gruppe wurde zudem diskutiert, dass im Forschungsfeld der Informatik mehr Diversität nötig sei, auch um unbeabsichtigte Verzerrungen zu verhindern, etwa wenn Alexa nach den grössten Genies gefragt werde, sollten nicht nur Männer genannt werden, damit Mädchen schon früh in den Bereichen Mathematik und Naturwissenschaften gefördert werden.

Die Diskutierenden der französischsprachigen Gruppe waren sich einig, dass sowohl dem Staat, den Bürgern und auch zivilen Organisationen die Verantwortung zukomme, zu kontrollieren, wie und was für Systeme private Unternehmen verwenden. Die Schweiz sei dazu zu klein; vielmehr sollte auf europäischem Niveau die Frage der Regulation diskutiert und umgesetzt werden. Zudem sollten die Server in Europa stehen. Zugleich solle Unternehmen in Europa die Möglichkeit gegeben werden, genügend Daten zu erheben, damit sie datenschutzfreundliche Konkurrenzprodukte entwerfen können. Die bestehenden Gesetze

würden die Unternehmen teilweise zu stark bei der Entwicklung hindern. Allfällige Gesetze sollten zudem technikneutral sein.

Eine Person aus der deutschsprachigen Gruppe sprach sich zudem für die Abkehr von der Gratismentalität aus: Wenn Nutzende bereit wären, einen geringen Betrag (z.B. USD 3) monatlich zu bezahlen, könne ein System ohne personalisierte Werbung auskommen. Hier müssten ggf. Staat und Bildungsinstitutionen die Bürgerinnen und Bürger für einen bewussteren Konsum sensibilisieren.

4.2.2. Polizeiliche Überwachung

Tabelle 9: Zusammenstellung der Fokusgruppe zu «polizeiliche Überwachung»

Thema	Sprache	Teilnehmende
Polizeiliche Überwachung	Deutsch	5
	Französisch	3

4.2.2.1. Kurzzusammenfassung

Die Teilnehmenden beider Diskussionsgruppen standen dem Einsatz der Gesichtserkennungstechnologie durch die Polizei sehr kritisch gegenüber. In der französischsprachigen Gruppe waren die Bedenken gross, dass die Polizei mit dieser Technologie zu viel Macht erhalte; eine Massenüberwachung durch den Staat müsse zwingend verhindert werden. Die deutschsprachige Gruppe argumentierte, der Technikeinsatz greife zu stark in die Privatsphäre der Menschen und ihre Grundrechte ein.

V.a. wurde kritisch reflektiert, ob die Überwachungsmassnahmen wirklich die gewünschte Sicherheit gewährleisten, indem sie Anschläge oder Verbrechen verhindern können oder doch nur eine «gefühlte» Sicherheit vermitteln.

Die Mehrheit der deutschsprachigen Teilnehmenden konnte sich aufgrund eigener Erfahrungen oder der genannten «Überwachungssituation» in der eigenen Stadt identifizieren. Breite Zustimmung erhielt das Argument, dass der Aspekt Vertrauen in die Regierung und/oder in die Technologien mit der Befürwortung oder Ablehnung von Überwachungsmassnahmen seitens polizeilicher Stellen korreliert.

In der französischsprachigen Diskussionsgruppe konnten sich die Teilnehmenden einigen, dass ein Einsatz für klar definierte und gesetzlich geregelte Zwecke denkbar wäre. Weitere wichtige Aspekte waren: Privacy by Design, gute Ausbildung der Polizei und eine Überprüfung der Anwendung der Technologie durch unabhängige Dritte.

4.2.2.2. Nachteile

Die Teilnehmenden der französischsprachigen Gruppe waren sich einig, dass der Einsatz der besprochenen Technologien die Gefahr eines Machtungleichgewichts beinhalte. Ein Teilnehmer fand es gefährlich, solche Instrumente in die Hände von Verantwortlichen zu geben, die bereits tendenziell autoritär seien. Die Überwachung nehme ohnehin stetig zu und es sei von grosser Bedeutung, dass eine Massenüberwachung verhindert werde. Das Argument der Terrorismusbekämpfung sei stark vereinfachend, von Angst geleitet und nicht rational. In der Gesellschaft müssen gewisse Freiheiten bestehen bleiben und dafür sei auch ein gewisses Risiko zu akzeptieren.

In der deutschsprachigen Gruppe kamen Befürchtungen in Bezug auf die Datensicherheit und -bearbeitung auf, etwa darüber, welche Software die Polizei für die Analyse biometrischer Daten nutzt, an welche Akteure Informationen im Hintergrund ggf. weitergereicht werden und wie der Datenaustausch zwischen polizeilichen Stellen und privatem Sektor geregelt ist.

Darüber hinaus erkannte eine Teilnehmerin in der polizeilichen Überwachung die Gefahr, dass alle Bürger und Bürgerinnen unter Generalverdacht gestellt werden und damit kein Fokus mehr auf die Verbrechensaufklärung oder Kriminalprävention gelegt wird. Diskutiert wurde zudem über die sozial(psychologisch)en Auswirkungen von Beobachtungs- und Überwachungsdruck. Demnach könne das Gefühl, überwacht zu werden, Veränderungen im sozialen Verhalten oder der öffentlichen Kommunikation nach sich ziehen. Es wurde befürchtet, dass die Gesichtserkennung Menschen daran hindern könnte, an Demonstrationen teilzunehmen. Daraus resultiere eine Gefahr für die Versammlungsfreiheit und somit auch für die demokratische Grundordnung. Weiter sahen die Teilnehmenden das Risiko, dass mit der Zeit ein Gewöhnungseffekt eintreten könnte.

In der französischsprachigen Gruppe wurde befürchtet, dass sowohl Technologien als auch die anwendenden Autoritäten einem Bias unterliegen könnten, nicht perfekt funktionieren und so zur Diskriminierung gewisser Personengruppen führen. Auch eine Manipulation der Videos (Deepfakes) wurde angesprochen. Wenn der Einsatz der Gesichtserkennung ohne unabhängige Kontrolle erfolge, seien derartige Manipulationen nicht auszuschliessen und auch nicht überprüfbar. Ebenfalls als problematisch angesehen wurde der automatische Datenabgleich gewisser Technologieanwendungen mit Gesichtsbildern aus dem Internet. Dies wäre gesetzlich verboten.

Beim Einsatz der Technologie auf offener Strasse wäre die Zahl der Direktbetroffenen im Vergleich zu den möglichen positiven Aspekten sehr hoch. Es gäbe keine guten Gründe für eine derartige Massenüberwachung. Der Einsatz der Technologie für die Suche von Personen mit suspektem Verhalten wurde dabei als besonders gefährlich eingestuft (Precrime), ebenfalls die Verwendung einer digitalen Bewegungsanalyse der betroffenen Personen.

4.2.2.3. Vorteile

In der deutschsprachigen Gruppe argumentierte ein Teilnehmer, dass er die Gesichtserkennung zur Identifikation von Personen befürworten würde, wenn er so vor Anschlägen und

Kriminalität geschützt sei. Ein Argument für die Ausweitung von Überwachungsmassnahmen sei der Schutz der öffentlichen Sicherheit. Auf Rückfrage der Moderation nach verhältnismässigen Überwachungsmassnahmen nannte ein Teilnehmer die polizeiliche Fahndung, während ein anderer gar keinen verhältnismässigen Einsatz sah, sondern immer nur eine Verletzung verschiedener Grundrechte. Die Mehrheit der Teilnehmenden unterstützte das Argument und bewertete es ebenfalls als inakzeptabel, dass Grundrechtseingriffe vorgenommen werden.

In der französischsprachigen Gruppe wurden keine Vorteile beim Einsatz dieser Technologie gesehen. Diese Aussage wurde dahin gehend relativiert, dass die Technologien zur Suche von verlorengegangenen Personen (z.B. Kindern oder Personen mit einer Alzheimererkrankung) oder auch Verdächtigen nützlich sein könne. Jedoch dürfe dieser Aspekt nicht als Vorwand dienen, eine allgemeine Massenüberwachung einzuführen. Es müsste auch die Freiheit bestehen, gewisse Grenzen zu übertreten; man sollte nicht konstant überwacht werden. Es müsse also klar geregelt werden, wo, wann und wie diese Technologien eingesetzt würden.

Für die Diskussionsteilnehmenden kam die Anwendung am Flughafen oder bei Grenzkontrollen – also in klar abgegrenzten Orten und Anwendungen – infrage. Mit der entsprechenden Automatisierung könnte der Prozess beschleunigt und so mehr Komfort für die Reisenden geschaffen werden. Eine Teilnehmerin fand, dass der Einsatz dieser Technologie an Bahnhöfen möglicherweise das Sicherheitsgefühl der Bürger erhöhen könnte. Ein anderer sah in der Videoüberwachung den Vorteil, dass man weniger Angst vor Diebstahl haben müsste. Jedoch waren sich die Teilnehmenden einig, dass diese Art des Sicherheitsgefühls nicht wirklich nötig sei. Ein weiteres mögliches Einsatzgebiet wurde für den Stadioneingang gesehen, um Hooligans den Einlass zu verwehren.

4.2.2.4. Empfehlungen

In Momenten der Krise, etwa nach einer Terrorattacke, könnte das Einführen dieser Technologie verlockend wirken; es wäre jedoch sehr wichtig, sie nicht unter dem Einfluss von Angst zu implementieren. Auch bei bestehendem Einsatz (Grenzkontrollen) sollte überlegt werden, ob dieser im Einzelfall tatsächlich notwendig ist oder ob er aus purer Gewohnheit erfolge. Die Teilnehmenden waren sich einig, dass in der Gesellschaft gewisse Freiheiten bestehen bleiben müssten und dafür ein gewisses Risiko zu akzeptieren wäre. Entsprechend solle kein flächendeckender Technologieeinsatz erfolgen, solange der konkrete Nutzen der Technologie nicht deutlich geworden ist und keine ausreichende gesellschaftliche Debatte hierüber stattgefunden hat. Befürchtet wurde etwa, dass Technologiebefürworter aus Politik und polizeilichen Stellen von einer Technologieeuphorie angetrieben würden, wodurch das tatsächliche Potenzial überschätzt werde. Daher solle im Rahmen der geforderten gesellschaftlichen Debatte genau diskutiert werden, welche Technik für welchen Zweck eingesetzt werden soll und wie gut sie zur Lösung des jeweiligen Problems geeignet ist.

Insb. in der französischsprachigen Diskussionsgruppe gab es weitgehende Zustimmung hins. des möglichen Technologieeinsatzes im Einzelfall. Für jeden Einsatz sollte aber vorgängig eine entsprechende Entscheidung gefällt werden. Dafür benötigte es einen gesetz-

lich klar definierten Rahmen der erlaubten Anwendungen. Der Einsatz der Technologie durch die Polizei müsste durch eine unabhängige Stelle kontrolliert werden, etwa durch eine unabhängige Kommission des Parlaments wie beim NDB. Der Polizeibericht allein reiche nicht. Nur so könnte der neutrale Blick sichergestellt werden. Die Möglichkeit einer Kontrolle der Massnahmen durch die Bürger war allen Teilnehmenden sehr wichtig. Nebst einer unabhängigen Überwachungsstelle wurde auch dazu angeregt, für den Einsatz der Technologie auch andere Berufsgruppen mit einzubeziehen. So könnten z.B. Sozialarbeiter bei der Suche nach vermissten Personen beigezogen werden, um einer allfälligen Missbrauchsgefahr entgegenzuwirken.

Die Nutzer der Technologien müssten entsprechend ausgebildet werden. Die Polizisten müssten sich darüber im Klaren sein, dass Gesichtserkennung keinesfalls perfekt funktioniert und dass auch bei einem positiven Treffer noch gewisse Unsicherheiten bestehen könnten.

Ein weiterer bedeutender Punkt in der Diskussion war die Transparenz. Es müsste jederzeit klar sein, was mit den erhobenen Daten gemacht werde. Die Datenbank, welche für den Abgleich der Gesichtsbilder erstellt würde, müsste gesetzeskonform erstellt werden. Die Datensicherheit hätte angemessen ausgestaltet zu sein und die Daten sollten auf einem Server in Europa gespeichert werden.

Technische Zugangsgrenzen wären vorzuziehen, da diese einfacher zu kontrollierten sind. So sollte das Prinzip Privacy by Design implementiert werden. Eine spätere Verwendung der Daten sollte technisch ausgeschlossen werden (etwa mithilfe *hash function*).

4.2.3. Authentifizierung via Stimme

Tabelle 10: Zusammenstellung der Fokusgruppe zur «Authentifizierung via Stimme»

Thema	Sprache	Teilnehmende
Authentifizierung via Stimme	Deutsch	5

4.2.3.1. Kurzzusammenfassung

Eingangs führte das Projektteam in das Thema am Beispiel der Stimmauthentifizierung bei Banken ein. Insgesamt standen die Mitglieder der Fokusgruppe dem biometrischen Authentifizierungsverfahren kritisch gegenüber: Alle fünf Teilnehmenden äusserten zum einen Misstrauen in Bezug auf die Verlässlichkeit des technologischen Systems und zum anderen in Bezug auf die Kompetenzen der Organisationen, die die Technik einsetzen und personenbezogene Daten bearbeiten. Vereinzelt sprachen die Teilnehmenden grösseres Vertrauen gegenüber staatlichen Stellen oder im medizinischen Einsatzkontext aus. Grosse Befürchtungen wurden in Bezug auf die Datensicherheit und der recht- und zweckmässigen Verwendung der Daten artikuliert. Nur ein Teilnehmer der Gruppe hatte bereits selbst Erfahrungen mit der Technik gemacht, weshalb nach dem Fachinput der Moderation erst einmal

bekannte Einsatzbereiche besprochen wurden. Bereits bekannt waren den Teilnehmenden Anwendungsszenarien im Bereich von Zugangs- oder Zugriffskontrollen sowie beim Kundenservice, Onlinebanking oder Vertragsabschlüssen, z.B. im Bereich Mobilfunk. Insgesamt schien es jedoch unklar, was das Authentifizierungsverfahren via Stimme genau leisten soll, welche Sicherheitsanforderungen bestehen, welche Anforderungen an die Systeme gestellt werden müssten und welchen Mehrwert die Authentifizierung für Nutzende bringen kann.

Vor diesem Hintergrund taten sich die Teilnehmenden der Fokusgruppensitzung schwer damit, konkrete Empfehlungen für den Umgang mit der Technologie abzugeben. Vielmehr wurden grundlegende Fragen in Bezug auf den Einsatz der Technologie aufgeworfen, die es vor einem Einsatz zu klären gelte.

4.2.3.2. Vorteile

Ein Teilnehmer erklärte, dass Stimmbiometrie und Stimmprofile aus seiner Sicht eine sicherere Authentifikationsmethode als bspw. ein Passwort darstellen, da der Biomarker «Stimme» deutlich individueller sei als eine Kombination aus Zahlen und Worten. Eine Teilnehmerin widersprach dem und erwiderte, dass auch Stimmprofile veränderbar seien, sodass die Möglichkeit für Missbrauch bestünde. Sie sah die Gefahr darin, dass eine Stimmaufnahme missbräuchlich und mit dem Ziel, Zugriff zu einem Konto oder Profil zu erlangen, verwendet werden könnte. Insgesamt vertrat die Gruppe mehrheitlich die Ansicht, dass eine Authentifizierungsmethode via Biomarker «sicherer» als herkömmliche Methoden, wie PINs oder Sicherheitsfragen, sein könnte. Die Gruppe diskutierte zudem, dass derartige biometrische Authentifizierungssysteme aus Kundinnen- und Kundensicht komfortabler erscheinen, als Passwörter zu generieren oder Sicherheitsfragen für eine Zugangsberechtigung zu beantworten.

4.2.3.3. Nachteile

Über die Hälfte der Gesprächszeit wurde darauf verwendet, mögliche Risiken durch eine missbräuchliche Verwendung (z.B. Identitätsmissbrauch) der Technik zu diskutieren. Die Teilnehmenden kamen zu dem Schluss, dass ähnlich wie bei der polizeilichen Gesichtserkennung auch diese Anwendung nur nach vorheriger Analyse der zu schützenden Aspekte wie Privatheit und Datensouveränität (und deren Gefährdung) eingesetzt werden sollte. Bei biometrischen Verfahren, wie der Stimmauthentifizierung, müsse sichergestellt werden, dass biometrische Merkmale nicht ohne Kenntnis der Betroffenen erfasst werden. Die Fokusgruppe sprach sich dafür aus, dass die Technik nicht als ausschliessliche Authentifizierungsmöglichkeit verwendet werden sollte und stattdessen stets Alternativen zur Verfügung stehen sollten. Auch vertrat die Mehrheit der Teilnehmenden die Ansicht, dass die Systemsicherheit und der Datenschutz für die Nutzenden bedeutsamer seien als ein potenzieller Komfort oder eine Zeitersparnis bei der Authentifizierung. Insgesamt erschien allen Teilnehmenden unklar, welches Kundensegment angesprochen werde und welchen genauen Nutzen der Anwendungsfall «Authentifizierung via Stimme» bringen solle.

Ein grosses Risiko der Stimmerkennung sahen die Teilnehmenden der Fokusgruppe darin, dass die Stimme aufgrund von Tagesform, Hintergrundgeräuschen etc. nicht erkannt werden könnte. Es bestanden zudem grosse Bedenken im Hinblick auf einen unberechtigten Zugang, den sich Dritte durch bestehende «Tonschnipsel» oder Mitschnitte verschaffen könnten. Fragen der Datensicherheit und des Datenschutzes wurden nicht nur in Bezug auf die Anwendung, sondern auch in Bezug auf die Datenbearbeitungsprozesse – aufseiten der Anbieter (Bank, Dienstleister) – diskutiert. Den Teilnehmenden war hierbei unklar, inwieweit die Datenbearbeitung seitens der Dienstleister (Bank beim Onlinebanking) vorgenommen oder an Dritte fremdvergeben wird.

4.2.3.4. Empfehlungen

Die Teilnehmenden waren der Ansicht, dass beim Einsatz der Technologie das Risiko, die Verantwortungszuschreibung und das damit verbundene Aufkommen bei Schadensersatz aufseiten der Dienstleister (bzw. Akteuren/Unternehmen, die die Technologie einsetzen) und nicht aufseiten der Kundinnen und Kunden liegen müssten. Die technische Zuverlässigkeit und Datensicherheit der Stimmauthentifizierung müsse überprüft werden und gewährleistet sein. Wichtig war den Teilnehmenden auch, dass Transparenz darüber hergestellt wird, ob die Daten vom Unternehmen bearbeitet werden, das die Technologie einsetzt oder ob sie an Dritte weitergereicht werden. Ebenso wurde die Bedeutung der Gewährleistung einer Widerspruchsmöglichkeit und von alternativen Authentifizierungsmethoden angesprochen.

4.2.4. Gewaltprävention und -aufklärung in Sportstadien

Tabelle 11: Zusammenstellung der Fokusgruppe zur «Gewaltprävention in Sportstadien»

Thema	Sprache	Teilnehmende
Stadionüberwachung	Deutsch	5

4.2.4.1. Kurzzusammenfassung

Diskutiert wurde in der Fokusgruppe über zwei Einsatzszenarien für die automatisierte Gesichtserkennung im Stadion: zum einen das Erkennen von bereits bekannten straffällig gewordenen Personen, die durch die Gesichtserkennung im Stadion identifiziert werden sollen. Bereits als Gefährder (Hooligans) bekannte Personen würden so mittels Gesichtserkennung am Eingang erkannt und präventiv am Zutritt gehindert. Zum anderen könnte die Technik dafür eingesetzt werden, im Stadion stattgefundene Gewalttaten aufzuklären und eine Zuordnung und/oder Identifikation der in einem Delikt involvierten und straffällig gewordenen Personen vorzunehmen (Aufklärung).

Zunächst diskutierten die sechs Teilnehmenden ausführlich über die technischen Rahmenbedingungen zur Stadionüberwachung. Besonders viele Rückfragen der Teilnehmenden er-

gaben sich in Bezug auf die technischen Grundlagen, den Einsatz und die Anwendung der automatisierten Gesichtserkennung.

Im Zentrum der Diskussion standen immer wieder folgende übergeordnete Fragestellungen: Was kann die Technologie überhaupt? Und daran anknüpfend: Wofür genau soll sie eingesetzt werden? Der Tenor der Diskussion war, dass die Gesichtserkennung im Stadion eine unverhältnismässige Massnahme sei, für die es keine ausreichend transparenten (gesetzlichen) Grundlagen und Vorgaben zum Schutz der Privatheit und konkrete Vorteile für die öffentliche Sicherheit gibt.

4.2.4.2. Nachteile

Eine Teilnehmerin plädierte dafür, grundsätzlich nicht darauf zu blicken, was die Technologie kann, sondern die Problemlösung als Leit- und Zielbild für den Technologieeinsatz zu nehmen. Demnach handle es sich bei Hooligans sowie Straftäterinnen und Straftätern in der Fussballszene um ein Extrembeispiel und eine Minderheit, die mithilfe der Technik überwacht werden solle. Da der Stadionbesuch jedoch in den Bereich einer privaten Freizeitunternehmung vieler Menschen zähle, solle nicht ohne Weiteres ein Grundrechtseingriff vorgenommen werden. Dieses Argument erhielt grossen Zuspruch. Hierbei wurde insb. betont, dass Freiheit auch ein Gefühl sei und Menschen nicht ständig und überall einer Überwachung ausgesetzt sein wollten.

Der anlasslose Einsatz von Videoüberwachung mit derartigem Eingriff in die Persönlichkeitsrechte stünde – laut mehrheitlicher Meinung der Teilnehmenden – ausser Verhältnis zum Zweck. Weil es sich bei Stadiongewalttätern um eine sehr kleine Minderheit handle, sei die Erfassung der Gesichter aller Besucher ungerechtfertigt. In diesem Zusammenhang wurde auch über die hohe Fehlerquote des Systems gesprochen, die den Einsatz der Technik noch stärker infrage stellte. Die Teilnehmenden diskutierten, dass es eine hohe Fehleranfälligkeit bei der Zuordnung von Personen gebe, da Lichtverhältnisse, Blickwinkel und Bewegung variieren und einen Einfluss auf die Identifizierbarkeit hätten. Auch das Geschlecht, die Hautfarbe und das Alter hätten einen Einfluss auf die Erkennungsleistung des Systems. Ferner waren die Teilnehmenden der Meinung, dass die Gefährdung durch gewaltbereite Personen in Sportstadien in der öffentlichen Wahrnehmung überschätzt würde. Folglich wurde auch der Nutzen des Technologieeinsatzes als eher gering angesehen. Die Teilnehmenden (die recht gut über Gesichtserkennung informiert waren) nahmen allerdings zugleich an, dass ein grosser Teil der Bevölkerung den Einsatz von Gesichtserkennung in Sportstadien akzeptabel fände, weil zu wenig Bewusstsein über die Risiken vorhanden sei.

Als problematisch erachteten die Teilnehmenden zudem, dass die Hoheit der Einlasskontrollen und damit auch die zur automatisierten Gesichtserkennung in den Händen von privaten Unternehmen liegen würde. So sei zu befürchten, dass die Weiterbearbeitung und der Umgang mit den Daten nicht nachvollziehbar sei. Als weiterer kritischer Punkt wurde in diesem Zusammenhang die Datensicherheit der durch die Software der automatisierten Gesichtserkennung generierten Informationen diskutiert. Eine Teilnehmerin äusserte, dass sie sich unwohl fühlt, wenn private Betreiber Datenbanken mit sensiblen Daten besässen. Die wenigsten privaten Betreiber hätten ihrer Meinung nach Aspekte der Datensicherheit,

Hacking und des Datenschutzes «im Griff». In diesem Zusammenhang wurde auch infrage gestellt, ob und inwiefern der Einsatz solcher Systeme den vielen Betroffenen transparent gemacht wird. Einige Teilnehmende erklärten, dass sie mehr Vertrauen hätten, wenn die Daten in staatlichen Händen liegen bzw. der Einsatz staatlich reguliert würde. Ein Teilnehmer ergänzte, dass nicht klar sei, inwieweit die Daten miteinander verknüpft und mit welchen Stellen geteilt werden. Es wurde insb. befürchtet, dass heute unproblematische Daten auf schleichende Weise in wenigen Jahren problematisch werden (z.B. im Zuge eines politischen Machtwechsels) und somit die bei einem Stadionbesuch erhobenen Daten später möglicherweise negativ auf z.B. Arbeitnehmer zurückfallen könnten. Diesem befürchteten Kontrollverlust stimmten auch die übrigen Teilnehmenden zu und nahmen es als ernste Gefahr wahr, dass Betroffene den Überblick verlieren, welche privaten Anbieter welche Daten über sie sammeln, bearbeiteten und womöglich Profile erstellen und weiterverkaufen (Sekundärnutzung). Die daraus potenziell resultierenden Schäden wurden als unabsehbares Risiko angesehen, das bereits heute adressiert werden sollte. Eine Teilnehmerin äusserte zudem die Befürchtung, dass es Ausweichbewegungen geben werde, sobald der Einsatz automatisierter Gesichtserkennung und die Standorte der Überwachungskameras bekannt würden, sodass an Gewalttaten Interessierte andere Wege finden würden, Straftaten im Stadion oder ausserhalb zu verüben.

4.2.4.3. Vorteile

In Bezug auf mögliche Vorteile der automatisierten Gesichtserkennung im Sportstadion diskutierte die Gruppe, dass die Technologie das Aufspüren von Gewalttätern und die Nachvollziehbarkeit von Delikten im Stadion seitens der Behörden erleichtern könnte. Auch wurde angerissen, dass die Kenntnis vom Vorhandensein eines Gesichtserkennungssystems eine abschreckende Wirkung erzielen und somit präventiv gegen Gewaltmassnahmen wirken könnte.

4.2.4.4. Empfehlungen

Die Teilnehmenden waren sich die Diskussionszeit über einig, dass der massive Eingriff in die Persönlichkeitsrechte der Stadionbesucher keiner Verhältnismässigkeit entspräche. Zum einen wurde keine dringende Notwendigkeit zur Überwachung aller Stadionbesucher zu Präventions- und Aufklärungszwecken von Straftaten mittels Gesichtserkennung gesehen. Zum anderen waren die Teilnehmenden der Ansicht, dass Straftäter auch mittels herkömmlicher Massnahmen ausreichend gut identifiziert werden könnten. Hierbei brachte ein Teilnehmer ein, ob es nicht reichen würde, eine Art «Fahndungsliste» bei der Einlasskontrolle – den Namen auf dem Ticket mit Angabe auf polizeiliche Referenzliste – abzugleichen. Eine andere Idee, die diskutiert wurde, war, dass die bekannten Gewalttäter/Hooligans bspw. mit einem Fingerprint im System hinterlegt sind und so ein Abgleich bei der Einlasskontrolle mit polizeilichen Referenzdatenbanken möglich wird. Dies würde zum einen dazu führen, dass nicht alle Stadionbesucher erfasst werden und zum anderen nur ein Biomarker und nicht die gesamten biometrischen Daten und Bewegungsprofile von Dritten in einem intransparenten System erfasst werden. Darüber hinaus stellte sich ein Teilnehmer die Frage,

ob es nicht darum gehen müsse, das Problem (Warum begehen Menschen Straftaten im Stadion?) mittels Aufklärungskampagnen und Deeskalationsmassnahmen anzugehen und nicht die Symptomatik lösen zu wollen. Über die konkrete Ausgestaltung solcher Massnahmen zur Deeskalation von Gewalt in Sportstadien wurde jedoch nicht weiter diskutiert.

Ferner wurde über die Rolle des Staates bei der Datensammlung und Datenbearbeitung gesprochen. Einig waren sich die Teilnehmenden dabei, dass, wenn die Technologie zum Einsatz kommen sollte, diese nicht aus dem Ausland eingekauft und verwendet werden darf. Hier wurde insb. ein Verlust der persönlichen Datensouveränität befürchtet. Vielmehr müssten von staatlicher Seite verbindliche Standards erarbeitet werden, denen die Technologie entsprechen müsse, bevor sie eingesetzt werden darf. Zudem müsste bei der Datenbearbeitung darauf geachtet werden, dass die Hoheit bei staatlicher Stelle liegt. Ein Zugriff auf diese Daten wäre dann nur staatlichen Sicherheits- und Ordnungsbehörden gestattet.

Für den Fall, dass automatisierte Gesichtserkennung in Sportstadien doch zum Einsatz kommt, formulierten die Teilnehmenden eine Reihe von Vorschlägen: So müsste die Abgrenzung zwischen staatlicher und privatwirtschaftlicher Datenhoheit vorgenommen und die Datenbearbeitung unter staatliche Aufsicht/Hoheit gestellt werden, insb. durch konkrete Gesetze zum Schutz der Betroffenen. Die Definition des Zweckes einer solchen automatisierten Gesichtserkennung müsste sehr genau formuliert sein, um Zweckentfremdungen wirksam auszuschliessen. Und es müsste eine Verpflichtung zur transparenten Information über den Einsatz, z.B. durch Hinweisschilder, die Infos über Zweck, Treffergenauigkeit, Speicherdauer und -ort der Daten, enthalten.

4.2.5. Erkennung physischer Krankheiten

Tabelle 12: Zusammenstellung der Fokusgruppe zur «Erkennung physischer Krankheiten»

Thema	Sprache	Teilnehmende
Erkennung physischer Krankheiten	Französisch	5

4.2.5.1. Kurzzusammenfassung

Die Diskussionsgruppe war sich einig, dass ein Tool zur Unterstützung von Ärzten sinnvoll sein könnte. Wichtig erschien ihr, dass es nur unterstützend zu einer ärztlichen Diagnose verwendet wird und dass die Daten nicht an Dritte weitergegeben werden sollten.

Bezüglich einer in der Bevölkerung frei nutzbaren Diagnose-App war die Gruppe deutlich kritischer. Hier erschien das Risiko einer falschen Anwendung und einer falschen Diagnose als hoch, weil befürchtet wurde, dass medizinischen Laien Fehler bei der Nutzung unterlaufen könnten. Um Fehlinterpretationen vorzubeugen, wurde das korrekte Funktionieren einer solchen App, die Transparenz der Datenbearbeitung sowie die mögliche Weitergabe der Daten an Dritte als essenzielle Faktoren angesehen, die gewährleistet sein müssten. Zudem wurde ein obligatorisches Zertifizierungsverfahren für diese Art von Apps gefordert.

4.2.5.2. Vorteile

Die Gesprächsgruppe war sich einig, dass der Einsatz dieser diagnostischen Anwendung ein unterstützendes Tool für Ärzte sein könnte, dies jedoch nur unter der Voraussetzung, dass das Tool die Diagnose nicht abschliessend bestimmt. Ein Teilnehmer merkte an, dass derartige Anwendungen es wahrscheinlich erlauben würden, viel mehr Daten miteinander in Verbindung zu bringen, als eine einzelne Person es je vermögen würde. Eine Teilnehmerin relativierte dies dahin gehend, dass der Arzt den Vorteil habe, den Gesamtkontext und auch den Patienten persönlich zu kennen. Einig waren sich die Teilnehmenden zudem dahin gehend, dass unterstützende Geräte aufgrund des Ärztemangels immer wichtiger werden könnten.

Bei der App-Variante wurde für positiv befunden, dass so der Zugang zu einer medizinischen Abklärung erleichtert werden könnte, wenn Betroffene möglicherweise nach einem entsprechenden Resultat einer App eher einen Arzt aufsuchen.

4.2.5.3. Nachteile

Ein Teilnehmer empfand die elektronische Erfassung von medizinischen Daten als grundsätzlich beunruhigend. So könnten insb. Krankenkassen versuchen, vermehrt zusätzliche Daten zu sammeln und miteinander zusammenzuführen (z.B. zusätzlich aus Smart Watches). Darauf basierend könnten im Bereich der Privatversicherungen Prämien angepasst werden, was diskriminierend wäre. Das Zusammenführen unterschiedlicher Daten könne jedoch dann vorteilhaft sein, wenn das elektronische Patientendossier nur beim Arzt verbliebe.

Auch bei einer App betrachteten die Gesprächsteilnehmenden die Gefahr der Datenweitergabe an Dritte und eine mangelnde Transparenz der Datenbearbeitung als problematisch.

Besonders betont wurde auch die Gefahr von Falschdiagnosen durch die App, die falsche Sicherheit oder unnötige Besorgnis hervorrufen könnten. Die falsch-positive Diagnose erschien den Teilnehmenden weniger belastend als eine falsch-negative Diagnose. Hier könnte wichtige Zeit zur Behandlung einer Erkrankung verloren gehen, da sich Betroffene in falscher Sicherheit wiegten. Im Besonderen wurde hierbei die Gefahr angesprochen, dass das Auffinden seltener Krankheiten möglicherweise schwieriger sein könnte, weil es aufgrund der Seltenheit schwieriger sei, eine ausreichende Datenbasis aufzubauen.

Die Teilnehmenden gaben an, solchen Apps dann eher zu vertrauen, wenn sie von gemeinwohlorientierten Forschungseinrichtungen stammen als von gewinnorientierten Unternehmen. Weil zugleich erwartet wurde, dass künftig eher privatwirtschaftliche Anbieter den Markt kontrollieren würden, äusserten die Teilnehmenden Befürchtungen hins. des Umgangs mit ihren sensiblen Gesundheitsdaten.

4.2.5.4. Empfehlungen

Den Diskutanten war es auch bei dieser Anwendung wichtig, Transparenz herzustellen und die Datenbearbeitung klar zu regeln.

Bei einer Erhebung der Daten durch den Arzt und der Erstellung eines Dossiers müsste absolut transparent sein, was mit den Daten genau passiert. Die Funktionsweise des Diagnosetools sollte den Patienten zudem verständlich erklärt werden.

Ärzte sollten miteinander eine Konvention erstellen, wie und wann diese Technologien eingesetzt werden. Auch sollten Ärzte bezüglich der Anwendung der Technologien extra geschult werden. Zudem müsste eine Qualitätskontrolle dieser Systeme eingeführt werden. Ebenfalls müsste ausreichend geforscht und experimentiert werden. Bei der Entwicklung der Systeme sollte medizinisches Personal eng mit eingebunden werden. Es wäre auch sicherzustellen, dass der Einsatz dieser Diagnosetools die Interaktion zwischen Arzt und Patient nicht behindert. Auch sollten diese Anwendungen nur unterstützend und nie als ausschlaggebendes Indiz für eine Diagnose verwendet werden.

Vor der Bereitstellung einer Diagnose-App müsste ein Verfahren durchlaufen werden. In diesem müsste die Qualität der Algorithmen betreffend medizinische Aspekte und die weitere Verwendung der medizinischen Daten überprüft werden. Die Teilnehmenden befanden das Erstellen und Erwerben eines speziellen Zertifikats für medizinische Apps als wünschenswert. Die Nutzer sollten zudem immer und klar über den Zweck und die Anwendung informiert werden. Die Daten sollten möglichst nur auf dem Gerät gespeichert und nicht auf einen Server übertragen werden.

Es sei in diesem Bereich keine nationale Strategie zu verfolgen. Es erfordere vielmehr eine weltweite Strategie, um grosse Unternehmen, die diese Diagnosetools erstellen, zu kontrollieren.

4.2.6. Erkennung psychischer Krankheiten

Tabelle 13: Zusammenstellung der Fokusgruppe zur «Erkennung psychischer Krankheiten»

Thema	Sprache	Teilnehmende
Erkennung psychischer Krankheiten	Deutsch	5

4.2.6.1. Kurzzusammenfassung

In der Diskussionsgruppe bestand Einigkeit darüber, dass es nach einer grundsätzlichen gesellschaftlichen Debatte über den Umgang, die Nutzung und die Auswirkungen des Einsatzes derartiger neuer Medizintechnologien bedarf.

Grundlegende Fragen wurden hingegen zur wissenschaftlichen Aussagekraft der Ergebnisse bzw. Diagnosen aufgeworfen. Besonderen Raum nahmen dabei normative Implika-

tionen wie «das Recht auf Nichtwissen» ein. Grundsätzlich unterliegen Ärzte einer Aufklärungs- und Informationspflicht gegenüber ihren Patientinnen und Patienten. Sollte bspw. die Technik vor Ausbruch einer Erkrankung eingesetzt werden und sollten für die Person keine Optionen zur Therapie mehr bestehen, böte das Recht auf Nichtwissen Schutz, um die Belastung zu mindern und die Lebensqualität (bis zum Ausbruch der Erkrankung) zu wahren. Diskutiert wurde zudem, dass unterschieden werden müsse, ob die Analyse im Rahmen einer ärztlich durchgeführten Diagnostik oder als Selbstdiagnostik anhand von Apps von Privatpersonen durchgeführt wird. Letztlich sprach sich die Fokusgruppe mehrheitlich für die Verwendung der Software im Rahmen einer ärztlichen Diagnostik aus.

4.2.6.2. Vorteile

Die Teilnehmenden sahen einen grossen Nutzen von neuen Technologien im Gesundheitsbereich, besonders in Bezug auf eine personalisierte Diagnose und die Früherkennung von Krankheiten. Daher wurde die Analyse von Stimme, Sprach- und Gesichtsausdrücken zum Zwecke der Erkennung und Behandlung von psychischen Krankheiten ebenfalls prinzipiell befürwortet. Eine Teilnehmerin führte dazu ein Beispiel aus dem persönlichen Umfeld an, bei dem eine Autismus-Diagnose zu spät erfolgte: Hätten derartige Technologien tatsächlich eine frühzeitige Autismus-Diagnose oder zumindest Hinweise (die noch durch Ärzte bestätigt werden müssen) ermöglicht, hätten Therapie und Betreuung viel früher einsetzen und die Lebensqualität somit merklich gebessert werden können. Auch die übrigen Teilnehmenden waren überzeugt von den Vorteilen einer Früherkennung von psychischen Krankheiten.

Darüber hinaus vertrat eine Teilnehmerin die Meinung, dass das Vertrauen in die Gesundheitsversorgung sowie in Ärztinnen und Ärzte gesunken sei und die Selbstdiagnose eine Möglichkeit darstelle, sich selbst um die eigene Gesundheit zu kümmern. Auch wurde diskutiert, dass solche Technologien für Minderheiten, die aus verschiedenen Gründen nicht regelmässig zum Arzt gehen, den Zugang zur Gesundheitsversorgung verbessern könnten.

4.2.6.3. Nachteile

Die Teilnehmenden befürchteten, dass Aussagen über (insb. mehrdimensionale) psychische Krankheitsbilder nicht ohne Weiteres von Laien interpretiert werden könnten. In Verknüpfung mit dem Gedanken, dass das Vertrauen in Ärzte gesunken sei, wurde befürchtet, dass auf diese Weise gefährliche Fehldiagnosen zustande kommen könnten. In diesem Zusammenhang wurde auch diskutiert, dass Faktoren, wie bspw. Hardwarevoraussetzungen (älteres Mobiltelefon, defektes Mikrofon) sowie die subjektive Tagesform, das Ergebnis einer solchen Diagnostik beeinflussen könnten, ohne dass sich Nutzende darüber im Klaren wären.

Eine Teilnehmerin brachte ein, dass neugierige Menschen die App auch ohne Anlass oder Symptomatik ausprobieren würden und dann mit (Fehl-)Diagnosen umgehen müssten. Dies könne eventuell zur Folge haben, dass sich Menschen in ihrem sozialen Verhalten verändern und so ihre Lebensqualität massiv eingeschränkt wird. Die Mehrzahl der Teilnehmenden sprach sich dafür aus, dass solche Apps nicht als Selbstdiagnoseinstrument

zugänglich gemacht werden sollten. Vielmehr sollte vorab die wissenschaftliche Evidenz solcher Anwendungen sichergestellt werden. Der Prozess der Interpretation der Daten und die Vermittlung der Diagnosen durch ärztliches Personal dürfe laut der Mehrzahl der Teilnehmenden nicht unterschätzt werden. Die grosse Mehrheit der Teilnehmenden sprach sich deshalb dafür aus, dass die Software nur in Verbindung mit einer ärztlichen Diagnostik eingesetzt werden sollte.

Darüber hinaus diskutierte die Gruppe, dass die Verwendung solcher Apps (im Falle der Selbstdiagnostik) nur ein weiteres Symptom der Optimierungsgesellschaft sei. Das aktive Bemühen um die eigene Gesundheit zur Erfüllung der eigenen Optimierungserwartungen würde nicht nur zu weiteren Belastungen in Bezug auf das eigene Konkurrenz- und Anspruchsdenken führen. Zusätzlich wurde befürchtet, dass es zu gesellschaftlichen Spaltungstendenzen führen könnte, indem die Chancenungerechtigkeit und soziale Ungleichheit verschärft werden, wenn bspw. Arbeitgebende nur noch besonders gesunde Menschen rekrutieren und auch Krankenkassen ihre Anforderungen für die Aufnahme oder Zahlung von Leistungen erhöhen.

4.2.6.4. Empfehlungen

Die Mehrzahl der Teilnehmenden war der Ansicht, dass stimm-, sprach- und gesichtserkennungsbasierte Software zur Erkennung psychischer Krankheiten als Medizinprodukt eingestuft werden und spezifischen Vorschriften und Standards unterliegen sollte, bevor es auf dem Markt mit einer entsprechenden Kennzeichnung zugänglich wird. Die Zertifizierung müsse aus Perspektive der Teilnehmenden als Voraussetzung dafür gelten, dass Patientinnen und Patienten mit der Software oder dem Medizinprodukt therapiert werden dürfen. Dafür wurde der Staat in der Verantwortung gesehen, entsprechende Vorkehrungen für einen sicheren Einsatz zu treffen. In diesem Zusammenhang wurde auch der Anspruch formuliert, dass digitale Tools, deren Einsatz und Erprobung noch am Anfang stehen, nur ein Hilfsmittel in der Patientenversorgung sein und keine ärztliche Diagnostik ersetzen dürften. Ein weiteres Augenmerk wurde darauf gelegt, dass die Software von Anfang an «datenschutzfreundlich» programmiert und entwickelt werden sollte, damit die besondere Sensibilität von Gesundheitsdaten bereits im Technikdesign Berücksichtigung findet.

Ferner wurde einvernehmlich für essenziell befunden, dass die entsprechenden Anwendungen erklärbare Diagnosen generieren, sodass die Ergebnisse für die Patientin oder den Patienten nachvollziehbar vermittelt werden können. In diesem Zusammenhang diskutierten die Teilnehmenden auch, dass die Robustheit und Zuverlässigkeit der Datenübertragung und der rechtlich konforme Einsatz der Medizintechnik sichergestellt werden muss.

Im Hinblick auf die korrekte Interpretation der Softwareergebnisse waren sich die Teilnehmenden einig, dass es nicht ausreiche, Laien zu informieren. Stattdessen wurde gefordert, dass auch die digitalen Kompetenzen des medizinischen Fachpersonals (auch im Hinblick auf den datenschutzbewussten Umgang mit Patientinnen- und Patientendaten) ausgebaut werden sollten, weil die Interpretation KI-basierter Softwareergebnisse eine neue Herausforderung darstelle.

4.2.7. Emotionserkennung und Aufmerksamkeitsanalyse

Tabelle 14: Zusammenstellung der Fokusgruppe zu «Emotionserkennung und Aufmerksamkeitserkennung»

Thema	Sprache	Teilnehmende
Emotionserkennung/Aufmerksamkeitsanalyse	Französisch	5

4.2.7.1. Kurzzusammenfassung

In der Fokusgruppe wurde sowohl über die Aufmerksamkeitsanalyse in der Schule als auch über mögliche Einsätze der Emotionserkennung bei Bewerbungsgesprächen, am Arbeitsplatz und im Auto diskutiert.

Die Diskussionsgruppe hatte gegenüber jedwedem Einsatz der Emotionserkennung grosse Bedenken. Der Vorteil eines Einsatzes in der Schule wurde als sehr gering gewertet. Der daraus entstehende Druck für die Schüler oder auch für Arbeitnehmende an einem Arbeitsplatz wäre zu gross und zudem fiele das Gefühl des ständigen Überwachtwerdens negativ ins Gewicht. Als sinnvoll angesehen wurde der Einsatz der Aufmerksamkeitsanalyse in Fahrzeugen für das Erkennen einer Übermüdung des Fahrers. Für diesen Einsatz setzte die Gruppe jedoch voraus, dass die Daten nicht an Dritte (z.B. Versicherungen) weitergegeben würden. Insgesamt wurden bei diesem Szenario (aufgrund der möglichen Missbrauchsgefahr und dem geringen Nutzen für den Betroffenen, dessen Gesichtszüge analysiert würden) starke Vorbehalte angebracht. Sollte es sich in bestimmten Situationen als sinnvoll erweisen, müsse der Einsatz sehr klar reglementiert werden.

4.2.7.2. Vorteile

Ein möglicher positiver Anwendungsfall wäre die Aufmerksamkeitsanalyse in einem Flugzeug oder in einem Auto. So könnten Unfälle aufgrund der Übermüdung des Fahrers verhindert werden. Daneben wurde die Durchführung einer ersten Vorauswahl von geeigneten Personen in Bewerbungsgesprächen als eine potenziell sinnvolle Möglichkeit angesehen.

Auch die Möglichkeit einer Unterstützung von Schülerinnen und Schülern durch die Technologie wurde teilweise als positiv bewertet.

Weitere Vorteile der Technologieanwendung wurden als wenig klar angesehen.

4.2.7.3. Nachteil

Die Idee einer Anwendung von Aufmerksamkeitsanalysen in Schulen fand eine Teilnehmerin traurig. Es sei nicht gut, wenn die Schüler konstant überwacht würden. Die Technologie würde zu viel Druck auf die Kinder ausüben, sei zu einschneidend und lasse keine Freiheiten. Zwei der Teilnehmenden waren sich einig, dass sie ihre Kinder nicht auf eine Schule

mit der Anwendung dieser Technologie schicken würden. Sie möchten ja selbst auch nicht konstant überwacht werden. Weiterhin sei es zweifelhaft, ob diese Anwendung im Unterricht überhaupt einen Vorteil darstelle. Eine gute Lehrkraft benötige so ein System wohl kaum. Im Falle einer zu grossen Schüleranzahl pro Lehrperson gäbe es auch andere Lösungen. Ein weiterer Teilnehmer meinte, dass es darauf ankomme, ob er den entsprechenden Lehrkräften vertrauen könnte. Er würde viele Fragen haben zum Einsatz des Systems, bevor er diesem im Klassenzimmer zustimmen würde (wie funktioniert der Algorithmus etc.). Ähnlich äusserte sich die vierte Diskussionspartnerin; ausschlaggebend sei auch, was die Lehrperson mit diesen Informationen machen würde. Es sei z.B. wichtig, dass Emotionen nicht negativ gewertet würden. Allenfalls könnte das System unterstützend eingesetzt werden. Insgesamt waren aber alle Fokusgruppen-Mitglieder bezüglich des Einsatzes dieser Technologie in Schulen ablehnend.

Auch wurde von einem Teil der Teilnehmenden grundsätzlich angezweifelt, ob eine künstliche Intelligenz Emotionen tatsächlich korrekt erkennen könnte. So bestünden auch kulturelle Unterschiede und selbst ein Lächeln könne ja unterschiedlich gedeutet werden. Zudem könnten Emotionen auch verfälscht werden, man könne z.B. weinen, ohne traurig zu sein. Dem fügte eine andere Teilnehmerin zu, dass bei äusserlich nicht sichtbaren Einschränkungen, wie z.B. einer Störung aus dem Autismusspektrum, das Erkennen von Emotionen sehr schwierig sein könne. Hier wäre die Gefahr, dass sich der Algorithmus täuscht, besonders hoch. Würde die Emotionserkennungssoftware etwa in einem Bewerbungsgespräch eingesetzt, könnte dies zu einer Diskriminierung führen.

Ein breiter Einsatz dieser Anwendung in Bewerbungsverfahren könnte dazu führen, dass Schüler und Schülerinnen und Studierende dahin gehend ausgebildet würden, den «richtigen» Gesichtsausdruck zu machen, um in Bewerbungsgesprächen erfolgreich abzuschneiden.

Auch für den Einsatz am Arbeitsplatz wurde die Technologie kritisch gesehen, da Überwachungssysteme zu zusätzlichem Druck führen würden.

Die Diskussionsgruppe besprach ebenfalls die Anforderung an die Transparenz der Datenbearbeitung und die Möglichkeit einer Datenbekanntgabe an Dritte. Ein Teilnehmer fügte an, er habe Bedenken, dass diese Technologie auch eingesetzt werden könnte, um Personen zu manipulieren. Schliesslich könnte das Durchschauen der Emotionen vieles über einen Menschen verraten. Daher müsste man Gesichtsbilder so schützen können, dass sie mithilfe dieser Technologien nicht erkannt werden. Ein weiterer Teilnehmer äusserte sich dahin gehend, dass es ja fraglich sei, ob eine Person, welche die Technologie einsetzt, dies immer im Interesse des Betroffenen tue. Es erschliesse sich ihm nicht, weshalb jemand seine Emotionen lesen sollte und zu welchem Zweck. Schliesslich seien Emotionen höchst persönlich.

So waren alle Diskussionsteilnehmenden auch hins. dieser Anwendung kritisch eingestellt und fanden, dass es v.a. auf den Einsatzbereich der Technologie ankäme. Bei einer Abwägung von positiven und negativen Aspekten dieser Anwendung würden aber die negativen Aspekte klar überwiegen.

4.2.7.4. Empfehlungen

Für den Einsatz dieser Technologie müssten die konkrete Situation und der tatsächliche Bedarf im Voraus analysiert werden. Die Anwendung könne wirklich nur in extremen Situationen gerechtfertigt werden und es bedürfe einer entsprechenden Kontrolle; so das Votum eines Teilnehmers.

Die Teilnehmenden waren sich darin einig, dass die Technologie grundsätzlich nicht durch die Polizei eingesetzt werden sollte, da die Wahrscheinlichkeit eines falsch-positiven Treffers zu hoch sei. Eine mögliche Anwendung wäre aber, die Technologie zu verwenden, um inhaftierte Personen freizusprechen. Allenfalls könnte sie eingesetzt werden, um innerhalb einer Aufklärung ein nächstes Indiz zu erlangen. Keinesfalls dürfte, auf der Technologie basierend, ein endgültiger Entscheid über die (Un-)Schuld einer Person gefällt werden.

Falls die Aufmerksamkeitsanalyse in der Schule eingesetzt würde, müsste sichergestellt werden, dass dies wohlwollend geschähe und dass es nur ein Hilfsmittel darstellte, um den Schülern damit eine individuelle Unterstützung zukommen zu lassen. Auch eine Anwendung der Technologien in Bewerbungsverfahren oder am Arbeitsplatz müsste kritisch überprüft werden.

Für die Entwicklung der Technologie und den Datenabgleich während des Einsatzes benötigte Daten müssten auf gesetzeskonformer Grundlage beschafft werden.

Eine Analyse der Emotionen sei allgemein kritisch zu sehen, da Emotionen Ausdruck der Persönlichkeit seien und damit eine Gefahr der Diskriminierung entsprechend hoch sei.

Die Privatsphäre und das Recht auf Datenschutz seien zu achten und Menschen seien in jedem Fall vor möglicher Manipulation zu schützen.

Wichtig war allen Teilnehmenden, dass die gesetzlichen Schranken, d.h., in welchem Rahmen diese Technologien eingesetzt würden, klar geregelt und auch entsprechend kontrolliert werden müssten.

4.2.8. Jedermann-Identifikation

Tabelle 15: Zusammenstellung der Fokusgruppe zur «Jedermann-Identifikation»

Thema	Sprache	Teilnehmende
Jedermann-Identifikation	Deutsch	4

4.2.8.1. Kurzzusammenfassung

Die Diskussionsgruppe äusserte sich durchweg sehr kritisch zur Jedermann-Identifikation. Nachdem in der ersten Hälfte eine Reihe von Bedenken geäussert wurden, drehte sich der Rest der Diskussion v.a. darum, wie dem als sehr problematisch empfundenen Anwendungsgebiet begegnet werden sollte. Mögliche Vorteile wurden erst geäussert, als seitens der Moderation explizit danach gefragt wurde. Generell taten sich die Teilnehmenden

schwer damit, überhaupt Nutzenpotenziale der Jedermann-Identifikation zu identifizieren. Bei der Diskussion von Handlungsempfehlungen erwies es sich als äusserst schwierig, den unerwünschten Folgen der Jedermann-Identifikation mit einfachen Mitteln zu begegnen. Den meisten Zuspruch erntete schliesslich der Vorschlag, das Tragen von Datenbrillen in der Öffentlichkeit bis auf Weiteres zu verbieten.

4.2.8.2. Nachteile

Eine Person eröffnete die Diskussion mit der Aussage, dass eine Welt mit Jedermann-Identifikation eine Welt wäre, «in der sie nicht leben möchte». Diese Aussage wurde ausnahmslos bejaht und bildete den Tenor der restlichen Diskussion.

Unter die als störend empfundene Aspekte fallen die befürchtete Erschwerung der Durchsetzung des Rechts am eigenen Bild, die vollständige Erosion der persönlichen Anonymität im öffentlichen Raum und die Möglichkeit der Entstehung einer Informationsasymmetrie zwischen denjenigen, die über die Mittel zur Nutzung der Technologie verfügen, und denjenigen, die dieser Form der Gesichtserkennung ausgesetzt werden. Insb. wurde befürchtet, dass dies zu vermehrtem und intensiverem Stalking führen könnte, weil Täter intime Informationen abgreifen und ihren Opfern so einfacher auflauern könnten. Eine Person äusserte Sympathie gegenüber einer durch die Jedermann-Identifikation unterstützte polizeilichen Rasterfahndung. Nachdem auch die Missbrauchsmöglichkeiten eines solchen Zugriffs diskutiert wurden, bewertete auch diese Person die Risiken höher als den Nutzen.

4.2.8.3. Vorteile

Auf die explizite Frage, ob auch positive Aspekte der Jedermann-Identifikation gesehen werden, äusserten die Teilnehmenden, dass es durchaus wünschenswert sein könnte, mit einer Datenbrille gewisse soziale Interaktionen zu vereinfachen. Auf einer Konferenz könnten bspw. fachliche Interessen nebst Namen der Personen eingeblendet werden, um den fachlichen Austausch zu erleichtern. Ebenso könnten Datenbrillen mit Gesichtserkennungsfunktion bei Dating-Veranstaltungen mit Einverständnis der Träger zur Kommunikation von Interesse verwendet werden. Für die Mehrzahl der Teilnehmenden war es aber zugleich klar, dass die wenigen positiven Aspekte in keinem Verhältnis zur Befürchtung der Dauerüberwachung durch Fremde stehen.

4.2.8.4. Empfehlungen

Ungefähr die Hälfte der Zeit diskutierten die Teilnehmenden, wie mit der Jedermann-Identifikation umgegangen werden sollte. Zunächst wurden technische Möglichkeiten diskutiert, darunter die Visualisierung der Aufnahmefunktion smarter Brillen mittels einer hell leuchtenden LED, sodass Betroffene unmittelbar erkennen könnten, dass sie gerade gefilmt werden. Schnell einigten sich die Teilnehmenden, dass Personen, die die Jedermann-Identifikation nutzen, mit hoher Wahrscheinlichkeit auch Möglichkeiten finden würden, einen solchen Warnhinweis zu deaktivieren.

Ein anderer Vorschlag sah vor, dass Betroffene, bspw. über ihre Smartphones, Störsignale aussenden, mit denen die Aufnahmen smarter Brillen gestört werden. Ein vergleichbarer Vorschlag sah vor, dass Betroffene dies auch durch Aufkleber im Gesicht mit Störmustern erreichen könnten. Wie schon beim Lösungsvorschlag zuvor wurde auch bei diesen Vorschlägen befürchtet, dass Personen, die die Jedermann-Identifikation betreiben, mit hoher Wahrscheinlichkeit technische Umgehungsmöglichkeiten finden würden. Bestenfalls wurde eine Art Katz- und Mausspiel zwischen der Entwicklung und Nutzung von Störungsmöglichkeiten und der Entwicklung verbesserter Gesichtserkennungsalgorithmen erwartet. So wurde erwartet, dass derartige Störungsmethoden eine Zeit lang Abhilfe bieten könnten, aber mit verbesserten Algorithmen, die den Umgang mit den Störmethoden erlernen, zunehmend nutzlos würden.

Auf der grundsätzlichen Ebene wurde an diesen Möglichkeiten kritisiert, dass sie die Schutzverantwortung auf die Betroffenen verlagern. Die Mehrzahl der Teilnehmenden war sich einig, dass eine angemessene Lösung ohne das Zutun der Betroffenen auskommen müsste.

Nachdem sich die Teilnehmenden darauf einigten, dass die vorgenannten Empfehlungen weitgehend wirkungslos und ohnehin nicht wünschenswert wären, wurde über zwei weitere Möglichkeiten des Umgangs mit der Jedermann-Identifikation diskutiert. Zunächst wurde darüber diskutiert, ob es zu einer sozialen Ächtung des Tragens von smarten Brillen in der Öffentlichkeit kommen könnte, sodass potenzielle Träger smarte Brillen aus sozialer Rücksichtnahme bzw. aus Furcht vor den sozialen Konsequenzen keine derartigen Brillen in der Öffentlichkeit tragen würden. Und falls doch, könnte eine Person, die sich davon gestört fühlt, dass jemand eine smarte Brille trägt, unter Verweis auf die geltende soziale Norm diese Person dazu auffordern, die Brille abzunehmen. Als zusätzlich vorteilhaft bei der sozialen Ächtung wurde die demokratische Komponente gesehen, da auf diese Weise ein regulatorischer Eingriff des Staates vermieden werden könnte und die Gesellschaft selbst die zu befolgenden Normen festlegen könnte. Seitens einer Person wurde eingewendet, dass eher zu erwarten sei, dass smarte Brillen sich bei zunehmend vielen Menschen durchsetzen werden und es daher voraussichtlich nicht zu einer sozialen Ächtung kommen werde. Eingewendet wurde aber auch, dass dieser Vorschlag seine Wirkmächtigkeit dann verlieren werde, sobald smarte Brillen durch die weitere Miniaturisierung der smarten Elemente irgendwann von herkömmlichen Brillen nicht mehr unterscheidbar werden.

Der weitreichendste, aber auch mehrheitlich als besonders sinnvoll erachtete Vorschlag war ein staatliches Verbot, smarte Brillen in der Öffentlichkeit zu tragen. Eine Person empfand dies als zu weitgehend und verwies darauf, dass dann auch der private Flugverkehr mit Drohnen verboten werden müsste. Zudem wurde auch bei diesem Vorschlag die grundsätzlich erschwerte Kontrollierbarkeit zu bedenken gegeben, die aus der zu erwartenden Ununterscheidbarkeit smarter von herkömmlichen Brillen resultiert. Diesem Argument wurde jedoch entgegengesetzt, dass bereits die Drohkulisse einer staatlichen Strafe eine abschreckende Wirkung auf die grosse Mehrzahl Trägerinnen smarter Brillen entfalten würde. Kritisch diskutiert wurde auch, dass die Teilnehmenden keinen nennenswerten Nutzen der Jedermann-Identifikation erkennen konnten und deshalb stark für ein Verbot eintraten. Daher wurde angemerkt, dass die Entstehung von Nutzungsweisen, die als nützlich wahrgenommen werden, dazu führen könnte, dass sich der Zuspruch für ein Verbot verringert.

4.3. Zusammenfassung der Ergebnisse und Schlussfolgerungen

Gemäss der von der Projektgruppe vorbereiteten Struktur unterteilten sich alle Fokusgruppen-Diskussionen in einen ersten Block, in dem Chancen bzw. Hoffnungen und Risiken bzw. Ängste in Erfahrung gebracht wurden, sowie in einen zweiten Block, in dem die Teilnehmenden nach ihren Wünschen für einen Umgang mit den zuvor identifizierten Risiken der Stimm-, Sprach- und Gesichtserkennungstechnologien in den jeweiligen Anwendungsgebieten gefragt wurden.

Insgesamt zeigte sich, dass bei allen betrachteten Anwendungsfeldern stets (und teils erheblich) mehr Nachteile als Vorteile gesehen wurden. Sowohl bei den Vorteilen als auch bei den Nachteilen gab es eine Reihe grundlegender Argumente. Im Folgenden werden nun die zentralen Argumente im Bereich der erkannten Vor- und Nachteile zusammengetragen. Im Anschluss werden die über mehrere Diskussionen hinweg genannten Empfehlungen zusammengeführt und Schlussfolgerungen aus den Fokusgruppen gezogen.

4.3.1. Vorteile aus Sicht der Teilnehmenden

Die Fokusgruppen-Teilnehmenden sahen eine Reihe von Vorteilen der Stimm-, Sprach- und Gesichtserkennungstechnologien. Dazu gehören (in absteigender Häufigkeit der Nennungen, vgl. auch Tabelle 21 im Anhang):

- Steigerung der (gefühlten) persönlichen Sicherheit (5)
- Steigerung des Komforts (4)
- Verbesserte Aufklärung bei Ermittlungen (3)
- Verbesserte (Früh-)Erkennung von Krankheiten (2)

Die **Steigerung der (gefühlten) persönlichen Sicherheit** wurde in der Diskussion zu smarten Lautsprechern, zu polizeilicher Überwachung und zur Emotionserkennung bzw. Aufmerksamkeitsanalyse sowie beim Thema Authentifizierung via Stimme als Vorteil genannt. Im Rahmen der Diskussionsrunden über smarte Lautsprecher wurde die Steigerung der persönlichen Sicherheit insofern als Chance wahrgenommen, als smarte Lautsprecher insb. ältere und beeinträchtigte Menschen dabei unterstützen könnten, z.B. nach einem Sturz Hilfe zu rufen oder den Hilferuf mittels automatischer Sturzerkennung zu automatisieren. Im Rahmen der Emotionserkennung bzw. Aufmerksamkeitsanalyse wurde die Steigerung der persönlichen Sicherheit darin gesehen, dass die Technologie im Falle der Übermüdung des Fahrers eine Warnung aussenden und somit Unfälle verhindern könnte. Bei der Authentifizierung via Stimme versprachen sich die Teilnehmenden mehr Sicherheit als bei herkömmlichen Authentifizierungsmethoden, wie PINs und Sicherheitsfragen. Im Hinblick auf den Einsatz von Gesichtserkennungstechnologien durch polizeiliche Stellen wurde diskutiert, dass der Technologieeinsatz in der Öffentlichkeit zu einem subjektiven Gefühl gesteigerter Sicherheit führen könnte.

Die **Steigerung des Komforts** wurde ebenfalls in den Fokusgruppen zu smarten Lautsprechern, in der Diskussion zu polizeilicher Überwachung sowie beim Thema Authentifizierung

via Stimme als Vorteil genannt. Bei den Diskussionen zu smarten Lautsprechern äusserten sich die Teilnehmenden zwar dahin gehend, dass die in den Lautsprechern integrierte Sprachsteuerung häufig bereits auch auf dem Smartphone vorhanden und damit teilweise überflüssig sei. Zugleich befanden sie, dass die Geräte durch die Übernahme kleinerer Tätigkeiten (Wechseln des Fernsehkanals, Smart-Home-Steuerung usw.) zunehmend zur Normalität werden und damit zu einer Komfortsteigerung führen würden. Im Kontext des polizeilichen Einsatzes von Gesichtserkennung wurde auf die Komfortsteigerung im Zusammenhang mit der Beschleunigung von Grenzkontrollen verwiesen. Die Authentifizierung via Stimme erachteten die Teilnehmenden als eine komfortable Alternative zu herkömmlichen Authentifizierungsmethoden, weil dadurch die Nennung von Passwörtern bzw. Beantwortung von Sicherheitsfragen entfalle.

Die **verbesserte Aufklärung bei Ermittlungen** wurde im Rahmen der Diskussion der Stadionüberwachung und im Rahmen beider Fokusgruppen zur polizeilichen Überwachung als Vorteil hervorgehoben. Beim Einsatz in Sportstadien wurde das Aufspüren von Gewalttätern und die Vereinfachung behördlicher Ermittlungstätigkeiten als Vorteil gesehen. Die Teilnehmenden stellten aber zugleich klar, dass sie die Erfassung der Gesichter aller nicht verdächtigen Personen als unverhältnismässig einstuften und sich gegenwärtig trotz der Vorteile gegen den Einsatz von Gesichtserkennung im Stadion aussprechen würden. Auch die Unterstützung polizeilicher Ermittlungstätigkeiten (insb. zur Suche vermisster Personen) wurde als Chance aufgeführt, die jedoch wieder relativiert wurde, weil damit nicht der Einführung einer generellen Massenüberwachung Vorschub geleistet werden dürfe. In der Diskussion über den polizeilichen Einsatz der Gesichtserkennung wurde die Polarisierung der Argumente klar. Während ein Teil des Diskutanten die Nutzung für die polizeilichen Ermittlungstätigkeiten als klaren Vorteil mit gesamtgesellschaftlichem Nutzen befürwortete, war der andere Teil wesentlich skeptischer und befand den Nutzen im Vergleich zu den Nachteilen als nicht ausreichend.

Die **verbesserte (Früh-)Erkennung von Krankheiten** wurde in den Fokusgruppen als eindeutig vorteilhaft bewertet. Die Teilnehmenden verwiesen darauf, dass sowohl die korrekte (Früh)Erkennung einer Krankheit zu begrüssen sei als auch bereits die Ausgabe von Hinweisen, die noch ärztlich überprüft werden müssten.

4.3.2. Nachteile aus Sicht der Teilnehmenden

Die Fokusgruppen-Teilnehmenden erkannten insgesamt weitaus mehr Nachteile bzw. Risiken als Vorteile. Zugleich waren sich die Teilnehmenden bei allen Anwendungsfeldern deutlich häufiger über die identifizierten Nachteile einig. Zu den mehrfach aufgeführten **Nachteilen** zählten (in absteigender Häufigkeit der Nennungen, vgl. auch Tabelle 22 im Anhang):

- Intransparenz (9)
- Wahrnehmung als unverhältnismässiger Privatheitseingriff (8)
- Befürchtung vor der Unzuverlässigkeit der Software (7)
- Furcht vor Diskriminierung und weiteren sozialen Folgen (7)
- Angst vor Manipulation (3)

- Bedenken hins. der Datensicherheit (3)
- Angst vor Technologieabhängigkeit (2)

Fehlende Transparenz wurde bei neun Diskussionen als zentraler Kritikpunkt angeführt. Je nach diskutiertem Anwendungsgebiet bezog sich diese Kritik auf unterschiedliche Aspekte. *Erstens* bezog sich die Kritik an der fehlenden Transparenz auf den intransparenten Einsatz von Stimm-, Sprach- und Gesichtserkennungstechnologien an sich. Im Kontext der Gesichtserkennung durch polizeiliche Stellen oder in Sportstadien wurde befürchtet, dass der Einsatz nicht oder unzureichend kenntlich gemacht würde. *Zweitens* wurde fehlende Transparenz im Hinblick auf den jeweiligen Technologieeinsatz bemängelt. Hier wurden fehlende Transparenz über die Art, Menge, Speicherung und Auswertung der erhobenen Daten, über die Möglichkeit der (un-)gewollten Weitergabe von Daten an Dritte sowie über die Möglichkeit der intransparenten und ungewollten Sekundärnutzung von Daten für andere Zwecke benannt. *Drittens* wurde mangelnde Transparenz im Hinblick auf die Nachvollziehbarkeit von Softwareergebnissen diskutiert: Dabei ging es primär um die Nachvollziehbarkeit der Ergebnisse im Kontext der Diskussionen zur Erkennung psychischer bzw. physischer Krankheiten, aber auch die Nachvollziehbarkeit der Ergebnisse, die smarte Lautsprecher ausgeben.

Die **Wahrnehmung als unverhältnismässiger Privatheitseingriff** wurde in acht Diskussionen genannt, lediglich im Zusammenhang mit der Erkennung psychischer bzw. physischer Krankheiten kam diese Perspektive so nicht vor. Über alle anderen Anwendungen hinweg waren sich die Teilnehmenden zunächst einig, dass der Einsatz von Stimm-, Sprach- und Gesichtserkennungstechnologien in jedem Fall einen Eingriff in die Privatheit darstellt. Diskutiert wurde dann zumeist darüber, ob es sich dabei um einen gerechtfertigten bzw. verhältnismässigen Eingriff handelt. Dass der Technologieeinsatz schliesslich von fast allen Teilnehmenden als unverhältnismässig bewertet wurde, lag an mehreren Gründen und lässt sich am besten als eine Art Quintessenz der Beurteilung der Summe aller jeweiligen Vor- und Nachteile verstehen. Dabei wurde *erstens* der Status quo des Technologieeinsatzes im Hinblick auf eine als mangelhaft wahrgenommene Transparenz stark kritisiert. *Zweitens* wurde diskutiert, dass auch ein transparenter Einsatz zu einem Freiheitsverlust (bspw. infolge der «Normalisierung» von Massenüberwachung) und damit zu einer ungerechtfertigten Einschränkung der Bevölkerung führen könne. *Drittens* wurde befürchtet, dass der Einsatz dieser Technologien zu unerwünschten Nebeneffekten, wie Diskriminierung und Manipulation, führen könnte. *Viertens* wurde in der polizeilichen Nutzung von Gesichtserkennung eine Ausweitung staatlicher Macht erkannt, der keine adäquaten Massnahmen des Aufbaus von Gegen- oder Kontrollmacht gegenüberstünden. Und *fünftens* wurde eine Diskrepanz zwischen dem wahrgenommenen Nutzen und dem dafür erforderlichen Technologieaufwand bzw. Eingriff in die Grundrechte und Privatheit erkannt. Generell lässt sich festhalten, dass die Teilnehmenden insb. kein Vertrauen in die konkrete Ausgestaltung bzw. den konkreten Einsatz der Technologien hatten, aber sich grundsätzlich mit dem Einsatz einverstanden zeigen könnten, wenn ihre Befürchtungen adressiert werden.

Die **Unzuverlässigkeit von Softwareergebnissen** wurde in sieben Diskussionsrunden als Nachteil genannt. Bei den Diskussionen über Stadionüberwachung und polizeiliche Überwachung wurde die Befürchtung formuliert, dass die falsch-positive Erfassung von Personen zu einer Beeinträchtigung unbescholtener Bürgerinnen und Bürger führen könnte. Bei

der Diskussion über smarte Lautsprecher wurde die mangelhafte Erkennung sprachlicher Akzente und Dialekte angesprochen, die die Nutzung der Geräte erschwert. Bei der Diskussion zur Stimmauthentifizierung wurde die Fähigkeit der Software grundsätzlich infrage gestellt, Personen anhand ihrer Stimme korrekt zu erkennen. Im Kontext der Emotionserkennung bzw. Aufmerksamkeitsanalyse bezweifelten die Diskutanten, dass die Software in der Lage sein würde, Emotionen korrekt zu interpretieren. Fatale Folgen für die Gesundheit Betroffener befürchteten die Teilnehmenden schliesslich im Rahmen der Diskussionen über die Erkennung physischer und psychischer Krankheiten, wenn entsprechende Apps falsch-positive oder falsch-negative Ergebnisse lieferten und dadurch die Nutzenden in falsche Sicherheit oder unnötige Besorgnis versetzten. Es wurde auch darauf hingewiesen, dass besonders falsch-negative Diagnosen problematisch seien und ihrerseits zu negativen gesundheitlichen Folgen führen könnten.

Die **Furcht vor Diskriminierung und weiteren sozialen Folgen** wurde in fünf Diskussionsrunden thematisiert. Bei einer Diskussion über smarte Lautsprecher wurde angenommen, dass Alexa grössere Schwierigkeiten hätte, Frauen oder Personen mit Beeinträchtigungen beim Sprechen zu erkennen. Im Kontext des Einsatzes von Emotionserkennung bei Bewerbungsgesprächen wurde Diskriminierung als Folge unzuverlässiger Softwareergebnisse befürchtet. Im Rahmen der Diskussion polizeilicher Gesichtserkennung wurde befürchtet, dass ein Bias in der Software zur Diskriminierung gewisser Gruppen führen könnte. Es wurde auch befürchtet, dass die automatisierte Erkennung psychischer Krankheiten die Tendenz zur Selbstoptimierung befördern und zu sozialer Diskriminierung bei all jenen führen könnte, die sich dieser Entwicklung verweigern. Bei der Diskussion über die Erkennung physischer Krankheiten wurde befürchtet, dass die Erhebung solcher Daten auch Begehrlichkeiten bei den Krankenkassen weckt und es im Rahmen der Privatversicherungen zu einer (Prämien-)Diskriminierung kommen könnte. Immer wieder wurde auch befürchtet, dass heute unproblematische Daten durch den technischen Fortschritt und die Verknüpfung mit anderen Datenbeständen weitreichendere Schlussfolgerungen erlauben und negativ auf die Menschen zurückfallen könnten.

Die **Angst vor Manipulation** wurde in vier Diskussionsrunden angesprochen. In den Diskussionen über smarte Lautsprecher befürchteten die Teilnehmenden, dass die Geräteher-steller durch die selektive Auswahl der Suchergebnisse oder von Werbung die Handlungen der Nutzenden manipulieren könnten. Bei der Emotionserkennung wurde die Gefahr gesehen, dass die Technologie, sofern sie präzise funktionieren sollte, zur Manipulation von Menschen eingesetzt werden könnte, weil sich aus dem Lesen der Emotionen viele Erkenntnisse über die Persönlichkeit eines Menschen schlussfolgern liessen.

Im Rahmen der Diskussionen zur Stadionüberwachung, zu smarten Lautsprechern und zur Authentifizierung via Stimme äusserten die Teilnehmenden **Bedenken hins. der Datensicherheit**. Dabei wurde insb. den für die Bearbeitung der Daten verantwortlichen Stellen nicht zugetraut, für eine ausreichende Datensicherheit zu sorgen. Zudem fand die Erwähnung dieser Bedenken im Kontext der Diskussion der unkontrollierten Weitergabe von Daten an Dritte und der Speicherung der Daten in der Cloud statt.

Schliesslich wurde im Rahmen der Diskussionen zu smarten Lautsprechern die **Angst vor Technologieabhängigkeit** als Nachteil genannt. In beiden Diskussionen verwiesen die Teilnehmenden darauf, dass der mögliche Übergang von einer haptischen Gerätebedienung zu

allgegenwärtiger Sprachsteuerung nachfolgende Generationen dahin gehend negativ betreffen könnte, dass diese nicht mehr Schreiben und Lesen lernen.

4.3.3. Zusammenfassung der wesentlichen Empfehlungen

Deutliche Gemeinsamkeiten über fast alle Fokusgruppen hinweg gab es auch im Hinblick auf die Empfehlungen (vgl. auch Tabelle 23). In fast allen Diskussionen waren sich die Fokusgruppen-Teilnehmenden einig, dass der **Gewährleistung von ausreichender Transparenz** im Hinblick auf den Einsatz von Stimm-, Sprach- und Gesichtserkennungstechnologien eine zentrale Rolle zukommt (van den Broek et al. 2017, S. 28–29).[180] Dabei bezogen sich die Empfehlungen der Teilnehmenden im Einzelnen auf unterschiedliche Dimensionen von Transparenz:

- Transparenz über den jeweiligen Einsatz
- Transparenz über die Art, Menge, Speicherung und Auswertung der erhobenen Daten
- Transparenz über die Weitergabe von Daten an Dritte und über eine mögliche Sekundärnutzung
- Transparenz über die Sekundärnutzung der Daten
- Transparenz im Hinblick auf die Nachvollziehbarkeit von Softwareergebnissen

Deutlich äusserten die Teilnehmenden ausserdem eine Vielzahl von **Erwartungen an die Politik**. Die im Rahmen von acht Gesprächsrunden diskutierte **Befürwortung spezifischer gesetzlicher Vorschriften** lässt sich dabei eher als generelle Empfehlung verstehen, auf die im Rahmen der einzelnen Diskussionen genannten spezifischen Herausforderungen mittels Regulierung zu reagieren. In diesem Sinne ist der dringende Wunsch nach Transparenz ein Element dieses Rufs nach staatlicher Regulierung. Weitere Erwartungen an die Politik sind:

- Gesetzliches Verbot spezifischer, besonders problematischer Zwecke (z.B. über smarte Lautsprecher verbreiteter gezielter politischer Werbung oder des Tragens von Datenbrillen in der Öffentlichkeit)
- Gewährleistung der technischen Zuverlässigkeit von Stimm-, Sprach- und Gesichtserkennungstechnologien (z.B. mittels Zertifizierung)
- Vorantreiben von Aufklärungskampagnen über Technologierisiken
- Sicherstellung der Kontrolle und Einhaltung der geltenden Gesetze durch unabhängige Dritte

Daneben wurde auch eine Reihe von **Handlungsempfehlungen** formuliert, die sich **an die Betreiber und Hersteller von Stimm-, Sprach- und Gesichtserkennungstechnologien** richten. Darunter insb.:

[180] In der einzigen Diskussion, in der dieser Aspekt nicht vorkommt, nämlich der deutschsprachigen Diskussion polizeilicher Gesichtserkennung, wurde der Nutzen des Einsatzes grundsätzlich angezweifelt, weswegen erst gar nicht über umsetzungsbezogene Massnahmen wie der Gewährleistung von Transparenz diskutiert wurde.

- Schulung des Personals, das die Technologie einsetzt, zum kritischen Umgang mit Daten bzw. Analyseergebnissen
- Förderung alternativer Methoden anstelle oder parallel zum Technologieeinsatz
- Gewährleistung der Datensicherheit
- Datenschutzfreundliches Technikdesign
- Bearbeitung von Daten auf den Endgeräten bzw. in der Schweiz

Weil die Teilnehmenden v.a. den Staat in der Verantwortung sahen, wurden nur wenige Empfehlungen in Bezug auf **gesellschaftliche Handlungsmöglichkeiten** formuliert. Die Vorschläge, wie das Aussenden von Störsignalen zur Beeinträchtigung der Gesichtserkennungsfunktion smarter Brillen oder das Bekleben des eigenen Gesichts, bauen auf dem Vorschlag zu einem Verbot des Tragens smarter Brillen in der Öffentlichkeit und wurden als alternative Schutzvorkehrung für den Fall des Scheiterns des Verbots vorgeschlagen.

4.4. Zwischenfazit

Die Fokusgruppen-Teilnehmenden erkannten bei den Anwendungen der Stimm-, Sprach- und Gesichtserkennung insgesamt weitaus mehr Nachteile bzw. Risiken als Vorteile und waren sich auch deutlich häufiger über die identifizierten Nachteile einig.

Zur Adressierung der Nachteile sahen die Teilnehmenden in erster Linie den Staat in der Pflicht. Dieser solle mittels weitgehender Transparenzregelungen, Verbote, Kontrollen und Aufklärungskampagnen die gesellschaftsverträgliche Nutzung von Stimm-, Sprach- und Gesichtserkennungstechnologien gewährleisten. In zweiter Instanz sollen auch die Betreiber von entsprechenden Anwendungen und die Technologiehersteller durch eigene Massnahmen zum vertrauenswürdigen Einsatz der Technologien beitragen, so etwa durch die Schaffung von Alternativen, Datensicherheit und Schulung des Personals. Der Gesellschaft selbst wollten die Teilnehmenden hingegen kaum unmittelbare Schutzverantwortung zuschreiben.

Insgesamt schienen die Fokusgruppen-Teilnehmenden nicht nur den verschiedenen Anwendungen von Stimm-, Sprach- und Gesichtserkennung zu misstrauen, sie zeigten auch einen Mangel an Vertrauen in die Fähigkeit und den Willen der Politik, bestehendes Recht durchzusetzen und, wo nötig, neues Recht zu erlassen. Neben dem Regulierungsbedarf wurde auch die Bedeutung einer intensivierten gesellschaftlichen Debatte über derartige neue Technologien und Anwendungen hervorgehoben. Nachdem Medien und Öffentlichkeit das Thema in den vergangenen Jahren bereits verstärkt aufgegriffen haben, ist nun die Politik gefordert, sich dieser Themen in verstärkter Weise anzunehmen – und zu gegebener Zeit der politischen Debatte auch wohlüberlegte Massnahmen folgen zu lassen.

5. Die Perspektive von Bürgerinnen und Bürgern in der Bevölkerungsumfrage

Neben der qualitativen Untersuchung der Bevölkerungsmeinung im Rahmen von Fokusgruppen wurde diese auch quantitativ mittels einer repräsentativen Bevölkerungsumfrage untersucht. Im Vordergrund standen dabei die Fragen, welche Aspekte der besprochenen Stimm-, Sprach- und Gesichtserkennungsanwendungen aus welchen Gründen als erwünscht oder unerwünscht gelten und wie Politik und Gesellschaft mit diesen Fragen umgehen sollten.

In Abschnitt 5.1 werden die Ziele und die Durchführung der Bevölkerungsumfrage vorgestellt. Im Anschluss (5.2) werden die Ergebnisse der Bevölkerungsumfrage pro Anwendungsgebiet vorgestellt und diskutiert. Jede dieser Diskussionen gliedert sich wiederum in die zwei Unterkapitel-Typen (1) Erwünschtheit bzw. Nutzung (im Falle smarter Lautsprecher) inkl. der Diskussion der Vor- und Nachteile sowie (2) Wünsche hins. der Ausgestaltung der jeweiligen Anwendungsfelder. Bei den Fragen zu den drei Anwendungsgebieten smarte Lautsprecher, Gesichtserkennung durch polizeiliche Stellen und Gesichtserkennung in Sportstadien kommt zusätzlich hinzu, welche Schutzmassnahmen seitens Betroffener umgesetzt werden (könnten). Allfällige statistische Auffälligkeiten hins. geschlechts-, alters-, bildungs- oder (sprach-)regionsbasierter Differenzen werden in den entsprechenden Abschnitten benannt. Gemeinsame und querliegende Vor- und Nachteile sowie Empfehlungen werden in Abschnitt 5.3 zusammengefasst und darauf basierende Schlussfolgerungen (5.4) gezogen.[181]

5.1. Ziele und Durchführung der Bevölkerungsumfrage

Die repräsentative Bevölkerungsumfrage dient dem besseren Verständnis der gesellschaftlichen Wahrnehmung von Stimm-, Sprach- und Gesichtserkennungstechnologien. Der Fokus lag darauf, zu erfahren, für wie wünschenswert die Befragten den Einsatz in den unterschiedlichen Anwendungsgebieten halten. Im Anschluss wurden die Gründe für die geäusserten Meinungen abgefragt und wie deren Nutzung akzeptabler gestaltet werden kann.

Die Umfrage wurde von einem professionellen Meinungsforschungsinstitut Mitte Oktober 2021 mit 1000 Teilnehmenden (Onlinepanel) durchgeführt. Dieses Sample war repräsentativ für die Schweizer Bevölkerung. Der Fragebogen wurde in einer deutschen, französischen und italienischen Fassung über das Internet angeboten.[182]

---

[181] Aus Gründen der besseren Lesbarkeit und Übersichtlichkeit weichen die Item-Beschriftungen teilweise von der Version ab, die den Befragten angezeigt wurde. Auf den vollständigen Datensatz der Befragung, inkl. der Originalformulierungen kann unter der folgenden Adresse zugegriffen werden: 10.5281/zenodo.6838773.

[182] Das Projektteam erkundigte sich zuvor bei mehreren Meinungsforschungsinstituten danach, ob der Fragebogen zusätzlich auch in rätoromanischer Sprache angeboten werden sollte. Weil die Institute übereinstimmend angaben, dass eine gute Abdeckung der rätoromanischsprechenden Bevölkerung über Deutsch und Italienisch gegeben sei, wurde davon Abstand genommen.

Erstellt wurde der Fragebogen im Frühjahr/Sommer 2021 und er enthielt 152 Fragen. Diese wurden auf zwei Gruppen von jeweils 500 Befragten verteilt, wobei jede Gruppe zu unterschiedlichen Anwendungsgebieten (Tabelle 16) gefragt wurde.[183] Angaben zur Repräsentativität nach Geschlecht, Schulbildung, Alter, Wohnort und Erstsprache der Befragten können entnommen werden.

Tabelle 16: Aufteilung der Anwendungsfälle pro Befragten-Gruppen

Befragte	**Anwendungsgebiet**
N=1000	• Smarte Lautsprecher
N=500	• Gesichtserkennung durch polizeiliche Stellen • Erkennung von physischen Krankheiten • Jedermann-Identifikation
N=500	• Gewaltprävention und -aufklärung in Sportstadien • Gesichtserkennung im Schulkontext zur Aufmerksamkeitsanalyse • Erkennung von psychischen Krankheiten • Authentifizierung via Stimme bei Telefonbanking

Tabelle 17: Demografie der Befragten (N=1000)

		Anzahl	**Prozent**
Geschlecht	Männlich	496	49,6
	Weiblich	497	49,7
	k. A.	7	0,7
Schulbildung	nicht akademisch	707	70,7
	akademisch	293	29,3
Alter	16–34	312	31,2
	35–54	394	39,4
	55–75	287	28,7
	k.A.	7	0,7
Wohnort	Städtisch	776	77,6
	Ländlich	224	22,4 %
Erstsprache der Befragten	Deutsch	637	63,7 %
	Französisch	321	32,1 %
	Italienisch	40	4,0 %

[183] Da erwartet wird, dass die Nutzung smarter Lautsprecher besonders stark zunimmt, wurden die Fragen zu diesem Anwendungsfall allen 1000 Befragten gestellt.

5.2. Ergebnisse der Bevölkerungsumfrage

5.2.1. Smarte Lautsprecher

Nutzung

Im Hinblick auf die Frage nach der Nutzung smarter Lautsprecher gaben 63 %[184] der Befragten an, dass sich kein solches Gerät in ihrem Haushalt befindet. 37 % besitzen oder nutzen einen smarten Lautsprecher, welche sich, wie in Abbildung 14 beschrieben, auf Selbst- oder Mitnutzung aufteilen.

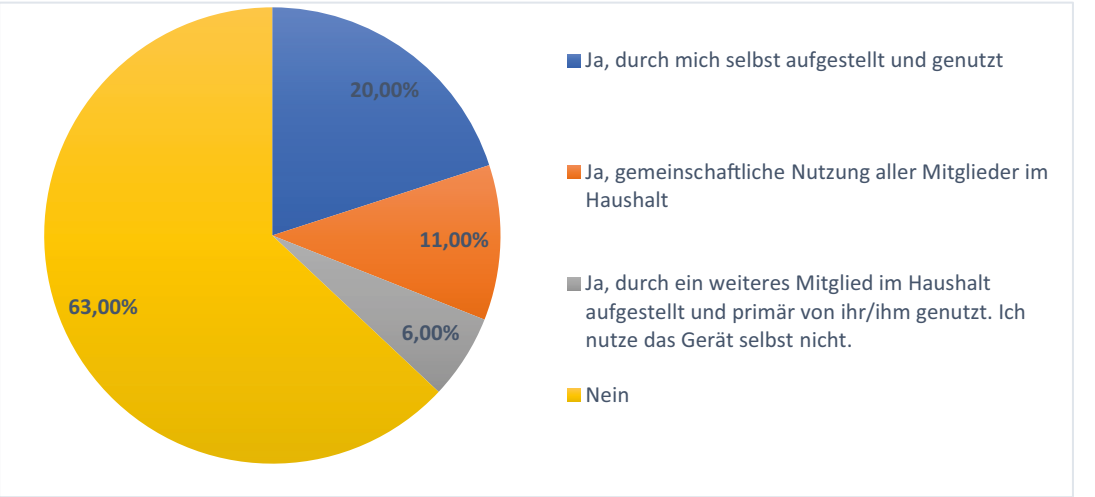

Abbildung 14: Auswertung zur Nutzung von smarten Lautsprechern

Männer nutzen smarte Lautsprecher signifikant häufiger (28 % bei p <0,01) als Frauen (13 % bei p <0,01) und auch bei den Altersgruppen zeigen sich signifikante Unterschiede: Sowohl bei der alleinigen als auch der gemeinschaftlichen Nutzung ist ein signifikant höherer Anteil der jüngsten Alterskohorte (30 % bzw. 16 %, beides bei p <0,01) vorzufinden, während die niedrigsten Werte bei den 55–75-Jährigen vorzufinden sind (14 % bei p <0,05 bzw. 5 % bei p <0,01). Signifikant sind ferner die Unterschiede nach Erstsprache der Befragten: Während nur 13 % (bei p <0,01) der französischsprachigen Befragten angegeben haben, einen smarten Lautsprecher selbst aufgestellt zu haben und zu nutzen und 75 % (bei p <0,05) keinen nutzen, liegt die Nutzung bei anderen insgesamt deutlich höher: In der italienischen Schweiz liegt der individuelle und gemeinschaftliche Nutzungsgrad bei 48 %, in der Deutschschweiz bei 35 % und in den französischsprachigen Regionen hingegen bei insgesamt 21 %.

Wie Abbildung 15 entnommen werden kann, sind Google Assistant (45 %) und Amazon Echo (29 %) unter den Befragten (n=204), die angegeben haben, dass sie einen selbst

[184] Alle Prozentangaben in diesem Dokument mit gerundete Werte, sodass Summen > 100% möglich sind.

aufgestellten smarten Lautsprecher nutzen, am weitesten verbreitet. Auf Platz drei folgt der Apple HomePod mit 17 % und unter Sonstiges fallen v.a. Geräte von Swisscom (9 %).

Knapp die Hälfte dieser Befragten gab an, dass sie ihr Gerät zwischen einem Monat und zwölf Monaten nutzen (49 %). 23 % nutzen den Lautsprecher erst seit weniger als einem Monat. Knapp 19 % zwischen einem und drei Jahren und nur 9 % seit mehr als drei Jahren.

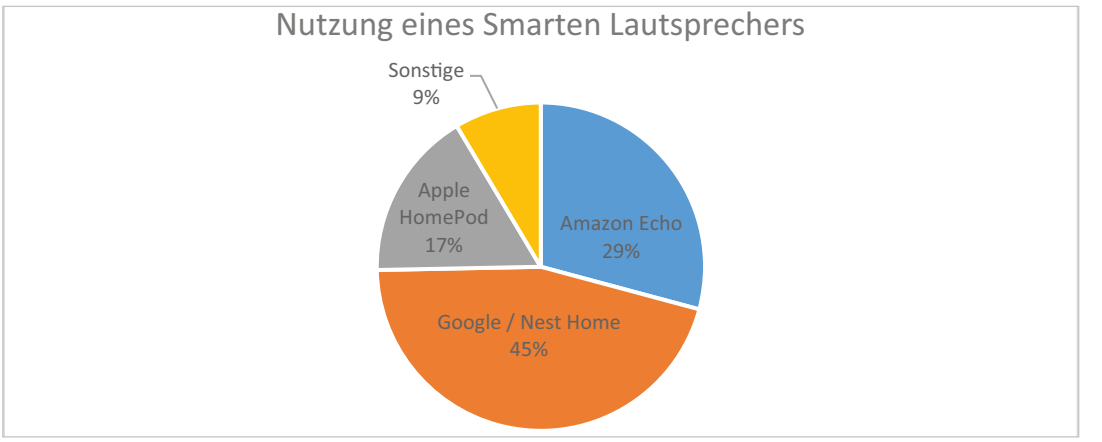

Abbildung 15: Verteilung der verschiedenen Arten Lautsprecher

Gründe für die Anschaffung

Von den Befragten (n=312), die zuvor angegeben hatten, einen smarten Lautsprecher entweder persönlich oder gemeinschaftlich zu nutzen (Abbildung 16), gaben 69 % an, dies aus technischer Neugier zu tun. 64 % der Befragten versprachen sich zudem eine Erleichterung des Alltags. Für 51 % spielte der günstige Preis der Geräte und für 48 % der Befragten die Nutzung von smarten Lautsprechern durch Freunde und Bekannte eine wichtige Rolle. Zwar gaben 39 % an, dass sie von Freunden zum Kauf überredet wurden bzw. ihnen ein smarter Lautsprecher geschenkt wurde, doch gaben zugleich 40 % an, dass dies für ihre Nutzung keine Rolle spiele. In den Freifeldern wurden als weitere Gründe die nutzerfreundliche Bedienung, die Steuerung des smarten Zuhauses, die gute Tonqualität sowie die Möglichkeit, im Falle eines medizinischen Notfalles einen Notruf absetzen zu können, genannt.

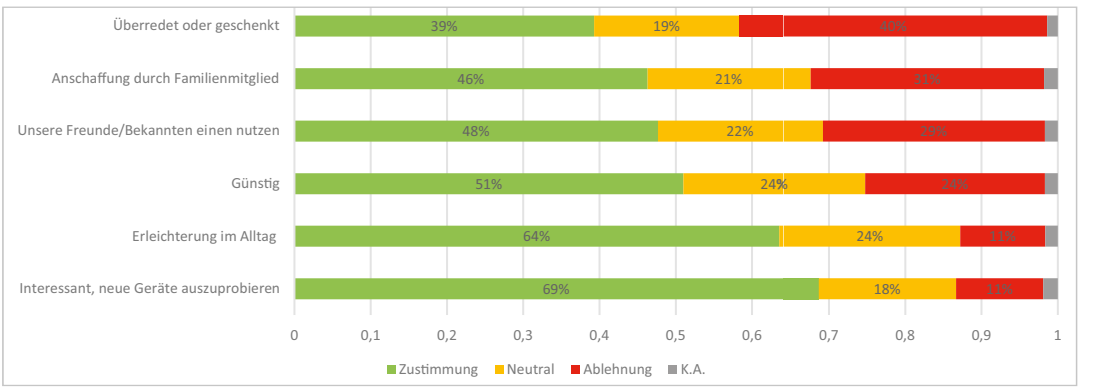

Abbildung 16: Gründe für die Anschaffung eines smarten Lautsprechers von Nutzern, die bereits einen besitzen (Mehrfachnennung möglich)

Jene Befragte (n=688), die angegeben hatten, derzeit keinen smarten Lautsprecher zu nutzen, wurden ausserdem dazu befragt, ob sie sich eine Nutzung in der Zukunft vorstellen könnten und welche Gründe für ihre Entscheidung ausschlaggebend wären. 41 % dieser Befragten gaben an, dass sie auch in Zukunft keine smarten Lautsprecher nutzen würden. 39 % gaben an, dass sie unsicher sind und 14 % bejahten die Frage einer möglichen künftigen Anschaffung (Abbildung 17).

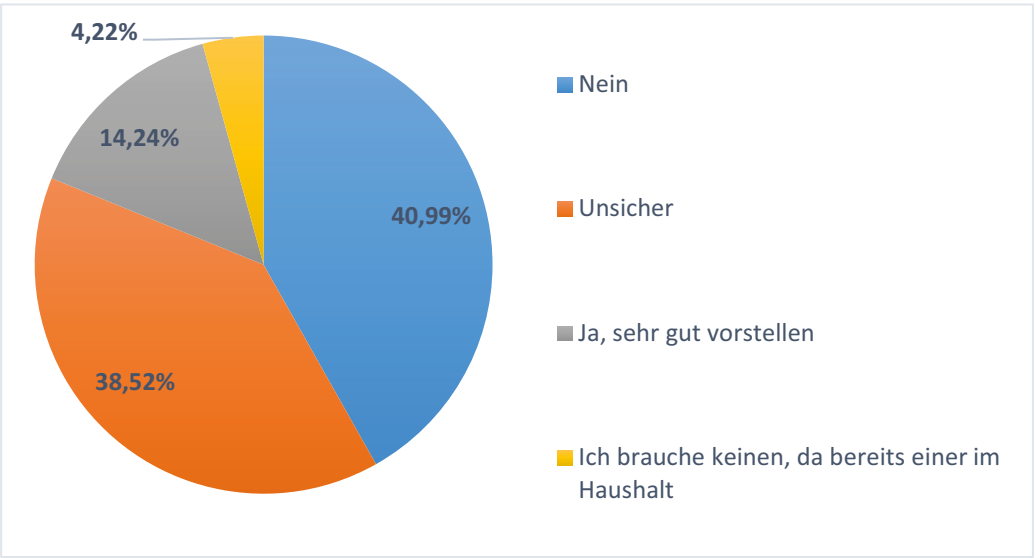

Abbildung 17: Vorstellungen zur Anschaffung eines smarten Lautsprechers

Befragte, die unsicher über die Anschaffung sind (n=265) (Abbildung 18), nannten als wichtigsten Grund für ihre Unsicherheit, dass sie die Vor- und Nachteile nicht ausreichend abgewogen hätten (65 %). Mehr als die Hälfte der Befragten verwies zudem auf Sorgen hins. der Sicherheit smarter Lautsprecher (55 %). Unter diesen Antwortenden waren es mit 75 % v.a. die Befragten in der Altersgruppe zwischen 16–34 Jahren, die Bedenken äusserten, während die 35–54-Jährigen und die 55–75-Jährigen mit 48 % und 47 % seltener entsprechende Bedenken nannten. 42 % gaben an, dass sie es schwierig finden, sich ausreichend über das Thema zu informieren. Ein mangelndes Interesse spielte nur für einen kleinen Teil Befragten eine Rolle: Immerhin 39 % gaben an, dass sie am Thema smarter Lautsprecher Interesse haben.

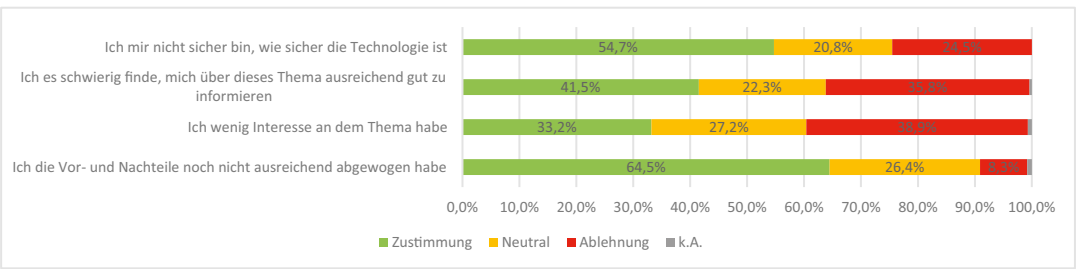

Abbildung 18: Gründe jener Befragten, die unsicher im Hinblick auf die Anschaffung eines smarten Lautsprechers sind (Mehrfachnennung möglich)

Befragte, die sich eine Anschaffung nicht vorstellen können (n=282) (Abbildung 19), führten mit 69 % an, dass sie *zu wenige* sinnvolle Einsatzzwecke für smarte Lautsprecher sehen. 57 % gaben sogar an, dass sie überhaupt keinen sinnvollen Einsatzzweck kennen.

Weitere wichtige Gründe waren Datenschutzbedenken und ein grundsätzliches Desinteresse an der Nutzung solcher Technologien (beide mit jeweils 67 %). 60 % der Befragten äusserten zudem ihre Besorgnis, dass smarte Lautsprecher alles, was gesprochen wird, mithören und aufzeichnen würden. Sowohl die Äusserung grundsätzlicher Datenschutzbedenken als auch die Sorge vor dem Mithören wurde von einer grossen Mehrheit der deutschsprachigen Teilnehmenden geäussert (75 % bzw. 73 %). Entsprechende Sorgen wurden seitens der französischsprachigen Teilnehmenden vergleichsweise seltener geäussert (59 % bzw. 47 %).[185]

Die Befürchtung, dass der Lautsprecher den eigenen Dialekt nicht verstehen könnte, ist für 21 % der Befragten ein Grund, sich keinen smarten Lautsprecher beschaffen zu wollen. Unter diesen befindet sich mit 29 % ein signifikant (p=<0,05) grösserer Anteil Befragter aus der Deutschschweiz als solche aus der Romandie (12 % bei p=>0,05) und der italienischen Schweiz (17 %).

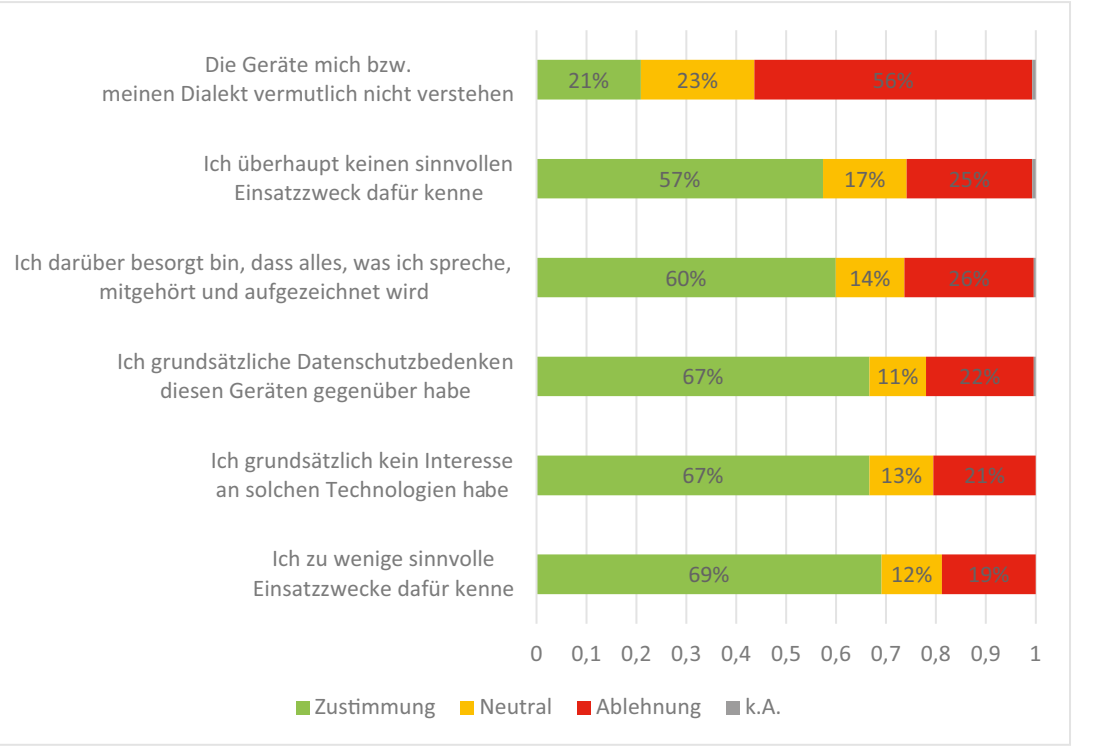

Abbildung 19: Gründe für die Nicht-Anschaffung eines smarten Lautsprechers (Mehrfachnennung möglich)

[185] Im Falle italienischsprachiger Befragter waren es in beiden Fällen 17 %, allerdings gilt auch hier dasselbe Problem, dass lediglich sechs italienischsprachige Personen im Sample enthalten waren und die Resultate folglich wenig Aussagekraft besitzen.

Bewertung der Möglichkeit der Emotionsanalyse

Zudem wurden alle Teilnehmenden (N=1000) gefragt, wie sie die Möglichkeit bewerten, dass aus Merkmalen der Stimme Informationen über ihre Emotionen abgleitet werden können (Abbildung 20 bis Abbildung 22). Die Mehrheit der Befragten sprach sich gegen die Emotionserkennung aus (Abbildung 20) und führte insb. das Unwissen darüber, wer die Daten erhält, und die Befürchtung negativer Konsequenzen in der Zukunft (60 %) sowie die Befürchtung, dass Unternehmen zu viel über die eigene Person erfahren (51 %), an. Die Befürchtung, dass eine solche Erkennung fehlerhaft wäre und vorwiegend für noch mehr personalisierte Werbung genutzt würde, teilten 43 % bzw. 48 % der Befragten. Jene Befragten, die den Einsatz von Emotionserkennung befürworten (Abbildung 21), verwiesen ungefähr gleichhäufig auf die Möglichkeit des verbesserten Trainings von Algorithmen für andere (z.B. medizinische) Zwecke (37 %) sowie auf Vorteile, die sich bei der Nutzung von Diensten über smarte Lautsprecher mittels verbesserter Personalisierung ergeben könnten (35 %). Hierbei ist es auffallend, dass insb. die Altersgruppe der 16–34-Jährigen den entsprechenden Einsatz befürwortet, mit 46 % (p=<0,01) für Trainingszwecke und mit 45 % (p=<0,01) für Personalisierungszwecke. Zugleich wurde der Technologieeinsatz signifikant weniger häufig von der Altersgruppe der 55–75-Jährigen befürwortet (27 % bei p=>0,01 bzw. 23 % bei p=>0,01).

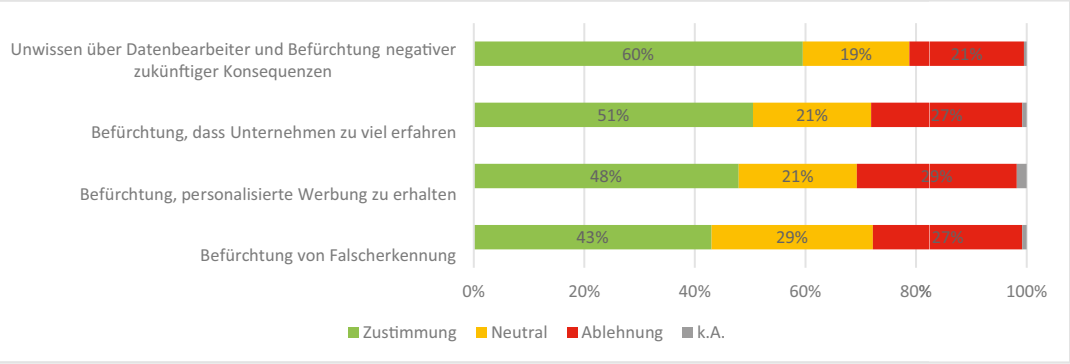

Abbildung 20: Gründe für die Ablehnung von Emotionserkennung (Mehrfachnennung möglich)

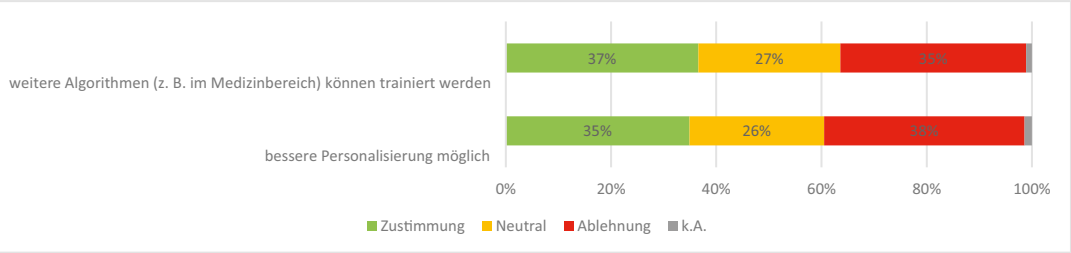

Abbildung 21: Gründe für die Befürwortung von Emotionserkennung (Mehrfachnennung möglich)

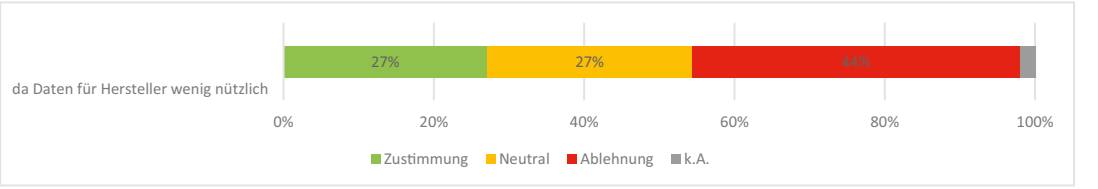

Abbildung 22: Gründe für Unklarheiten über Emotionserkennung

Schutzmassnahmen

Zuletzt wurden alle Besitzer eines smarten Lautsprechers (n=367) gefragt, ob und inwiefern sie Schutzmassnahmen getroffen haben, um ihre persönlichen Daten besser zu schützen (Abbildung 23). Hier gaben 27 % an, dass sie keinerlei Bedenken haben und daher keine Schutzmassnahmen getroffen haben. 15 % der Befragten gaben an, aufgrund von Bedenken Schutzmassnahmen zu treffen. Weitere 34 % gaben an, zwar selbst keinerlei Bedenken zu haben, aber trotzdem Schutzmassnahmen, z.B. für Gäste oder Mitbewohner, zu treffen. Schliesslich gaben 25 % der Befragten an, dass sie trotz vorhandener Bedenken keine Schutzmassnahmen getroffen haben, darunter mit 34 % signifikant häufiger (p=<0,05) Frauen.

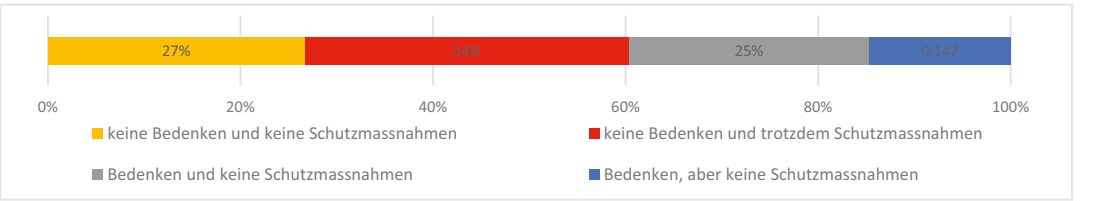

Abbildung 23: Bedenken und Schutzmassnahmen der Befragten

Von denjenigen, die Schutzmassnahmen getroffen haben (n=178) (Abbildung 24), wurde mit 66 % (p=<0,05) am häufigsten die Beschränkung der für Zusatzprogramme (Skills) erteilten Zugriffsberechtigungen genannt, gefolgt von der Vergabe eines nutzerprofilspezifischen Passworts (59 %), der Anpassung der Datenschutzeinstellungen (58 %), der Verhaltensänderung in Form der Nicht-Erwähnung sensibler Informationen bei Interaktionen mit dem Gerät (56 %) oder dem Abschalten des smarten Lautsprechers (56 %).

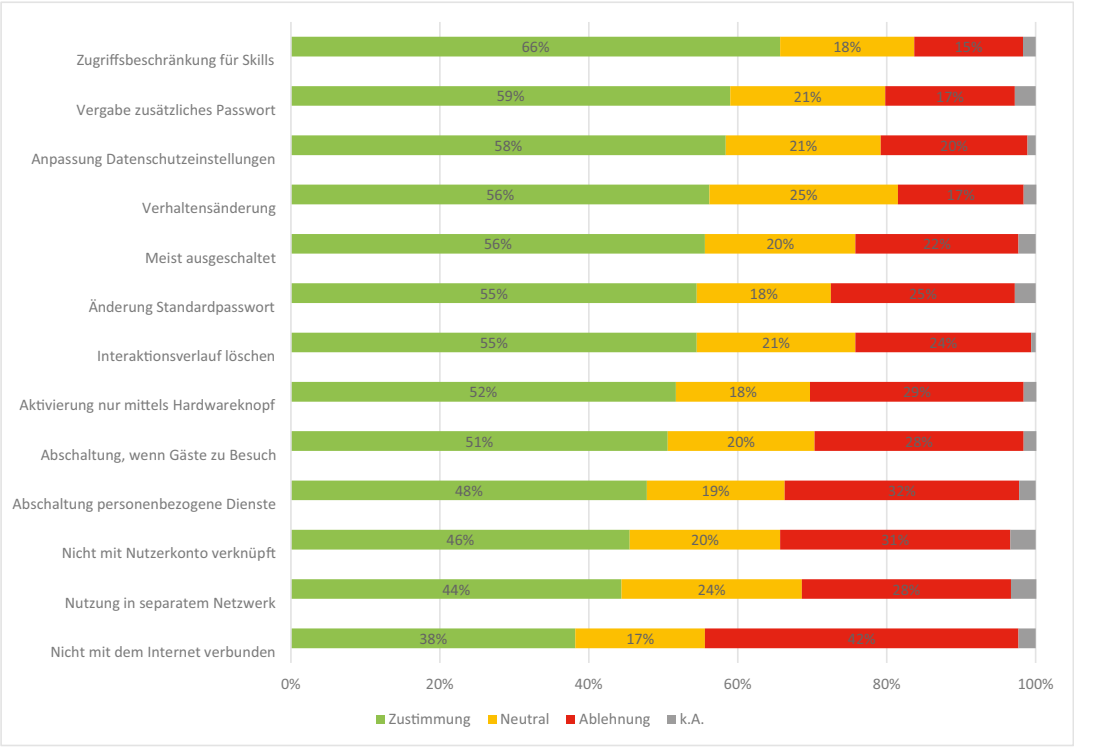

Abbildung 24: Getroffene Schutzmassnahmen (Mehrfachnennung möglich)

5.2.2. Gesichts- und Spracherkennung durch polizeiliche Stellen

Akzeptanz

Im Hinblick auf die Frage, wie die Schweiz mit Gesichts- und Spracherkennung durch polizeiliche Stellen umgehen sollte (Abbildung 25), äusserte die knappe Mehrheit der Befragten (33 %), dass sie keine klare Meinung zu dem Thema habe. 33 % sprachen sich für einen Einsatz und 31 % für ein Verbot bis auf Weiteres aus.

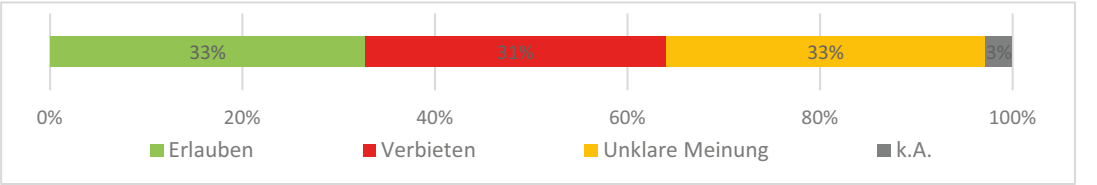

Abbildung 25: Akzeptanz zum Einsatz von Gesichts- und Spracherkennung durch die Polizei

Ein zeitweises Verbot wurde v.a. von Personen aus der jüngsten Altersgruppe der 16–34-Jährigen gefordert (39 % gegenüber 30 % der 35–54-Jährigen und 24 % der 55–75-Jährigen). Den grössten Zuspruch fand der Einsatz in der Altersgruppe der 55–75-Jährigen (38 % gegenüber 32 % der 25–54-Jährigen und 29 % der 16–34-Jährigen).

Von den 156 Befragten, die sich für ein Verbot ausgesprochen hatten (Abbildung 26), wurde mit 81 % die Furcht vor einem Missbrauch der erhobenen Daten für andere Zwecke und die Ablehnung eines grossen Einflusses von Maschinen und KI auf das öffentliche Leben mit 80 % am häufigsten als Gründe genannt. Ausserdem stimmten 78 % der Befragten der Aussage zu, dass Gesichts- und Stimmerkennung zu einer anlasslosen Massenüberwachung der Bevölkerung führe, und 74 % der Aussage, dass sie generell jede Form der Überwachung ablehnen. Die Aussagen, dass bestimmte Bevölkerungsgruppen diskriminiert werden könnten (69 %) und dass man selbst ohne Grund ins Visier der Fahnder geraten könnte (64 %), fanden vergleichsweise etwas weniger Zustimmung.

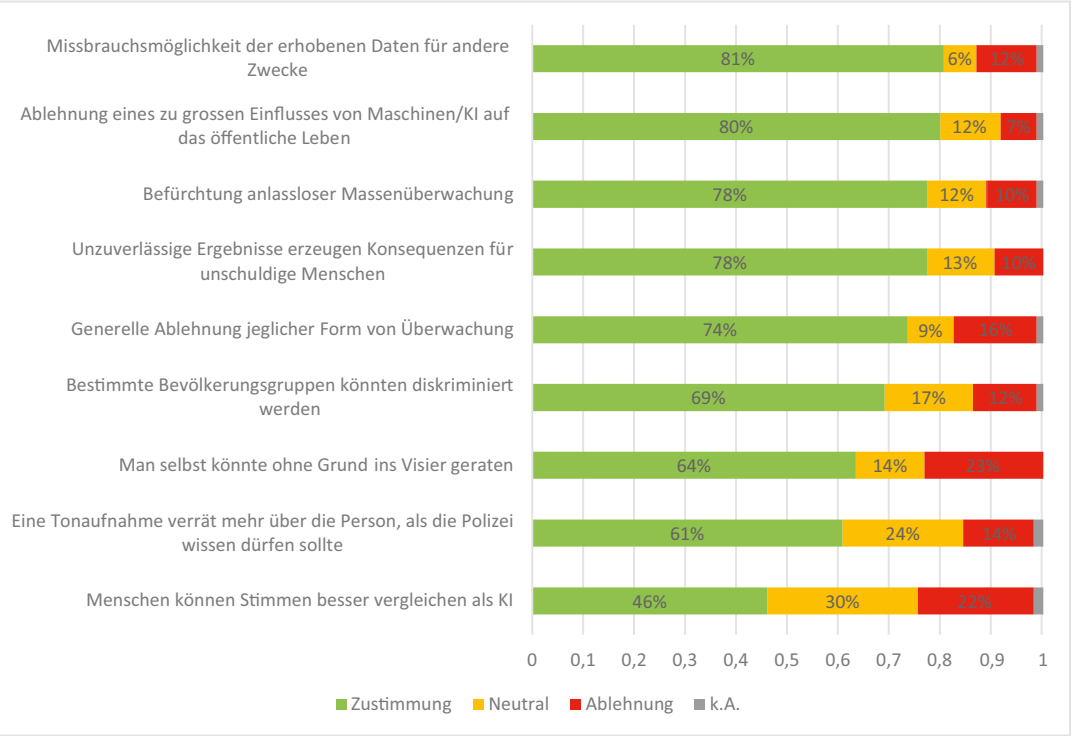

Abbildung 26: Begründungen derjenigen, die sich für ein Verbot von Gesichts- und Spracherkennung durch die Polizei aussprachen (Mehrfachnennung möglich)

Befragte, die zum Thema noch keine klare Meinung haben (n=166) (Abbildung 27), gaben überwiegend an, sich noch nicht ausreichend mit der Thematik befasst zu haben (76 %). 68 % der Befragten stimmten ausserdem der Aussage zu, dass es von den konkreten Einsatzzwecken abhängen werde und 67 % begründeten ihre unklare Meinung damit, ob die Daten nicht auch für andere Zwecke missbraucht werden könnten. Grössere Unklarheit bestand auch darüber, ob Gesichts- und Spracherkennungstechnologien überhaupt ein geeignetes Mittel zur Gefahrenabwehr sind und ob Menschen nicht in stärkerem Masse zu Diskriminierung neigen könnten als eine Software.

Die Befürworter des Einsatzes von Gesichts- und Spracherkennungstechnologien durch die Polizei (n=164) (Abbildung 28) gaben am häufigsten an, dass dies eine geeignete Methode zum Auffinden vermisster Menschen sei (88 %), gefolgt vom Einsatz gegen den Terrorismus (87 %) und der Aufklärung von Mord- und Vergewaltigungsfällen allgemein (85 %). 82 % ver-

traten dabei die Ansicht, dass bereits das Vorhandensein der Gesichtserkennung eine abschreckende Wirkung entfalten werde. 81 % stimmten zudem der Aussage zu, dass Gesichts- und Spracherkennung selbst eine geeignete Methode zur Bekämpfung von kleineren Delikten wie Taschendiebstahl seien. Als eine geeignete Methode zur Bekämpfung von Kinderpornografie bewerteten die Technologien 80 % der Befragten. Schliesslich bejahten 76 % die Aussage, dass jene, die sich angemessen verhalten, auch nichts vom Technologieeinsatz zu befürchten hätten.

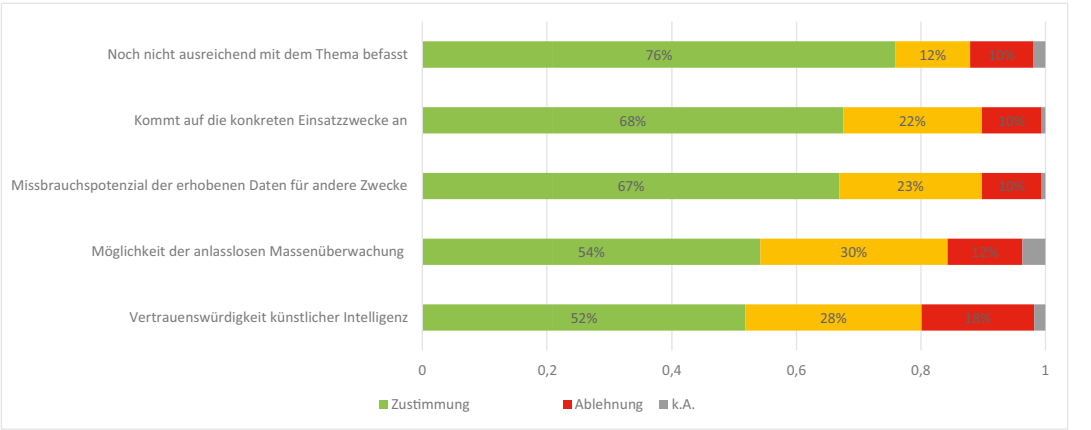

Abbildung 27: Top-5-Begründungen derjenigen, die unschlüssig über den Einsatz von Gesichts- und Spracherkennung durch die Polizei sind (Mehrfachnennung möglich)

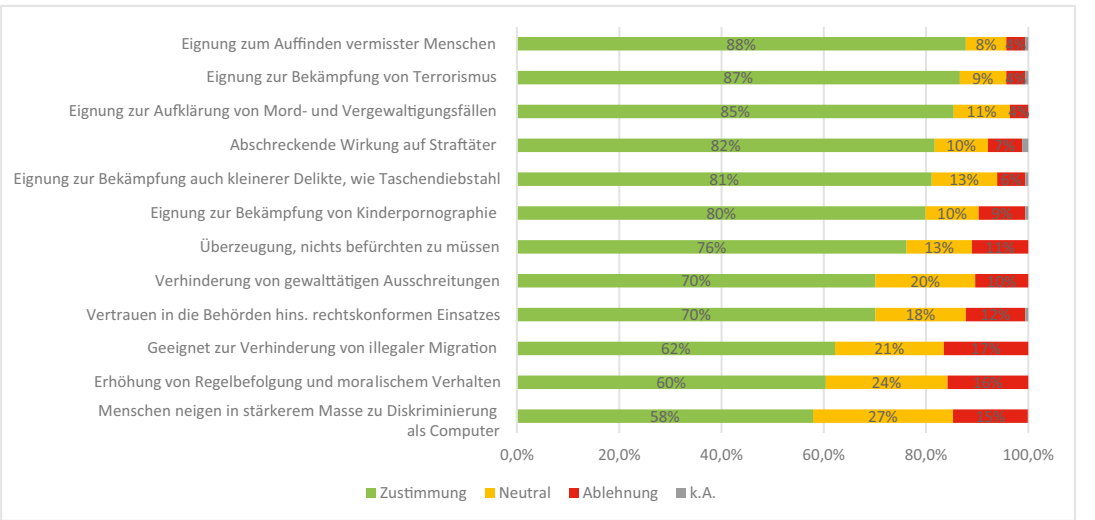

Abbildung 28: Begründungen derjenigen, die sich für den Einsatz von Gesichts- und Spracherkennung durch die Polizei aussprachen (Mehrfachnennung möglich)

Schutzmassnahmen vor Gesichts- und Spracherkennung durch polizeiliche Stellen

Alle Teilnehmende, die sich zuvor ablehnend oder unschlüssig bezüglich des Einsatzes der Gesichts- und Spracherkennung geäussert hatten, wurden dazu befragt, ob und inwiefern sie verschiedene Schutzmassnahmen vor Gesichts- und Spracherkennungstechnologien durch die Polizei ergreifen würden.

Die Befragten mit ablehnender Haltung (n=156) stimmten am häufigsten den Aussagen zu, dass sie Orte vermeiden würden, an denen Gesichts- und Spracherkennungstechnologien eingesetzt werden (67 %), dass sie sich bei den Datenschutzbehörden und anderen Stellen über den Einsatz beschweren würden (65 %) und andere warnen würden (64 %). 44 % stimmten der Aussage zu, dass sie politisch gegen den Einsatz tätig würden, während 35 % der Befragten dies ablehnten (Abbildung 29). Eine Maskierung oder ein anderes aktives Vorgehen gegen die Überwachung kam für diese Befragtengruppe mehrheitlich nicht infrage. Während diese Fragen von 31 % bzw. 33 % bejaht wurden, lehnten 49 % bzw. 47 % diese Form von Schutzmassnahmen ab.

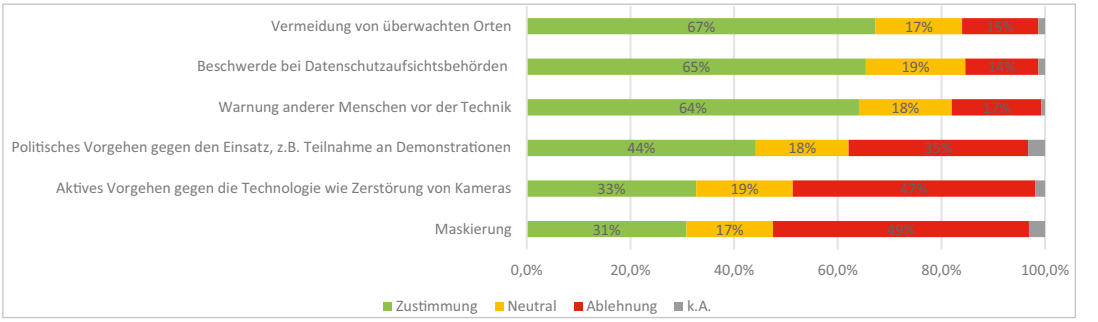

Abbildung 29: Meinungen der Einsatzgegner zu Schutzmassnahmen (Mehrfachnennung möglich)

Um einen Einblick zu bekommen, ob und inwiefern jene Befragte, die sich unschlüssig bezüglich des Technologieeinsatzes geäussert hatten (n=166), ebenfalls bereit wären, Massnahmen zum Schutz vor der Überwachung zu treffen, wurden dieselben Fragen auch an diese Gruppe gestellt. Lediglich die Aussage, dass Einsatzorte gemieden würden, erhielt eine knappe mehrheitliche Zustimmung von 34 %, wobei genau dieselbe Prozentzahl der Befragten diese Aussage ablehnen. Ansonsten wurden alle aufgeführten Schutzmassnahmen mehrheitlich abgelehnt. Auch bei dieser Befragtengruppe wurden die Maskierung (63 %) und ein aktives Vorgehen (62 %) am stärksten abgelehnt (Abbildung 30).

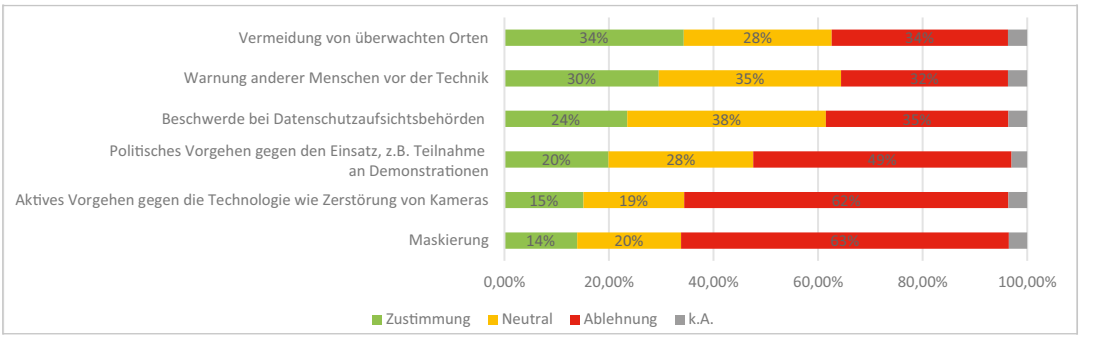

Abbildung 30: Top-5-Meinungen von unschlüssigen Personen zu Schutzmassnahmen (Mehrfachnennung möglich)

Mögliche Einsatzzwecke

Alle Teilnehmer (N=500) wurden dazu befragt, welche Einsatzzwecke der Gesichts- und Spracherkennung durch polizeiliche Stellen sie für sinnvoll und angemessen halten (Abbildung 31). Den grössten Zuspruch fand die Nutzung zum Auffinden vermisster Menschen (76 %), gefolgt von der Bekämpfung schwerer Straftaten wie Mord und Vergewaltigung (73 %). Grösseren Zuspruch erhielt auch der Einsatz zum Zwecke der Bekämpfung von Kinderpornografie (68 %), organisierter Kriminalität (68 %) sowie Terrorismus (68 %). 64 % befürworteten den Einsatz zur Bekämpfung von Kriminalität im Allgemeinen und 60 % auch den Einsatz zur Bekämpfung kleinerer Delikte wie Laden- oder Taschendiebstahl. 23 % sprachen sich zugleich gegen den Einsatz zur Bekämpfung kleinerer Delikte aus. Den geringsten Zuspruch erhielt der mögliche Einsatzzweck der Verhinderung illegaler Migration: Eine relative Mehrheit von 43 % sprach sich zwar dafür aus, 25 % äusserten sich zugleich allerdings unschlüssig und 30 % ablehnend.

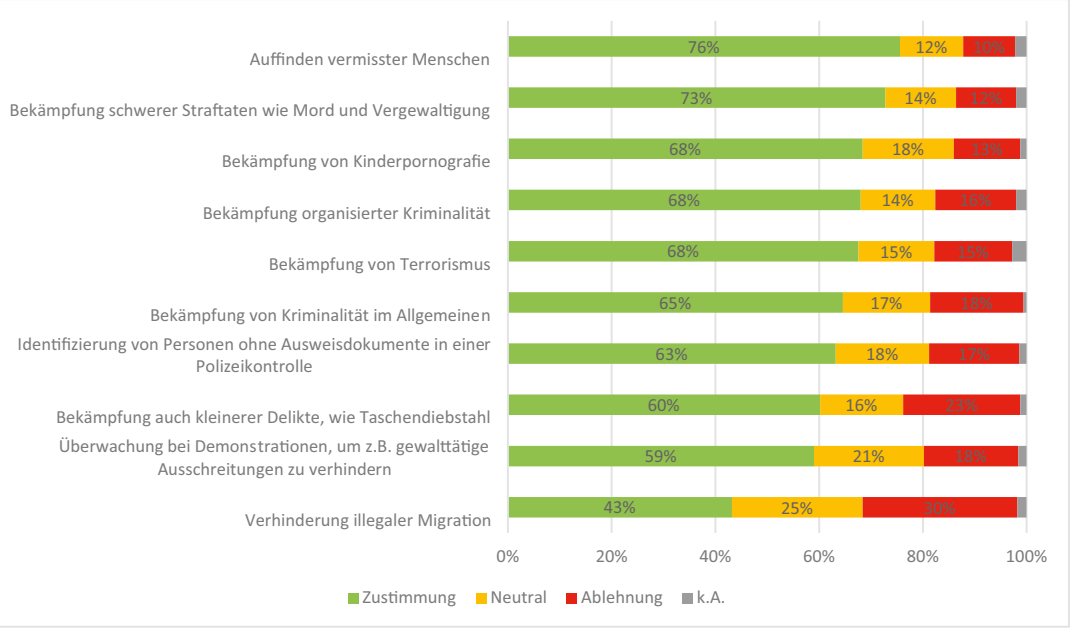

Abbildung 31: Vorstellungen zu möglichen Einsatzzwecken (Mehrfachnennung möglich)

Ausgestaltung des Einsatzes

Zudem wurden alle Teilnehmenden (N=500) dazu befragt, wie der Einsatz von Gesichts- und Spracherkennungstechnologien in der Schweiz ausgestaltet sein sollte (Abbildung 32). Die höchsten Zustimmungswerte erreichten die Gestaltungsvorschläge, dass lediglich autorisiertes Polizeipersonal Zugriff auf die verwendeten Gesichtserkennungssysteme haben sollte und dass ausserdem jeder Zugriff protokolliert werden sollte (81 %) sowie der Vorschlag, dass transparent über den Einsatz kommuniziert werden sollte (80 %). 73 % stimmten der Aussage zu, dass der Einsatz stets auf einer möglichst konkreten gesetzlichen Grundlage erfolgen sollte, und 72 % der Aussage, dass der Einsatz von unabhängigen Experten begleitet und regelmässig evaluiert werden sollte. 60 % befürworteten die Aussage,

dass ein Einsatz nur nach richterlicher Genehmigung erfolgen sollte. Dass der Polizei grosse Freiheiten beim Technologieeinsatz überlassen werden sollten, wurde von einer relativen Mehrheit von 46 % abgelehnt und von 31 % befürwortet.

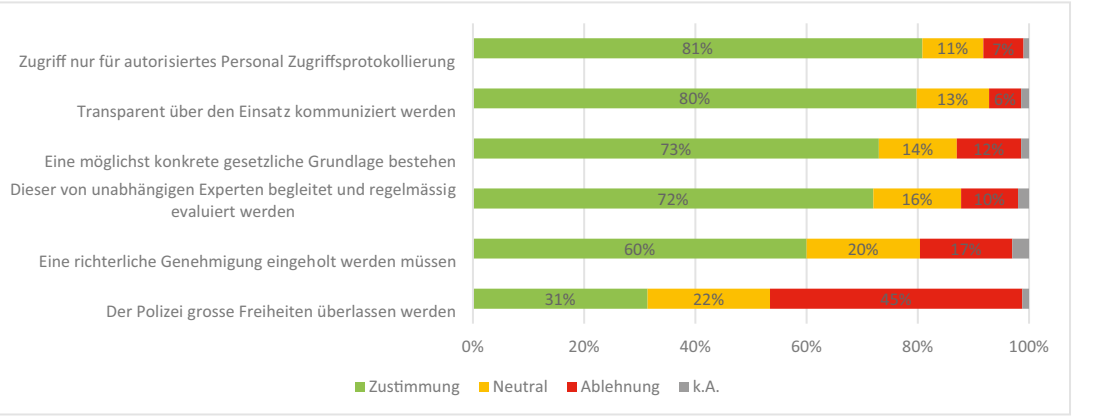

Abbildung 32: Ausgestaltung des Einsatzes von Gesichts- und Spracherkennung (Mehrfachnennung möglich)

Bewertung der Möglichkeit der Emotionsanalyse

Zuletzt wurden alle Teilnehmenden (N=500) gefragt, wie sie zum Einsatz von Gesichts- und Spracherkennungstechnologien zur Emotionserkennung stehen (Abbildung 33). 65 % der Befragten stimmten der Aussage zu, dass sie besorgt seien, dass die Software nicht korrekt funktioniert und dadurch unschuldige Personen ins Visier der Fahnder geraten könnten. 62 % halten Emotionen und Gefühle für zu komplex, als dass sie von einer Software korrekt erkannt werden könnten. Zugleich stimmten 59 % der Aussage zu, dass sie besorgt wären, dass Strafverfolgungsbehörden die Technologie anlasslos einsetzen könnten. Eine relative Mehrheit der Befragten stimmte ausserdem den Aussagen zu, dass sie sich im Falle eines solchen Technologieeinsatzes in jeder Polizeikontrolle sehr unwohl fühlen würde (49 %). Ungefähr ebenso viele Teilnehmende stimmten allerdings der Aussage zu, dass der Einsatz für sie in Ordnung wäre, weil sie nichts zu verbergen hätten (49 %). Eine knappe relative Mehrheit von 38 % lehnte die Aussage ab, dass der Einsatz für die Sicherheit der Gesellschaft förderlich wäre, während 38 % diese bejahten. Dass sie gegen den Technologieeinsatz demonstrieren würden, gaben 21 % der Befragten an. Eine knappe Mehrheit der Befragten (51 %) sprach sich jedoch gegen diese Form des Widerstands aus.

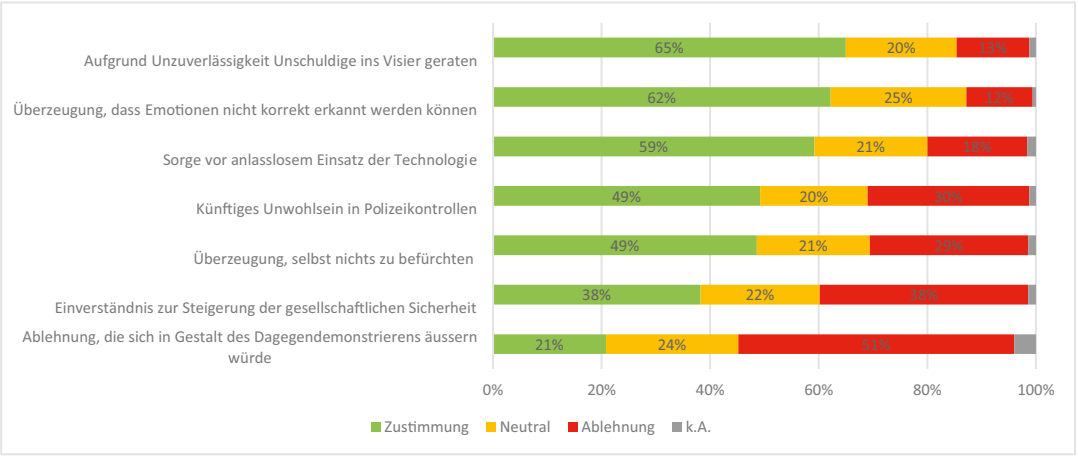

Abbildung 33: Bewertung der Möglichkeit der Emotionsanalyse durch die Polizei (Mehrfachnennung möglich)

5.2.3. Authentifizierung via Stimme bei Telefonbanking

Bedenken

Alle Befragten (N=500) wurden zunächst dazu befragt, wie sie den Einsatz von automatischer Stimmerkennungstechnologie zur Identitätsfeststellung von Anrufern beim Telefonbanking bewerten (Abbildung 34). Eine relative Mehrheit von 37 % gab an, grosse Bedenken zu haben. 33 % der Befragten äusserten, keine klare Meinung zum Thema zu haben, und 24 % der Befragten gaben an, keine Bedenken zu haben.

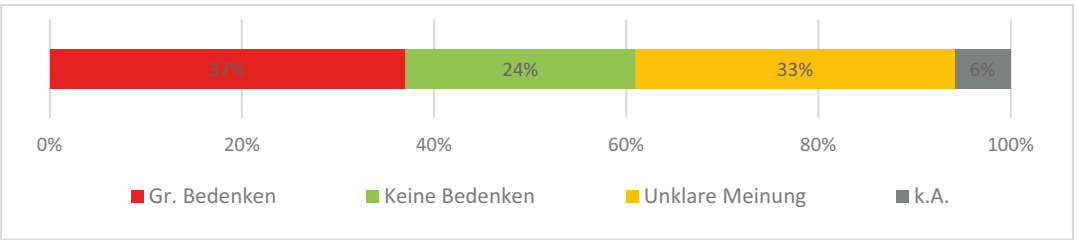

Abbildung 34: Bedenken beim Einsatz zur Authentifizierung beim Tele-Banking

Gründe für Bedenken

Personen, die grosse Bedenken hins. des Einsatzes äusserten (n=185) (Abbildung 35), empfinden es zum Grossteil als seltsam, dass ihre Stimme für derartige Zwecke verwendet wird (89 %). Zugleich stimmten 87 % der Befragten der Aussage zu, dass die automatisierte Stimmerkennung nicht notwendig sei, weil klassische Authentifizierungsverfahren (z.B. Name, Wohnort, Geburtsdatum oder PIN) ausreichend seien. Damit zusammenhängend äusserten 86 % der Befragten Zweifel am fehlerfreien Funktionieren der Stimmerkennung (z.B. aufgrund geringer Sprachqualität bei schlechtem Telefonempfang). Damit zusammenhängend äusserten 76 % der Befragten die Befürchtung, dass jemand ihre Stimme imitieren

könnte. Schliesslich stimmten 69 % der Befragten der Aussage zu, dass sie die Nutzung der Stimmanalyse für andere Zwecke befürchten, etwa zur Gewinnung von Informationen, z.B. über die individuelle Stimmung, den Gesundheitszustand oder Aufenthaltsort.

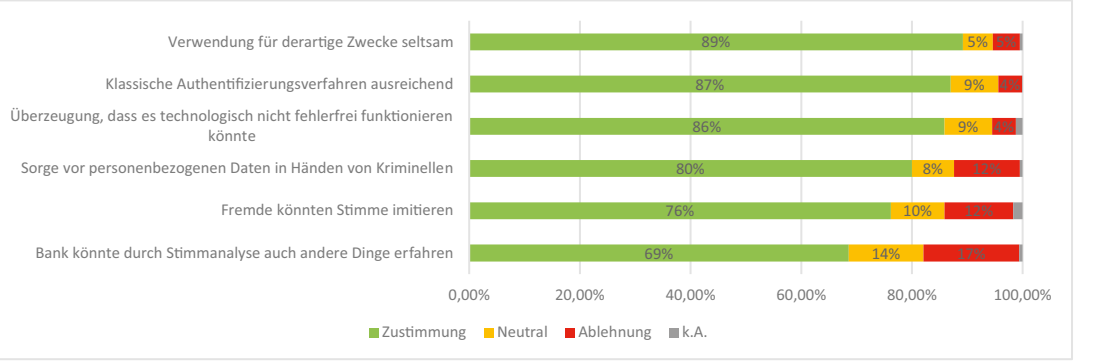

Abbildung 35: Begründungen von Befragten mit grossen Bedenken (Mehrfachnennung möglich)

Befragte, die angegeben hatten, keine klare Meinung zum Thema zu haben (n=166) (Abbildung 36), begründeten diese Meinung am häufigsten damit, sich noch nicht ausreichend mit dem Thema befasst (75 %) zu haben. Unsicherheiten im Hinblick auf die Zuverlässigkeit äusserten 71 % der Befragten. 69 % gaben an, sie hätten Zweifel an der Sicherheit derartige Systeme, insb. im Vergleich zu klassischen Authentifizierungsverfahren (63 %). Verunsicherung darüber, ob ihre personenbezogenen Daten in die Hände von Kriminellen gelangen könnten, gaben 65 % bzw. darüber, ob jemand ihre Stimme imitieren und dadurch Zugang zu ihren Daten erhalten könnte, 52 % der Befragten an. 53 % der Befragten gaben zudem an, unsicher zu sein, ob die Bank durch die Stimmanalyse auch andere Informationen über die Anrufer (z.B. ihre individuelle Stimmung, den Gesundheitszustand oder Aufenthaltsort) in Erfahrung bringen könnte.

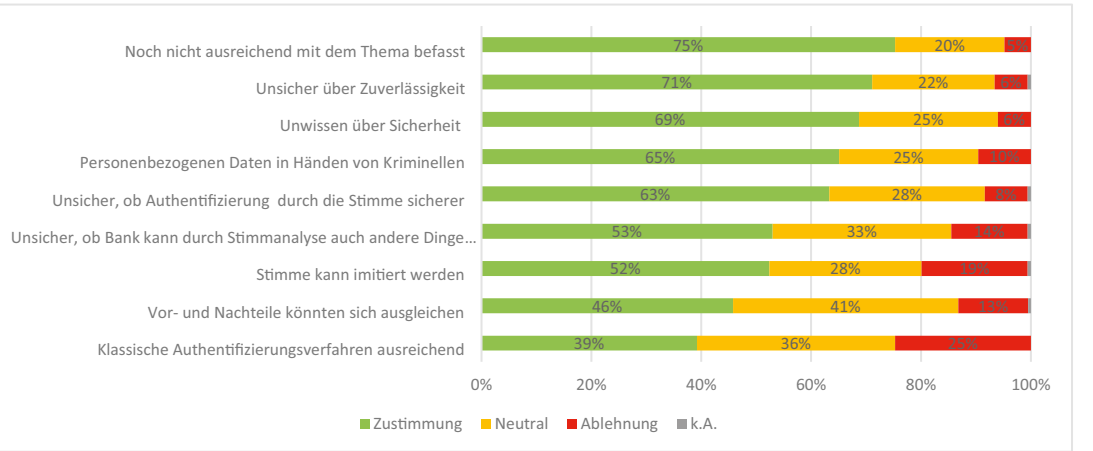

Abbildung 36: Top-5-Begründungen von Befragten, die unschlüssig sind (Mehrfachnennung möglich)

Befürworter des Einsatzes (n=120) (Abbildung 37) sehen in der Authentifizierung via Stimme mehrheitlich eine bequemere und schnellere Möglichkeit, sich gegenüber der Bank zu au-

thentifizieren, als durch die Angabe von Geburtsdatum und Name (80 %). 63 % der Befragten stimmten der Aussage zu, dass die Authentifizierung mittels Stimme sicherer sei als durch die Angabe persönlicher Daten. Zudem gaben 36 % der Befragten an, die Authentifizierung mittels Stimme zu befürworten, obwohl sie selbst keine Telefonbankingdienste verwenden.

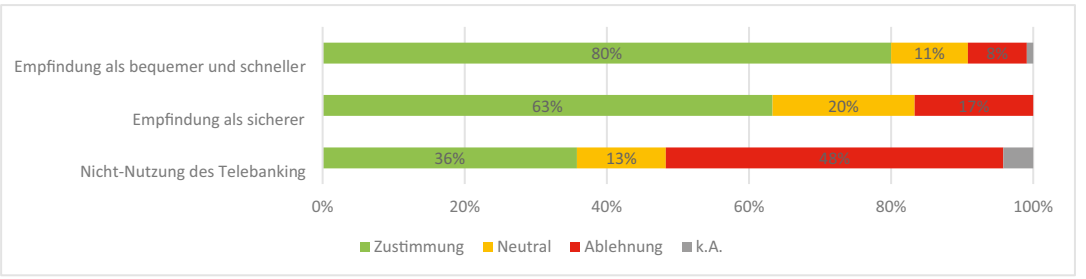

Abbildung 37: Begründungen von Befragten ohne Bedenken (Mehrfachnennung möglich)

Ausgestaltung des Einsatzes

Auf die an alle Teilnehmende (N=500) gerichtete Frage, wie Banken die Stimmerkennung zur Anruferauthentifizierung einsetzen sollten (Abbildung 38), gaben 73 % der Befragten an, dass ein Hinweis vor jedem Einsatz erfolgen sollte. 73 % wünschen sich, dass die Stimmanalyse ausschliesslich zur Stimmauthentifizierung und nicht zur Ableitung weiterer Informationen verwendet wird. 70 % der Befragten wünschen sich, dass die Authentifizierung mittels Stimme keine Pflicht sein darf und eine Widerspruchsmöglichkeit bestehen sollte. 69 % der Befragten sprachen sich dafür aus, dass der Einsatz nur unter der Aufsicht einer unabhängigen Instanz (etwa einer Datenschutzaufsichtsbehörde) erfolgen sollte. 67 % votierten dafür, dass zusätzlich zur Stimmerkennung zusätzliche Anmeldedaten wie ein Passwort abgefragt werden sollten. 56 % der Befragten sprachen sich ausserdem dafür aus, dass die zur Authentifizierung verwendete Stimme nicht als Tondatei gespeichert werden sollte, sondern abstrakt in Form eines Rechenwerts (Hash-Wert). Eine Weiterverwendung der erhobenen Daten seitens der Banken für andere Zwecke wurde seitens einer relativen Mehrheit von 44 % der Befragten abgelehnt, während sich 26 % dafür aussprachen.

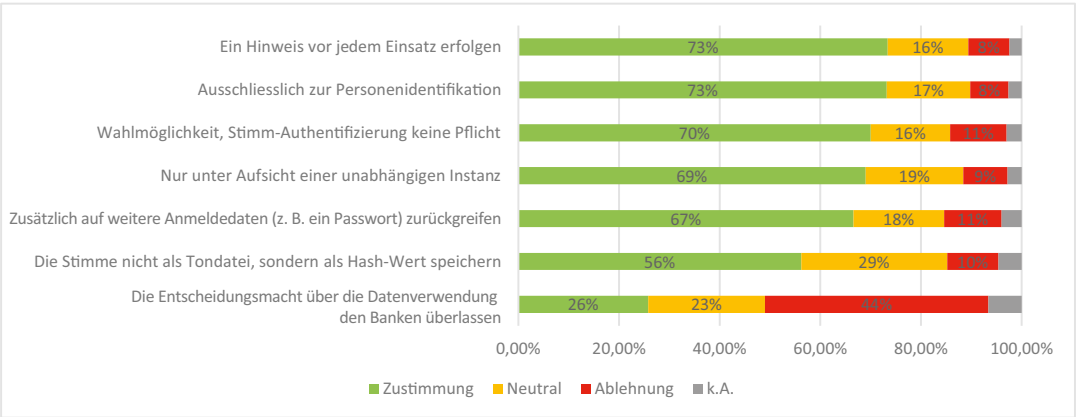

Abbildung 38: Ausgestaltung des Einsatzes von Gesichts- und Spracherkennung (Mehrfachnennung möglich)

5.2.4. Gewaltprävention und -aufklärung in Sportstadien

Bedenken

Eine knappe absolute Mehrheit der Befragten (50 %) sprach sich für den Einsatz der Gesichtserkennung aus (Abbildung 39). Dabei sprachen sich deutlich mehr ältere Befragte (55–75 Jahre) (64 %) für einen Einsatz aus als Befragte zwischen 16 und 34 Jahren (38 %). Einen Unterschied zwischen Befragten aus ländlichen oder städtischen Regionen und zwischen Männern und Frauen gab es hingegen nicht.

27 % der Befragten gaben an, keine klare Meinung zum Thema zu haben. Für ein Verbot bis auf Weiteres sprachen sich 19 % aus. Hier liegt der Anteil der jüngsten Alterskohorte (25 %) deutlich höher über dem der ältesten Alterskohorte (9 %). Ausserdem zeigt sich, dass deutschsprachige Befragte (25 %) eher zu einem Verbot tendieren als französisch- (10 %) und italienischsprachige Befragte (13 %).

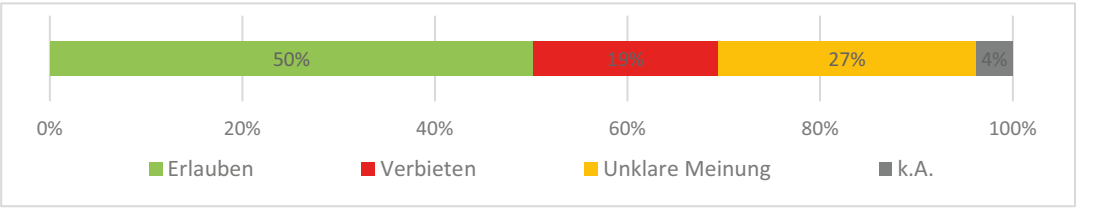

Abbildung 39: Meinungen zum Einsatz in Sportstadien

Gründe für Erlaubnis oder Verbot

Im Anschluss wurden die Teilnehmenden nach den Gründen für ihre jeweilige Meinung befragt.

Befürworter (n=251) (Abbildung 40) stimmten am häufigsten der Aussage zu, dass durch den Einsatz von Gesichtserkennung die Sicherheit in Stadien erhöht werde (94 %) und unerwünschte Personen besser/schneller erkannt werden könnten als durch Menschen (90 %). 83 % stimmten der Aussage zu, dass alle, die sich angemessen verhalten, auch nichts zu befürchten hätten, und ungefähr ebenso viele Befragte (83 %) teilten die Ansicht, dass bereits das Vorhandensein der Gesichtserkennungstechnologie eine abschreckende Wirkung auf Straftäter habe könne. Zudem stimmten 81 % der Befragten darin überein, dass Besucher freiwillig in ein Stadion gehen und daher akzeptieren müssten, dort überwacht zu werden. Dass Menschen in stärkerem Masse zu Diskriminierung neigten als ein Computer und durch Gesichtserkennungstechnologie Diskriminierung reduziert werden könnte, waren für 64 % der Befragten ein Grund für ihre Zustimmung. 57 % der Befragten gaben an, dass sie den Stadionbetreibern vertrauen, die Technologie verantwortungsvoll einzusetzen. Die Möglichkeit der Weiternutzung der Daten bspw. für Werbezwecke wurde von ebenso vielen Befragten abgelehnt und unterstützt (jeweils 32 %), während eine relative Mehrheit von 34 % keine klare Meinung dazu äusserte.

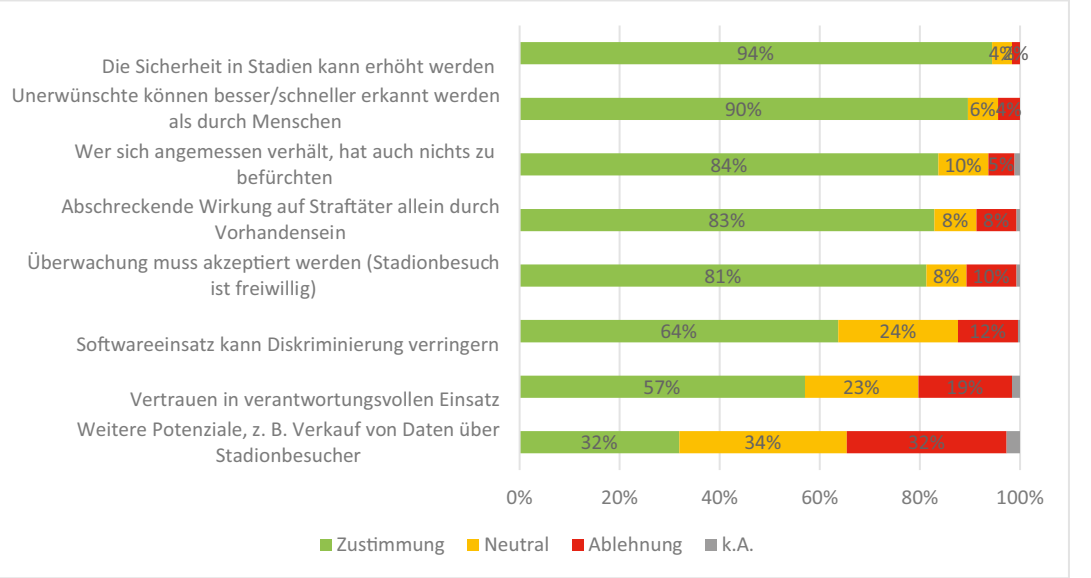

Abbildung 40: Gründe von Befürwortern des Einsatzes (Mehrfachnennung möglich)

Die meisten Befragten mit unklarer Meinung (72 %, n=134) (Abbildung 41) begründeten dies damit, dass sie sich noch nicht ausreichend mit dem Thema beschäftigt hätten. Ausserdem gab die absolute Mehrheit dieser Befragten an, dass es auf die konkreten Einsatzzwecke (59 %) bzw. die konkrete Gestaltung des jeweiligen Einsatzes ankomme (55 %), dass Unsicherheit darüber bestehe, ob Stadionbetreiber die Technologie verantwortungsvoll einsetzen würden (55 %) und ob damit die Sicherheit in Stadien erhöht werden könnte.

Befragte, die sich für ein Moratorium aussprachen (n=96) (Abbildung 42), führten als Gründe für ihre Ablehnung am häufigsten an, dass es sich bei der Gesichtserkennung im Stadion um eine anlasslose Massenüberwachung aller Stadiongäste handle (74 %) und es alternative Methoden der Verhinderung des Einlasses von Gewalttätern gäbe, die ebenso wirkungsvoll seien (74 %). 68 % befürchteten, dass die im Stadion erhobenen Daten für andere Zwecke, wie unerwünschte personalisierte Werbung, missbraucht werden könnten. 62 % stimmten der Aussage zu, dass die Gesichtserkennung keine zuverlässigen Ergebnisse liefere und daher auch unschuldige Menschen ins Visier geraten könnten. Damit zusammenhängend befürchteten 52 %, dass auch sie verdächtigt werden könnten. 58 % begründeten ihre Ablehnung damit, dass sie sich generell gegen jegliche Form von Überwachung aussprechen.

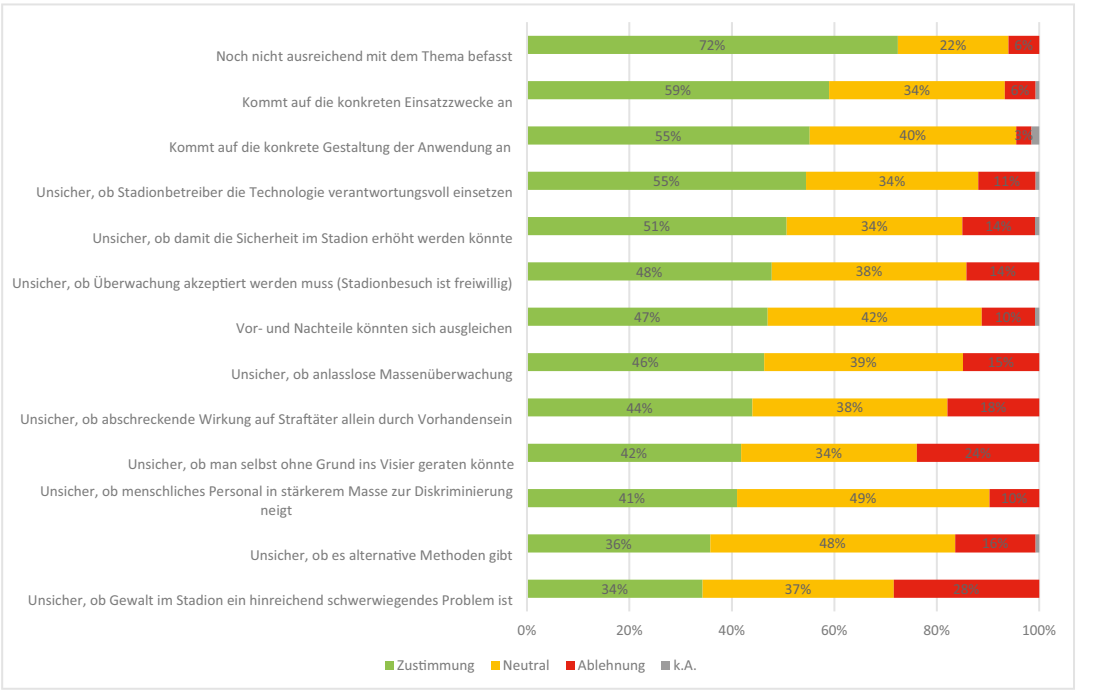

Abbildung 41: Gründe von Befragten, die unsicher über Einsatz sind (Mehrfachnennung möglich)

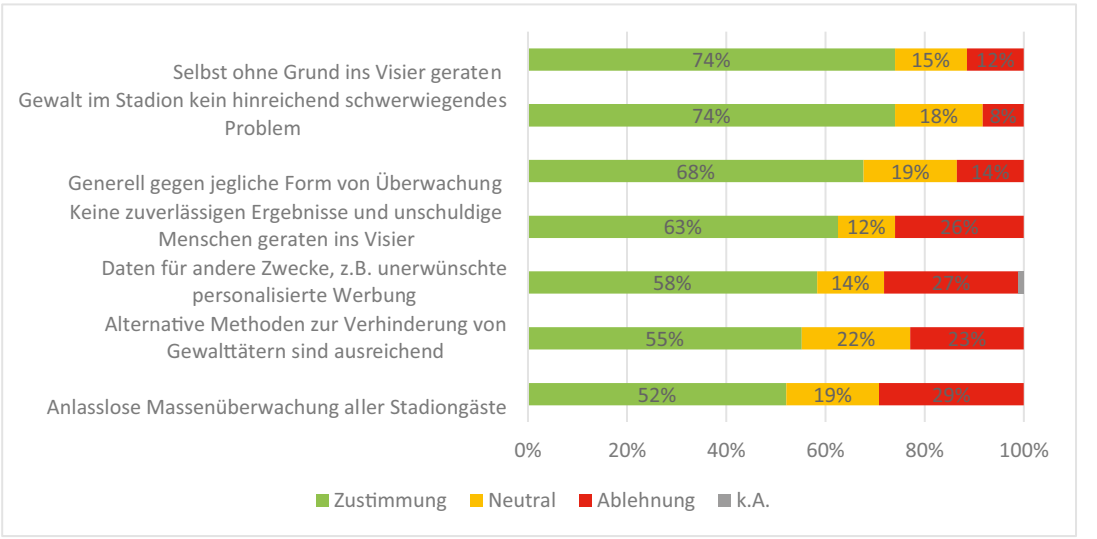

Abbildung 42: Gründe von Einsatzgegnern (Mehrfachnennung möglich)

Schutzmassnahmen

Zudem wurden die Personen, die sich gegen den Einsatz aussprachen (n=96) (Abbildung 43), danach gefragt, welche Schutzmassnahmen sie treffen würden, um der Überwachung bzw. Erkennung im Stadion zu entgehen. Demnach würden 69 % Bekannte, Freun-

de oder Familienmitglieder warnen und 59 % das Stadion selbst nicht mehr betreten. Die Hälfte aller Befragten (50 %) würde sich bei Datenschutzaufsichtsbehörden beschweren und knappe 42 % gegen den Technologieeinsatz politisch aktiv werden bzw. demonstrieren. Eine Maskierung als Schutzmassnahme kommt für 38 % der Befragten infrage, während etwa genauso (37 %) viele diese Vorgehensweise ablehnten.

Der Blick auf verschiedene Altersgruppen zeigt, dass 46 % der 16–34-Jährigen, aber nur 15 % der 55–75-Jährigen sich maskieren würden. 29 % der Befragten gaben an, aktiv gegen die Technologie (etwa das Zukleben oder Zerstören von Kameras) vorgehen zu wollen. Allerdings lehnte eine relative Mehrheit von 47 % derartige Massnahmen ab.

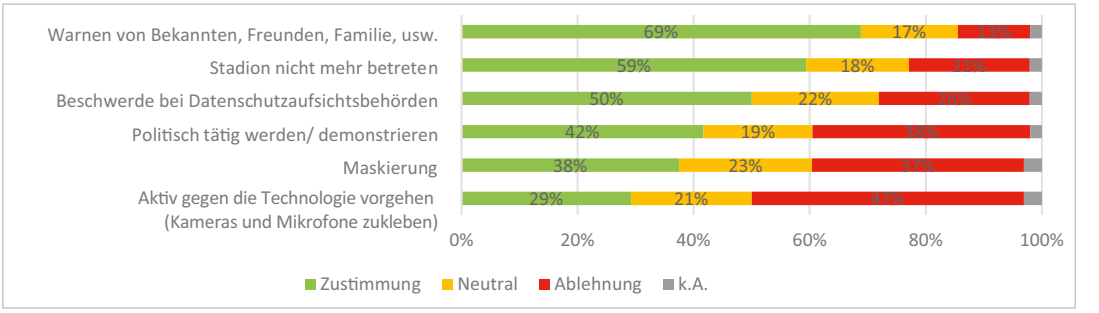

Abbildung 43: Schutzmassnahmen gegen Stadionüberwachung von Gegnern (Mehrfachnennung möglich)

Ausgestaltung des Einsatzes

Die Teilnehmenden (N=500) wurden dazu befragt, wie der Einsatz von Gesichtserkennungstechnologien in Stadien ausgestaltet sein müsste, um akzeptabel zu sein (Abbildung 44). Hier zeigt sich, dass knapp die Hälfte der Befragten überhaupt einen Einsatz befürwortet und 51 % Gesichtserkennung nur in sehr seltenen Ausnahmefällen und zeitlich befristet eingesetzt sehen möchten. Dies zeigt sich auch darin, dass 49 % der Befragten sich wünschen, dass Alternativen zur Gesichtserkennung zur Verfügung stehen sollten.

Ansonsten erreicht der Wunsch, dass der Technologieeinsatz transparent (Wer betreibt den Dienst? Was sind die Zwecke? Wo/wie lange werden die Daten gespeichert?) erfolgen sollte, die höchste Zustimmung (75 %). Eine grosse Mehrheit von 71 % sprach sich zudem dafür aus, dass Kameras überall im Stadion eingesetzt werden sollten, um Gewalttaten auch während des Spiels einer Person zuordnen zu können. Ungefähr ebenso viele Befragte (69%) sprachen sich dafür aus, dass der Einsatz von unabhängigen Experten begleitet und regelmässig evaluiert werden sollte und dass die Videoaufnahmen nicht bzw. nur für eine streng limitierte Dauer gespeichert werden dürften (69 %). Die Möglichkeit, der Gesichtserkennung widersprechen zu können, ohne dass dies negative Folgen wie eine Einlassverweigerung nach sich zieht, wünschte sich eine relative Mehrheit von 40,4 %, während 29,4 % sich dagegen aussprachen. Eine knappe Mehrheit stimmte ausserdem den Aussagen zu, dass den Stadionbetreibern beim Technologieeinsatz grosse Freiheiten überlassen werden sollten (38 %) und dass die Gesichtserkennung nicht an Minderjährigen durchgeführt werden sollte (35,4 %).

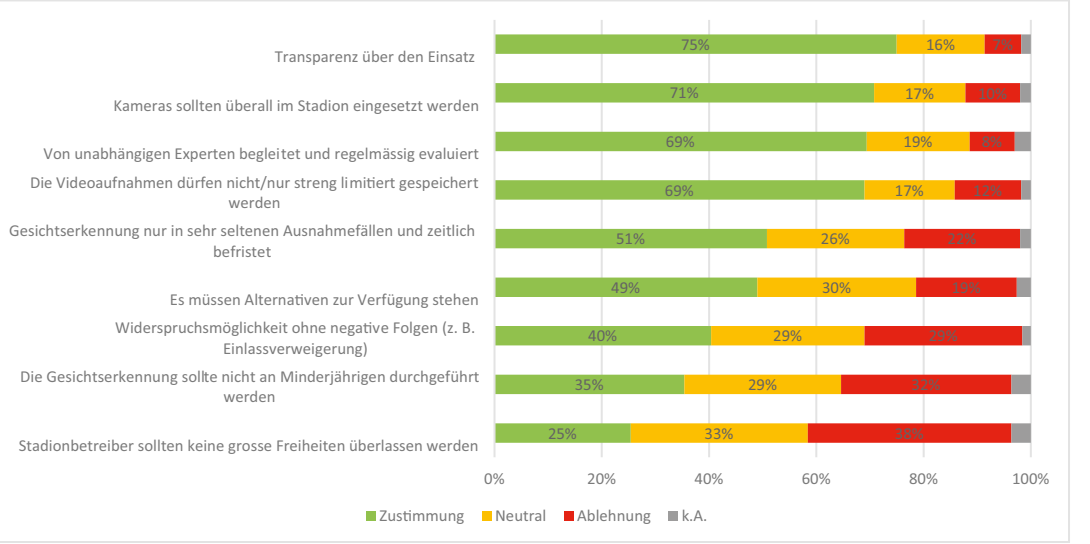

Abbildung 44: Meinungen zur Ausgestaltung des Einsatzes (Mehrfachnennung möglich)

Weitere Einsatzzwecke

Gefragt wurde (Abbildung 45) zudem, welche weiteren Einsatzzwecke (ausser zur Verhinderung von Stadiongewalt) sich die Befragten (N=500) vorstellen könnten. 78 % sprachen sich dafür aus, dass die Gesichtserkennung in Sportstadien auch für polizeiliche Zwecke, wie die Suche nach Kriminellen oder Vermissten, eingesetzt werden dürfen sollte. 67 % befürworteten den Einsatz zur Erkennung von sonstigem unerwünschtem Verhalten im Stadion (als Beispiel war die in Italien erprobte Erkennung rassistischer Sprachchöre genannt, vgl. Abschnitt 3.6.1). Mit grosser Mehrheit (63 %) wurde indes die Möglichkeit des Verkaufs von im Stadion mittels Gesichtserkennung erhobenen Daten an Werbetreibende abgelehnt.

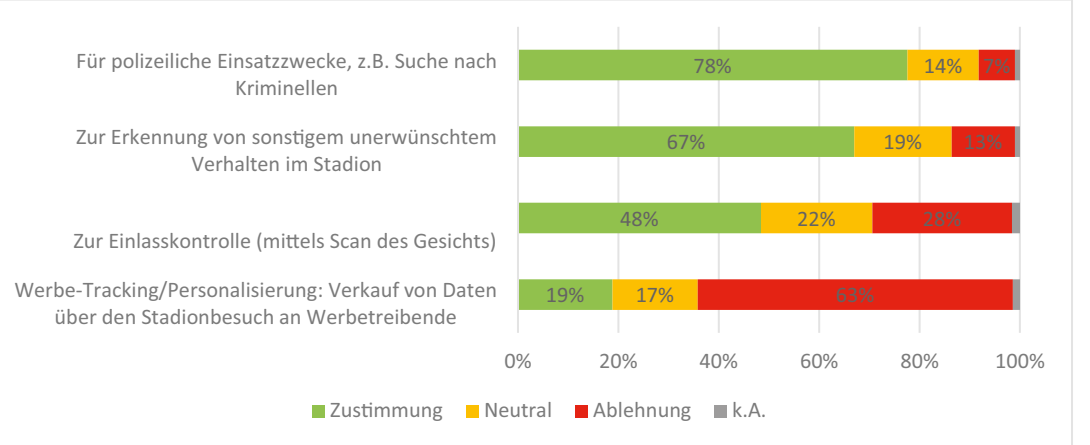

Abbildung 45: Meinungen zu weiteren Einsatzzwecken (Mehrfachnennung möglich)

Alternativen zur Gesichtserkennung in Sportstadien

Abschliessend wurden alle Teilnehmenden (N=500) (Abbildung 46) gefragt, welche Alternativen zur Gesichtserkennung sie sich in Stadien vorstellen könnten. Die höchste Zustimmung (75 %) erhielt hierbei der Vorschlag, generell mehr Sicherheitspersonal im Stadion einzusetzen. 68 % stimmten der Aussage zu, dass die Erkennung von bekannten Gewalttätern an den Eingängen mittels geschulten Sicherheitspersonals erfolgen könnte, und für 66 % kam die Ausstellung personalisierter Tickets in Betracht. Die Möglichkeit, an die Selbstdisziplin der Stadionbesucher zu appellieren, wurde von einer relativen Mehrheit von 45 % als Möglichkeit gesehen, während 27 % sich dagegen aussprachen.

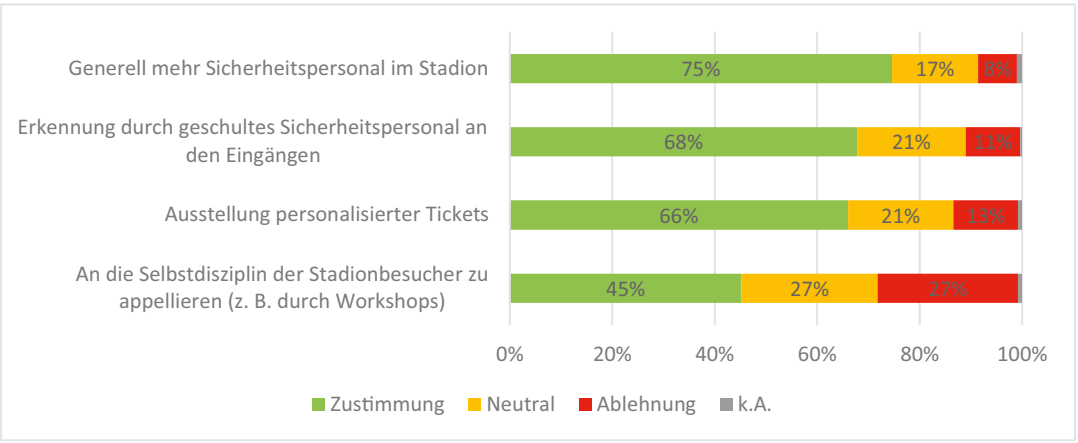

Abbildung 46: Meinungen zu Alternativen der Gesichtserkennung (Mehrfachnennung möglich)

5.2.5. Erkennung physischer Krankheiten

Bedenken

Die knappe Mehrheit der Befragten (37 %) gab an, keine klare Meinung zum Thema zu haben (Abbildung 47). Ungefähr gleich viele Befragte (37 %) gaben an, grosse Bedenken zu haben und 23 % stimmten der Aussage zu, dass sie keine Bedenken haben.

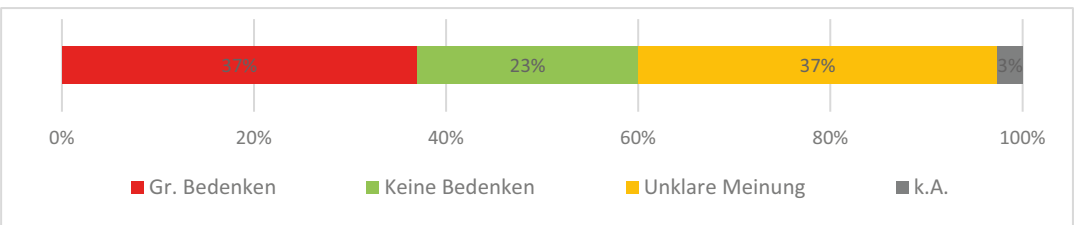

Abbildung 47: Bedenken über den Einsatz von Stimm- und Gesichtsanalyse zur Erkennung physischer Krankheiten

Gründe für Bedenken

Unter jenen, die grosse Bedenken haben (n=185) (Abbildung 48), wurde die Sorge, dass Krankenkassen Zugriff auf die Sprachaufnahmen des Smartphones erhalten, um mögliche Preisnachlässe zu geben, am häufigsten geäussert (86 %). Fast ebenso viele Befragte gaben ausserdem an, dass sie die Erstellung von Diagnosen ohne ihre Einwilligung befürchten (85 %) und die Zuverlässigkeit der Diagnose bezweifeln (84 %). In diesem Zusammenhang stimmten 81 % der Befragten der Aussage zu, dass ihre Bedenken darauf basieren, dass die Frage der Verantwortlichkeit für eine solche Fehldiagnose nicht geklärt sei. 80 % gaben an, dass ihrer Meinung nach keine ausreichenden organisatorischen Vorkehrungen getroffen werden könnten, um sensible Gesundheitsdaten ausreichend gut vor Missbrauch zu schützen, und 79 % stimmten der Aussage zu, dass die Verwendung von Gesichts- und Spracherkennung im medizinischen Bereich einen zu starken Eingriff in das Recht auf Privatsphäre darstelle. Eine generelle Ablehnung des Einsatzes von KI im Medizinbereich war für 58 % der Befragten ein ausschlaggebender Grund für ihre Bedenken. Zugleich verneinten 23 % diese Aussage.

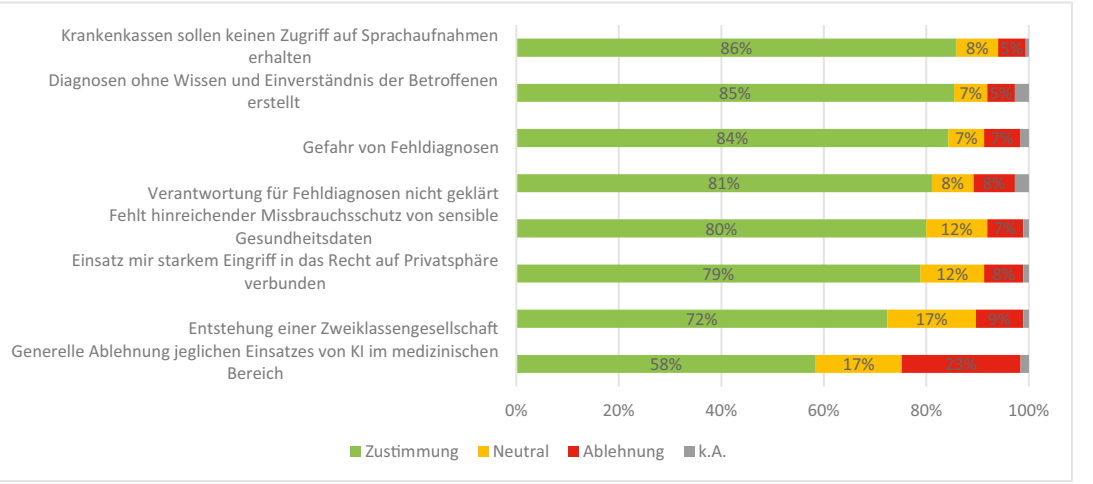

Abbildung 48: Gründe für Bedenken von Befragten mit grossen Bedenken (Mehrfachnennung möglich)

Befragte ohne klare Meinung zum Thema (N=187) (Abbildung 49) gaben als Grund hierfür an, dass sie dies nicht pauschal bewerten könnten, weil es vom konkreten Einsatzzweck abhänge (74 %), sie sich noch nicht ausreichend mit dem Thema befasst hätten (73 %) und es ausserdem auf die konkrete Gestaltung des Einsatzes ankomme (66 %). Zudem gaben sie an, unsicher zu sein, ob die Technologien aktuell zuverlässig genug funktionierten (56 %). Einen zu starken Eingriff in das Recht auf Privatsphäre erkannten 53 %. Weitere Begründungen für die unklare Meinung können Abbildung 49 entnommen werden.

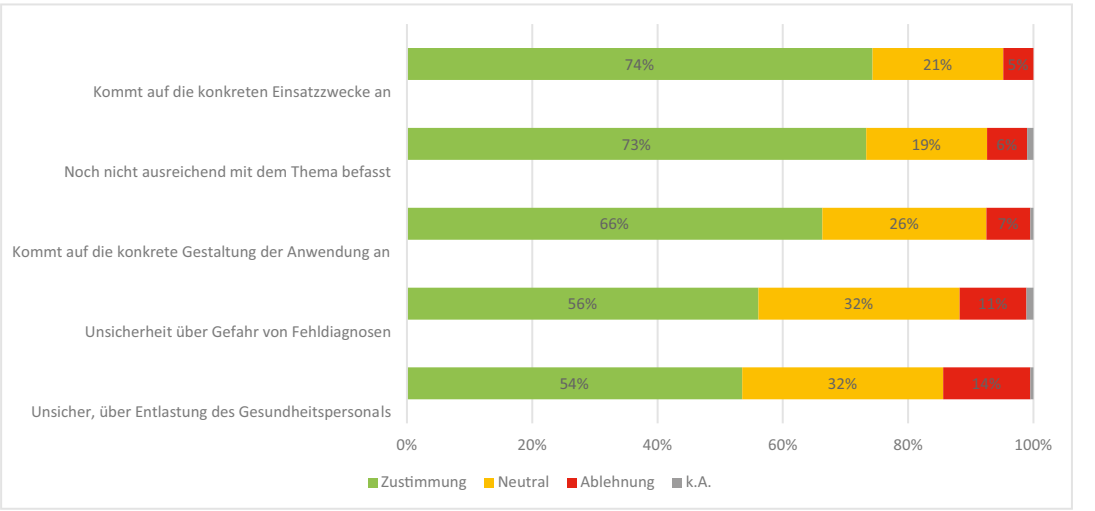

Abbildung 49: Top-5-Gründe für Bedenken von Befragten, die unschlüssig sind (Mehrfachnennung möglich)

Befragte ohne Bedenken (N=115) (Abbildung 50) nannten am häufigsten als Begründung, dass durch den Technologieeinsatz Menschenleben gerettet bzw. gesundheitliche Schäden abgewendet werden könnten (84 %). 70 % gaben ausserdem an, dass die Systeme lediglich als Unterstützung für den Arzt dienen und keine selbstständigen Diagnosen ausgeben würden. Grössere Zustimmungen erhielten ausserdem die Aussagen, dass der Technologieeinsatz zur Eindämmung der Covid-19-Pandemie beitragen könnte (64 %), zur Kostenersparnis im Gesundheitswesen beitragen könnte (64 %), zur Erhöhung medizinischer Kapazitäten beitragen könnte (63 %), durch entsprechende Systeme spezifische Gesundheitsdaten für Forschungszwecke im Dienste der Verbesserung der Gesundheitsversorgung gesammelt werden könnten (61 %) und die Früherkennung via Apps mitunter selbst durchgeführt werden könnte (59 %). Der Aussage, dass die Systeme genauere Diagnosen erlauben würden als durch das Gesundheitspersonal, stimmten weniger Befragte (44 %) zu, während 35 % die Aussage ablehnten. Die höchste Ablehnung (37 %) erhielt die Aussage, dass durch die Freigabe von Smartphone-Sprachaufnahmen eine Prämienverbilligung in der Krankenversicherung erfolgen könnte. Zugestimmt wurde dieser Aussage lediglich von 31 % der Befragten.

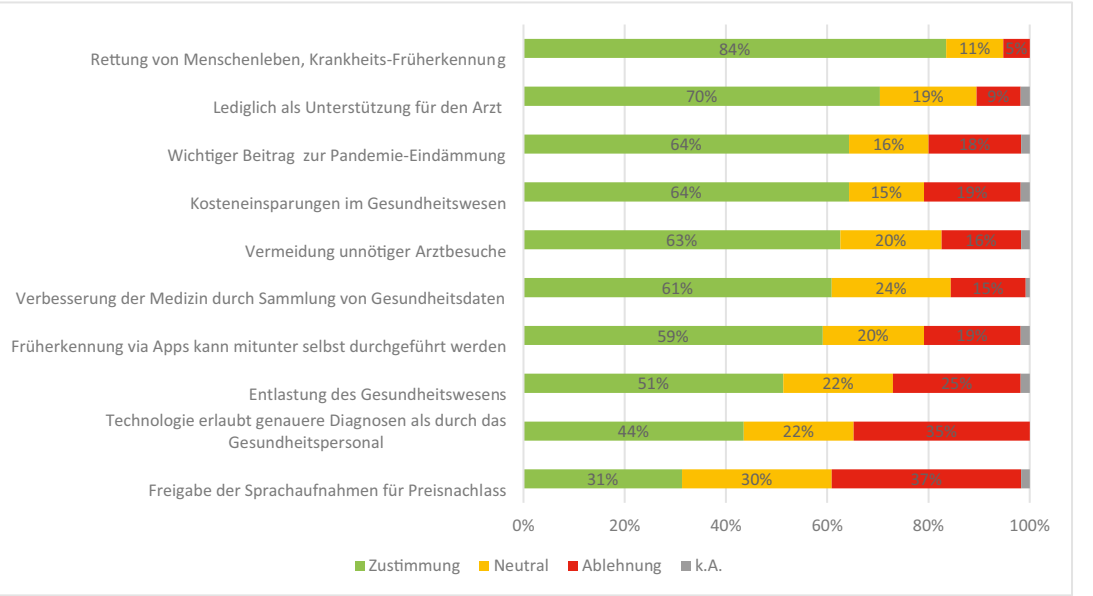

Abbildung 50: Gründe für Bedenken von Befragten ohne Bedenken (Mehrfachnennung möglich)

Ausgestaltung des Einsatzes

Abschliessend wurden die Teilnehmer (N=500) (Abbildung 51) gefragt, wie der Technologieeinsatz ausgestaltet sein sollte, um akzeptabel zu sein. Hier befürworteten 78 % der Befragten, dass erhobene Daten nicht an andere Stellen (wie Krankenkassen oder Arbeitgeber) weitergegeben werden dürften. Sehr hohe Zustimmungswerte erhielten ausserdem die Forderungen, dass transparent über den Technologieeinsatz informiert werden sollte (77 %), die Systeme nur als Unterstützung der ärztlichen Behandlung dienen und diese keinesfalls ersetzen sollten (76 %), es eine Widerspruchsmöglichkeit zum Technologieeinsatz geben sollte, ohne dass daraus Konsequenzen befürchtet werden müssten (74 %), und dass die Systeme nur dann eingesetzt werden sollten, wenn es verlässliche Erkennungsraten gibt (74 %). Dass die Systeme erst nach entsprechender Zertifizierung in Betrieb genommen (73 %) und auch während des Betriebs von unabhängigen Experten evaluiert werden sollten (71 %), wurde ebenfalls von einer grossen Mehrheit befürwortet. Etwas mehr als die Hälfte der Befragten (57 %) befürwortete ausserdem den Vorschlag, dass erhobene Daten nicht gespeichert werden dürften. Eine Mehrheit der Befragten (58 %) äusserte sich skeptisch gegenüber der Weiterverwendung der Daten zu anderen Zwecken. Lediglich die Weiterverwendung von Daten zu Forschungszwecken wurde von einer knappen Mehrheit von 52 % befürwortet. Ablehnung bei einer relativen Mehrheit der Befragten (34 %) erzeugte der Vorschlag, dass stimm- und gesichtserkennungsbasierte Krankheitserkennungstechnologien als Standarddiagnosemethode eingesetzt werden sollten. Während sich 33 % der Befragten unschlüssig gegenüber diesem Vorschlag äusserten, wurde er von 30 % befürwortet.

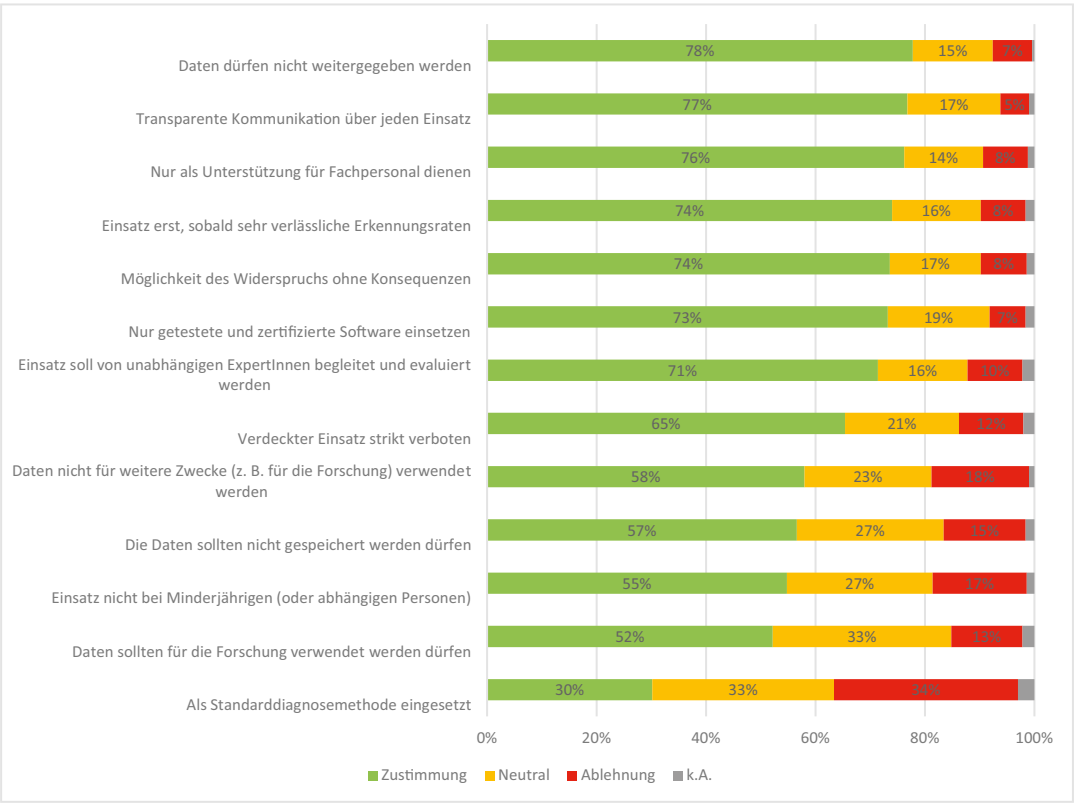

Abbildung 51: Ausgestaltung des Einsatzes von Gesichts- und Spracherkennung (Mehrfachnennung möglich)

5.2.6. Erkennung psychischer Krankheiten mittels Stimm-, Sprach- und Gesichtsanalyse

Bedenken

Die relative Mehrheit der Befragten (38 %) gab an, keine klare Meinung zum Thema zu haben. Grosse Bedenken äusserten 36 % und keine Bedenken 23 % der Befragten (Abbildung 52). Unter jenen Teilnehmern, die äusserten, grosse Bedenken zu haben, befanden sich deutlich mehr deutschsprachige Befragte (43 %) als französischsprachige (23 %) und auch etwas mehr als italienischsprachige Befragte (30 %).

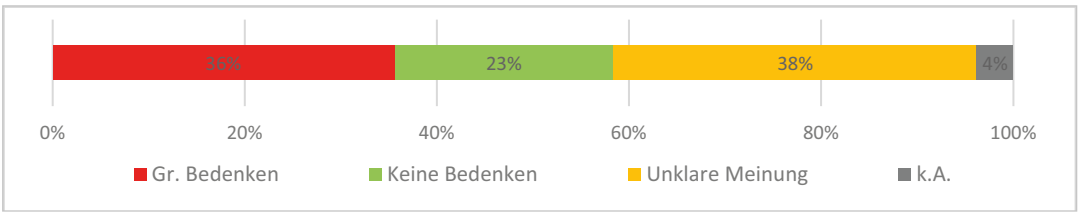

Abbildung 52: Bedenken über den Einsatz zur Erkennung von psychischen Krankheiten

Gründe für Bedenken

Von jenen Befragten, die grosse Bedenken äusserten (n=178) (Abbildung 53), gab eine grosse Mehrheit von 84 % an, dass psychische Krankheiten im Vergleich zu körperlichen Krankheiten nicht immer offensichtlich seien. Jeweils 83 % stimmten den Meinungen zu, dass keine hinreichenden organisatorischen Vorkehrungen getroffen werden könnten, um sensible Gesundheitsdaten ausreichend gut vor Missbrauch zu schützen, und dass die Technologie nicht zuverlässig genug funktioniere und daher keine ärztliche Diagnose ersetzen sollte. 82 % der Befragten gaben an, dass sie eine Diagnose ohne ihr Einverständnis und die Weitergabe der Daten an Krankenkassen befürchten. 81 % der Befragten stimmten der Aussage zu, dass psychisch Erkrankte aufgrund der damit verbundenen Probleme oft nicht in der Lage seien, eine bewusste Einwilligung zu erteilen. 80 % der Befragten stimmten der Aussage zu, dass die Verantwortung im Falle einer Fehldiagnose nicht geklärt sei, und 80 % befürchteten, dass es zu einer Stigmatisierung von Menschen mit psychischen Krankheiten kommen könnte. Eine generelle Ablehnung gegenüber jeglichem Einsatz von KI im medizinischen Bereich äusserte immerhin eine Mehrheit von 69 % dieser Befragten.

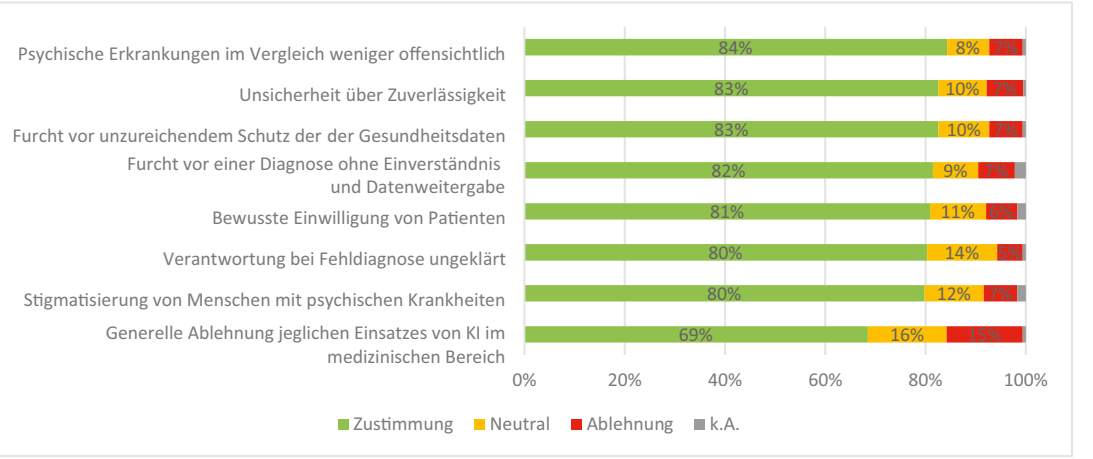

Abbildung 53: Gründe für Bedenken von Befragten mit grossen Bedenken (Mehrfachnennung möglich)

Befragte, die keine klare Meinung zum Thema haben (n=189) (Abbildung 54), gaben als Grund besonders häufig an, sich noch nicht ausreichend mit dem Thema beschäftigt zu haben (74 %). Zudem wurde darauf verwiesen, dass es auf die konkreten Anwendungen ankomme (71 %), und es wurden Zweifel darüber geäussert, wie technisch zuverlässig die Krankheitserkennung erfolge und dass deshalb eine ärztliche Diagnose nicht ersetzt werden sollte (67 %). 66 % gaben an, unsicher zu sein, ob dann eine Diagnosestellung auch ohne ihr Wissen und ihr Einverständnis erfolgen könnte. Verbunden damit gaben 66 % an, dass es auf die Einbindung des Arztes in den Prozess des Technologieeinsatzes ankomme. Unsicherheit äusserten die Befragten zudem darüber, ob es infolge des Technologieeinsatzes zu einer Stigmatisierung von Erkrankten kommen könnte (58 %), die Erhebung von Gesundheitsdaten der verbesserten Gesundheitsversorgung dienlich sein könnte (58 %), eine Früherkennung zum Wohle der Menschen möglich wäre (57 %), das Gesundheitspersonal dadurch entlastet werden könnte (55 %), ein ausreichend guter Schutz der Gesundheits-

daten vor Missbrauch möglich wäre (54 %) oder inwiefern eine individualisierte und damit verbesserte Gesundheitsversorgung möglich würde (53 %).

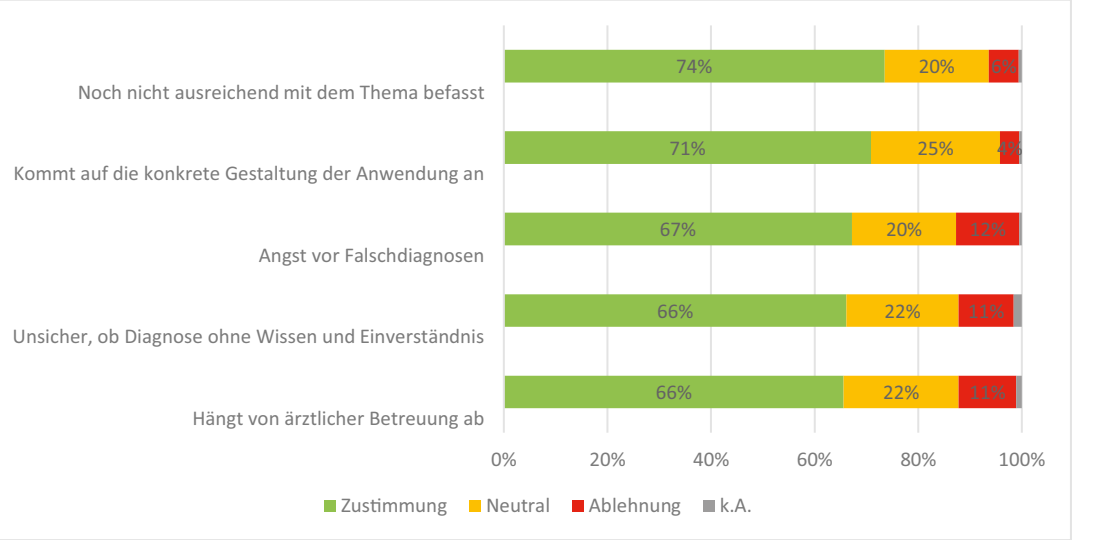

Abbildung 54: Top-5-Gründe für Bedenken von Befragten, die unschlüssig sind (Mehrfachnennung möglich)

Befürworter des Einsatzes von Stimm-, Sprach- und Gesichtserkennungstechnologien zur Erkennung psychischer Krankheiten (n=114) (Abbildung 55) sehen darin v.a. Potenzial für die Früherkennung von psychischen Krankheiten (75 %) und für eine individualisierte und damit bessere Gesundheitsversorgung bei psychischen Krankheiten (75 %). 69 % bzw. 66 % sehen darüber hinaus durch die Nutzung der gesammelten Daten für Forschungszwecke Potenzial für die Verbesserung der Gesundheitsversorgung insgesamt bzw. zur Verbesserung der Gesundheitsversorgung für psychische Krankheiten. Ebenfalls 66 % teilten die Meinung, dass der Einsatz eine sinnvolle Unterstützung der ärztlichen Behandlung sein könne. Kosteneinsparungen im Gesundheitswesen oder eine Entlastung des Gesundheitspersonals waren für jeweils 50 % der Befragten ein relevantes Kriterium für ihre Befürwortung. Den niedrigsten Zustimmungswert erhielt mit 40 % die Möglichkeit zur Freigabe von Sprachaufnahmen für eine Prämienverbilligung bei der Krankenversicherung. Mit 36 % erhielt diese Antwortmöglichkeit auch den höchsten Ablehnungswert.

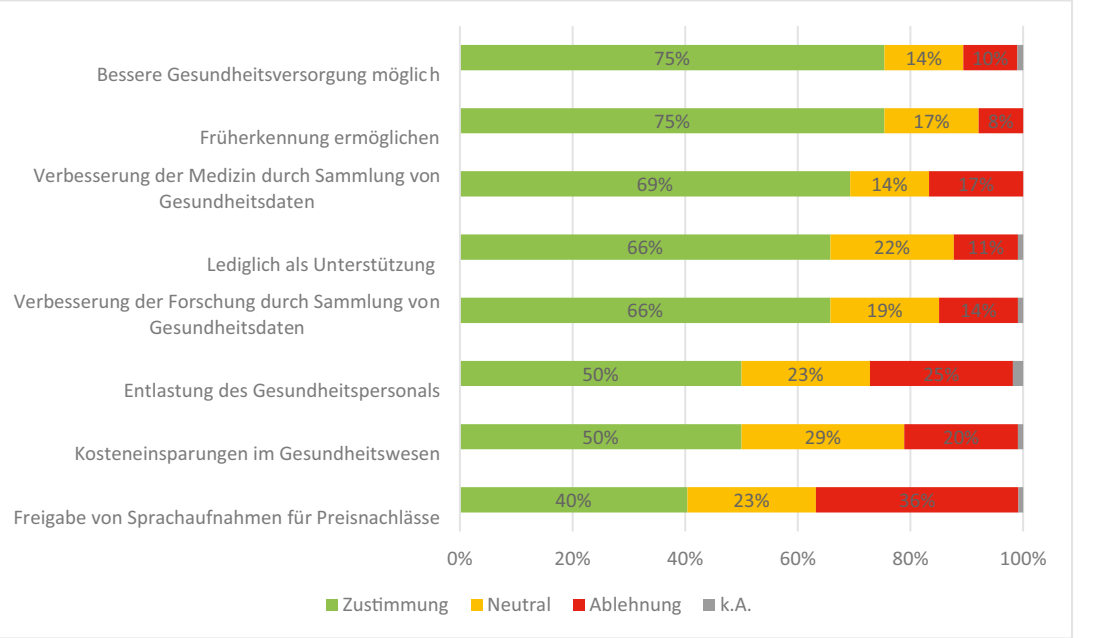

Abbildung 55: Gründe von Befragten ohne Bedenken (Mehrfachnennung möglich)

Ausgestaltung des Einsatzes

Abschliessend wurden alle Teilnehmenden (N=500) befragt, wie Stimm-, Sprach- und Gesichtserkennungstechnologien zur Erkennung psychischer Krankheiten eingesetzt werden sollten (Abbildung 56). Die höchste Zustimmung erreichte der Wunsch, dass die erhobenen Daten nicht an Stellen wie Krankenkassen oder Arbeitgeber weitergegeben werden (76 %). 73 % vertraten die Ansicht, dass derartige Systeme lediglich als unterstützendes Werkzeug für das Fachpersonal dienen und eine ärztliche Diagnose keinesfalls ersetzen sollten. Jeweils 72 % der Befragten sprachen sich zudem dafür aus, dass der Einsatz transparent erfolgen und eine Information über die Datennutzung erfolgen sollte, und dafür, dass ein Widerspruch ohne Konsequenzen zu befürchten möglich sein sollte. Gegen einen verdeckten Einsatz entsprechender Systeme sprachen sich 66 % aus. 70 % der Befragten waren dafür, dass der Einsatz von unabhängigen Experten begleitet und evaluiert werden sollte. Ebenso befürworteten 70 % der Befragten, dass der Einsatz nur unter Aufsicht einer unabhängigen Instanz wie einer Datenschutzaufsichtsbehörde erfolgen sollte. Dass die Technik erst dann eingesetzt wird, sobald verlässliche Erkennungsraten vorliegen, wurde von 70 % der Befragten befürwortet. Etwa ebenso hoch war die Zustimmung (69 %), dass entsprechende Systeme vor einem Einsatz auf Wirksamkeit getestet und zertifiziert werden sollten. Eine Mehrheit der Befragten (59 %) äusserte sich skeptisch gegenüber der Weiterverwendung der Daten zu anderen Zwecken. Lediglich die Weiterverwendung von Daten zu Forschungszwecken wurde von einer knappen Mehrheit von 52 % befürwortet. Eine Speicherung der Daten wurde von 59 % der Befragten abgelehnt. Zudem sprach sich eine knappe relative Mehrheit von 32 % der Befragten dagegen aus, dass die Technologie zu einer Standarddiagnosemethode wird.

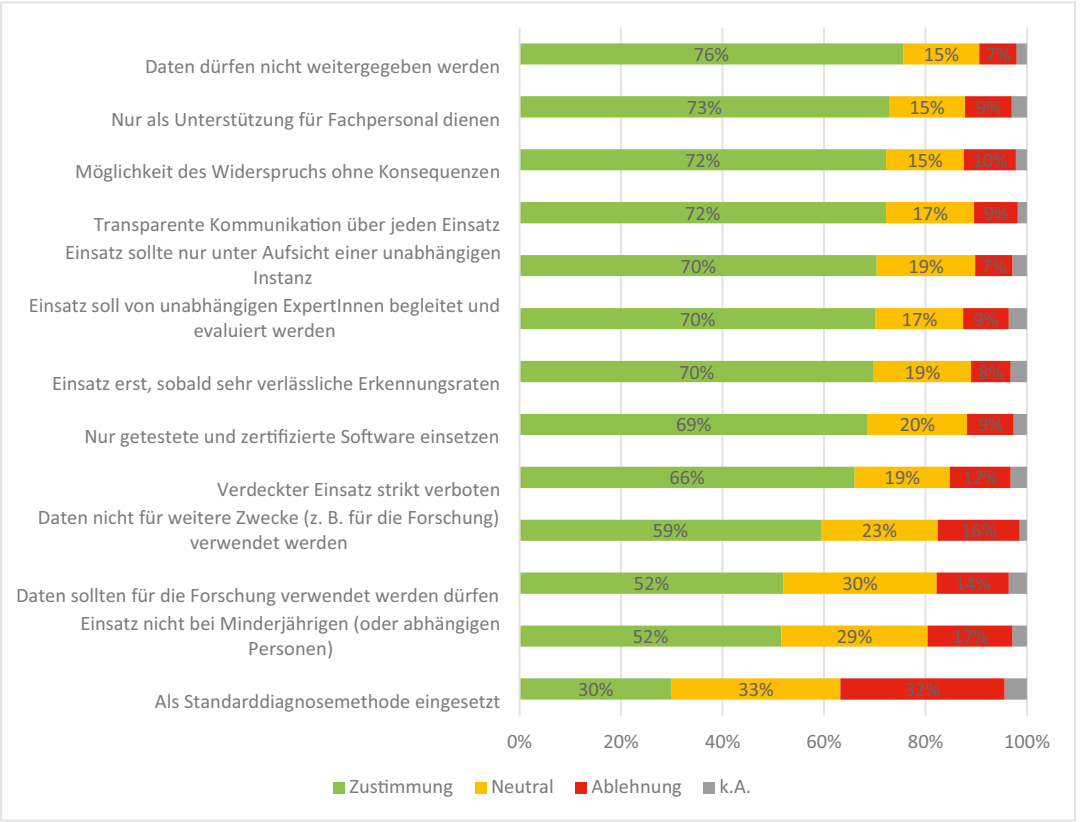

Abbildung 56: Meinungen zur Ausgestaltung des Einsatzes (Mehrfachnennung möglich)

5.2.7. Gesichtserkennung im Schulkontext zur Aufmerksamkeitsanalyse

Meinungen über den Einsatz

Die absolute Mehrheit der Befragten (56 %) befürwortete ein grundsätzliches Verbot der Aufmerksamkeitsanalyse in Schulen (Abbildung 57). Zu dieser Einstellung tendieren etwas mehr Frauen (60 %) als Männer (52 %) und etwas mehr Menschen aus städtischen Gegenden (58 %) gegenüber Menschen aus ländlichen Regionen (51 %). Während 28 % angaben, keine klare Meinung zum Thema zu haben, befürworteten 13 % der Befragten den Einsatz. Dabei sticht hervor, dass mit 18 % der männlichen Befragten gegenüber 9 % der weiblichen Befragten weitaus mehr Männer für den Einsatz stimmten.

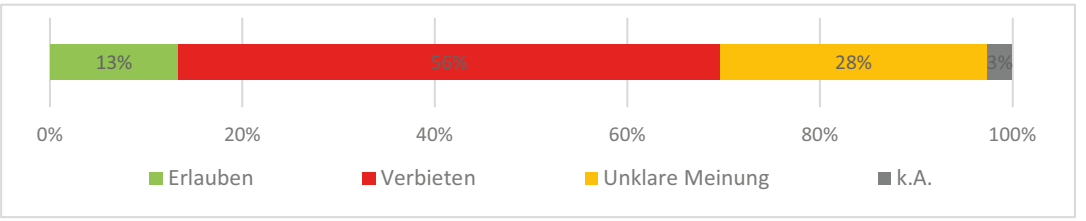

Abbildung 57: Meinungen zum Einsatz von Aufmerksamkeitserkennung in Schulen

Gründe für Ablehnung oder Befürwortung des Einsatzes

Im Folgenden wurden die Teilnehmenden nach den Gründen für ihre jeweilige Meinung befragt. Gegner des Einsatzes (n=281) gaben als Grund für ihre Ablehnung am häufigsten an (Abbildung 58), dass der Einsatz den Schulalltag oder den Lernerfolg der Minderjährigen stören (83 %), die freie Entwicklung zu mündigen Erwachsenen beeinträchtigen (82 %) und die Daten missbraucht werden könnten (79 %). 79 % teilten die Meinung, dass Minderjährige generell keinerlei Überwachung ausgesetzt werden sollten, und 73 % befürchteten, dass die Aufmerksamkeitsanalyse die Grundlage für weitere, noch invasivere Überwachungsmassnahmen sein könnte. Die Meinung, dass die Aufmerksamkeitsanalyse nicht ausreichend zuverlässig funktioniere, teilten 69 % der Befragten.

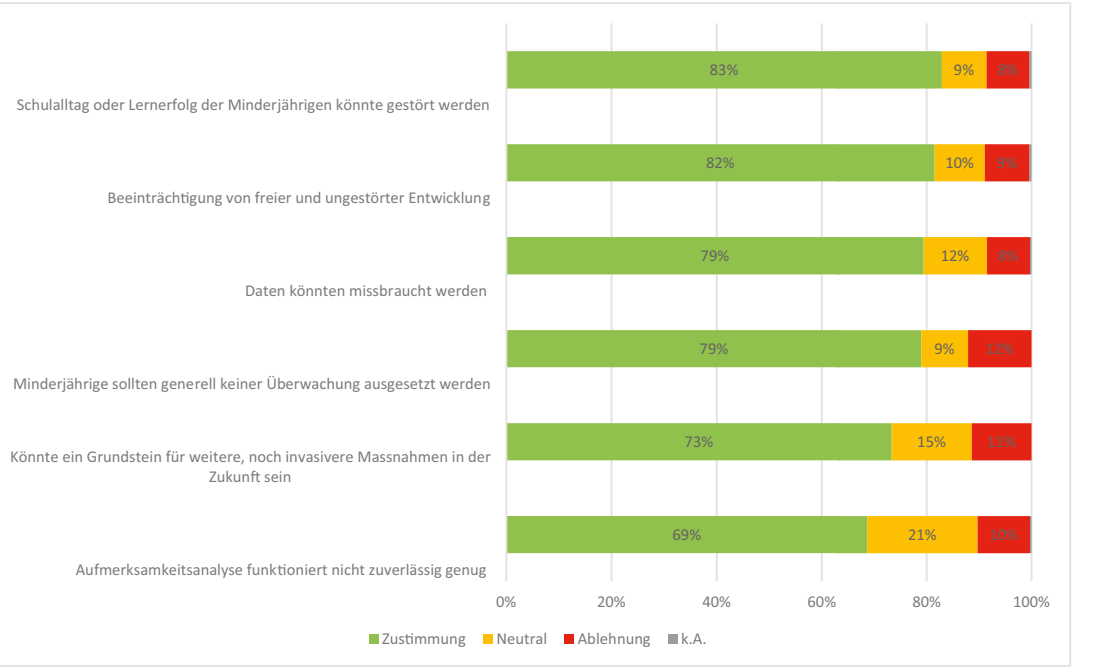

Abbildung 58: Gründe von Gegnern des Einsatzes (Mehrfachnennung möglich)

Befragte mit unklarer Meinung (n=139) gaben als Grund für ihre Meinung am häufigsten an (Abbildung 59), sich noch nicht ausreichend mit dem Thema befasst zu haben (67 %) und dass es auf die konkreten Einsatzzwecke (63 %) bzw. auf die konkrete Gestaltung des jeweiligen Einsatzes (62 %) ankomme.

Befürworter des Einsatzes (n=67) erachten die Aufmerksamkeitsanalyse von Schülerinnen und Schülern als wünschenswert (Abbildung 60), weil es zur besseren Disziplinierung der Schülerinnen und Schüler beitragen könne (69 %), die Bewertung der Schüler dadurch neutraler erfolgen könne, weil Lehrkräfte nicht imstande seien, alle Schülerinnen dauerhaft im Blick zu behalten (69 %), und die Technologie zugleich Kapazitäten bei den Lehrkräften freimachen könnte, sodass sie sich auf wichtigere Dinge, wie das Beantworten von Fragen oder die Betreuung leistungsschwacher Schüler, konzentrieren könnten (67 %). Schliesslich

stimmten 64 % der Befürworter der Aussage zu, dass die Aufmerksamkeitsanalyse in der Schule insgesamt zu einer besseren Schulbildung beitrage könne.

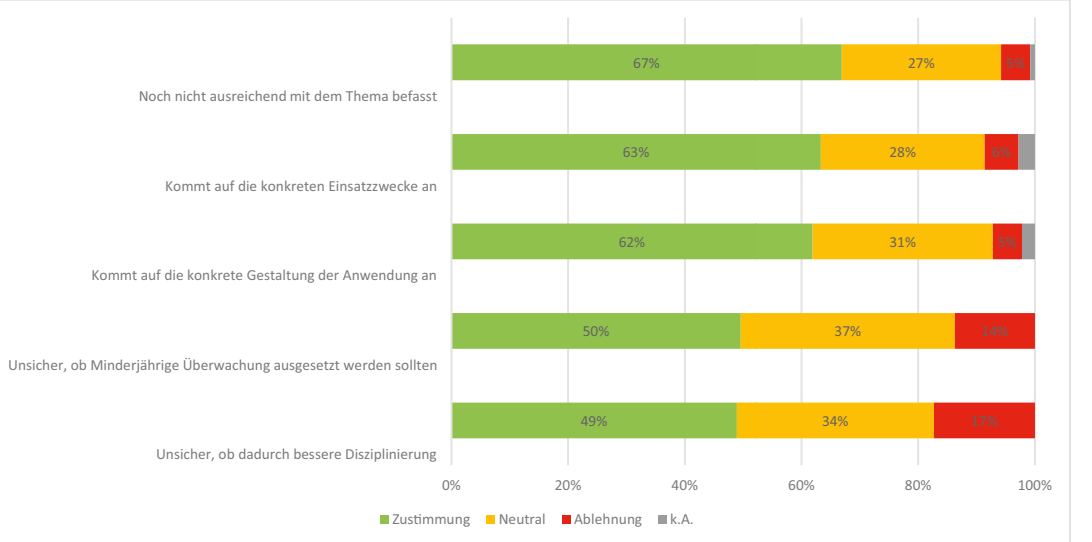

Abbildung 59: Top-5-Gründe von Personen, die unsicher über den Einsatz sind (Mehrfachnennung möglich)

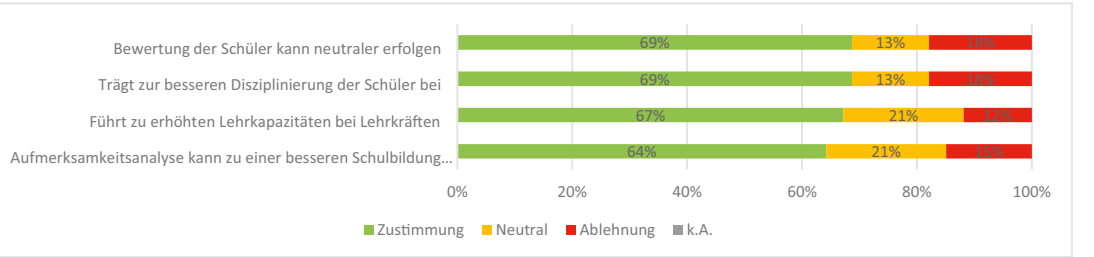

Abbildung 60: Gründe von Befürwortern des Einsatzes (Mehrfachnennung möglich)

Weitere Einsatzzwecke

Alle Teilnehmenden (N=500) wurden zudem danach befragt, für welche weiteren Zwecke die Gesichtserkennung in Schulen eingesetzt werden sollte (Abbildung 61). Eine Mehrheit fand den Einsatz von Gesichtserkennung zur Einlasskontrolle (60 %), zur Identifikation nicht autorisierter Personen an Schulen und der Verhinderung von Betrug bei (Online-)Prüfungen (45 %) für sinnvoll. Abgelehnt wurde der Einsatz von Gesichtserkennung hingegen in Bezug auf die Durchführung von Anwesenheitskontrollen bei Schülerinnen und Schülern (46 %) und zur Analyse der Aufmerksamkeit von Lehrkräften (52 %). Männer zeigten sich etwas häufiger mit diesen einverstanden als Frauen. So befürworteten 39 % der männlichen und 30 % der weiblichen Befragten den Einsatz zur allgemeinen Anwesenheitskontrolle. Ebenso befürworteten mehr Männer die Analyse von Lehrkräften (30 %) und den Einsatz zur Verhinderung von Betrug (51 %). Demgegenüber wurden die entsprechenden Einsatzzwecke von 19 % und 39 % der weiblichen Befragten begrüsst.

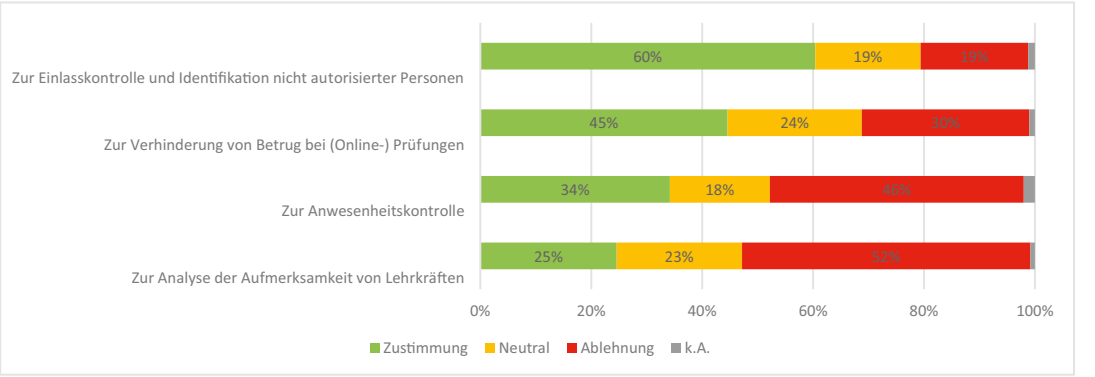

Abbildung 61: Meinungen zu weiteren Einsatzzwecken (Mehrfachnennung möglich)

Ausgestaltung des Einsatzes

Zum Abschluss wurden alle Teilnehmenden (N=500) danach gefragt, wie die Gesichtserkennung zur Bewertung der Aufmerksamkeit von Schülerinnen und Schülern im Falle eines Einsatzes ausgestaltet sein sollte (Abbildung 62). 74 % aller Befragten sprachen sich dafür aus, dass ein solcher Einsatz transparent kommuniziert werden müsste. 71 % der Befragten stimmten der Aussage zu, dass die Aufmerksamkeitsanalyse lediglich unterstützend eingesetzt werden dürfen sollte, sodass den Schülern im Falle von Unaufmerksamkeit keine negativen Konsequenzen entstehen könnten. 68 % teilten die Ansicht, dass die Schüler und Schülerinnen bzw. deren Erziehungsberechtigte die Möglichkeit haben sollten, der Gesichtserkennung zu widersprechen, ohne negative Konsequenzen zu befürchten. Schliesslich sprachen sich 67 % aller Befragten dafür aus, dass der Technologieeinsatz von unabhängigen Experten begleitet und evaluiert werden sollte, und 64 % dafür, dass die Technologie nur in Ausnahmefällen und zeitlich befristet, etwa an Brennpunktschulen, eingesetzt werden sollte.

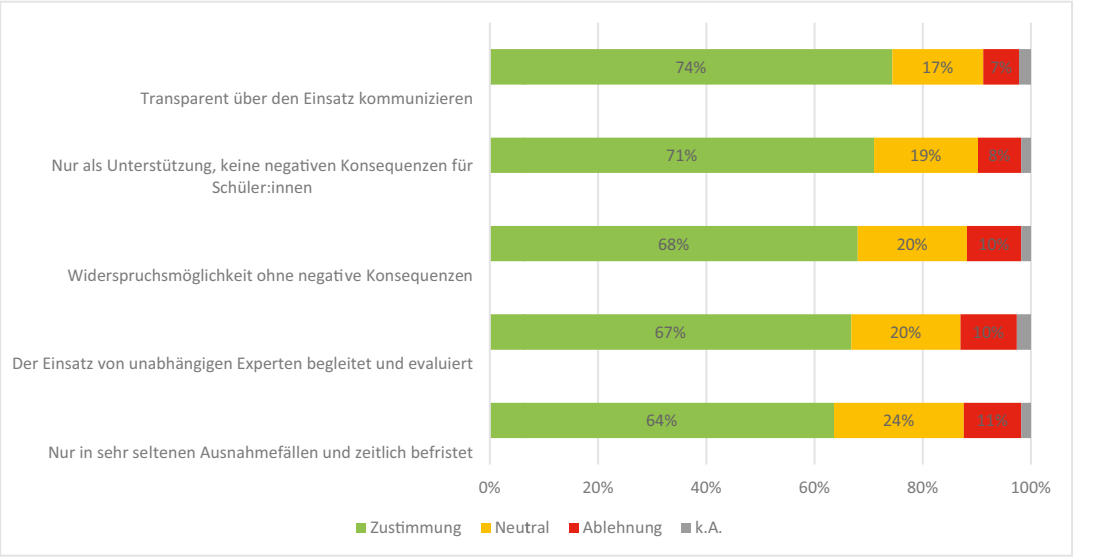

Abbildung 62: Meinungen zur Ausgestaltung des Einsatzes (Mehrfachnennung möglich)

5.2.8. Jedermann-Identifikation

Bedenken

Eine klare Mehrheit der Befragten (51,4 %) äusserte grosse Bedenken hins. der Jedermann-Identifikation (Abbildung 63). 31 % gaben an, keine klare Meinung zu haben, und 13 % äusserten, dass sie keine Bedenken hätten.

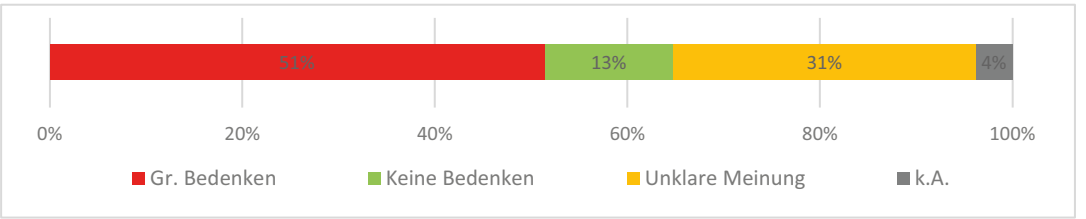

Abbildung 63: Bedenken hins. der Jedermann-Identifikation

Gründe für Bedenken

Als Nächstes wurden die Teilnehmenden nach den Gründen für ihre jeweilige Meinung befragt.

Unter denjenigen, die grosse Bedenken geäussert hatten (n=257), wurden die Bedenken am häufigsten damit begründet (Abbildung 64), dass mittels der Jedermann-Identifikation personenbezogene Informationen in die Hände von Kriminellen gelangen könnten und so Identitätsraub ermöglicht oder vereinfacht wird (91 %). Daneben gaben 90 % der Befragten an, dass sie befürchten, sich im Falle der Verbreitung der Jedermann-Identifikation nicht mehr anonym in der Öffentlichkeit bewegen zu können. Etwa ähnlich häufig wurde auch der Aussage zugestimmt (89 %), dass das Stalking von Menschen mittels der Jedermann-Identifikation vereinfacht würde.

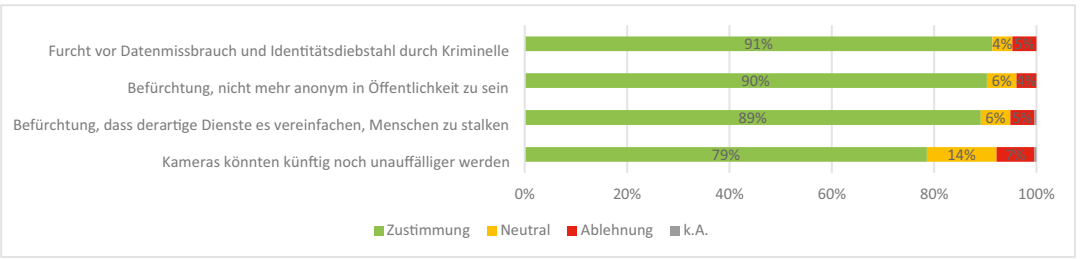

Abbildung 64: Gründe von Befragten mit grossen Bedenken (Mehrfachnennung möglich)

Gefragt nach den Gründen dafür, weshalb sie keine klare Meinung zum Thema haben (Abbildung 65), gaben die meisten Befragten aus dieser Teilgruppe (n=157) an, dass sie denken, dass die Bewertung von den konkreten Anwendungen abhängen werde (75 %) und dass sie sich noch nicht ausreichend mit dem Thema befasst hätten (73 %). Zudem stimmten 66 % der Befragten der Aussage zu, dass es von der Zuverlässigkeit der Identifikation abhängig sein werde.

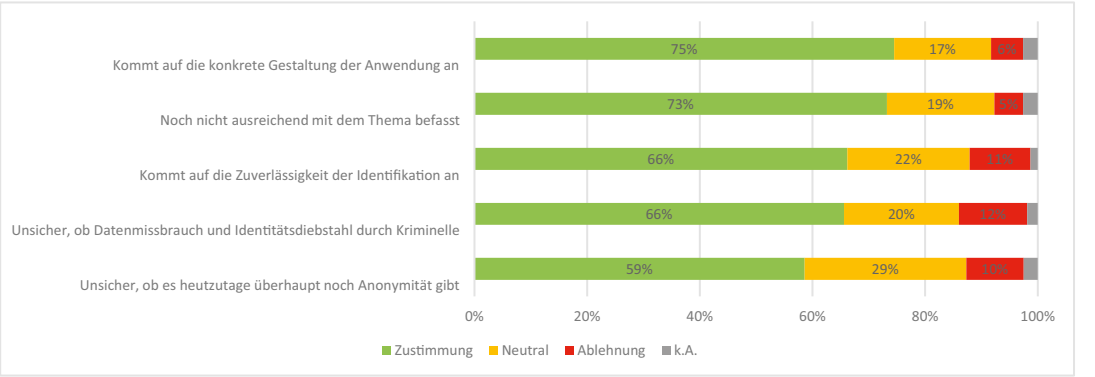

Abbildung 65: Top-5-Gründe von Befragten mit unklarer Meinung (Mehrfachnennung möglich)

Jene Befragte, die angegeben hatten, keine Sorgen vor der Jedermann-Identifikation zu haben (n=67), begründeten ihre Meinung am häufigsten damit (Abbildung 66), dass es heutzutage sowieso keine Anonymität mehr gäbe (76 %). Ausserdem teilten 70 % die Aussage, dass sich durch die Jedermann-Identifikation viele spannende Nutzungszwecke ergeben könnten. 63 % gaben an, dass man sich als Bürger einer westlichen Demokratie keine Sorgen machen müsse.

Die Aussage, dass Privatsphäre bzw. Anonymität ein veraltetes Konzept und daher nicht mehr schützenswert sei, wurde von einer relativen Mehrheit von 42 % geteilt (bei einer Ablehnung von 37 %). 43 % der Befragten verneinten die Aussage, dass sie deshalb keine Befürchtungen hätten, weil sie sich ohnehin darum bemühten, keine Bilder im Internet zu posten (Bejahung bei 36 %).

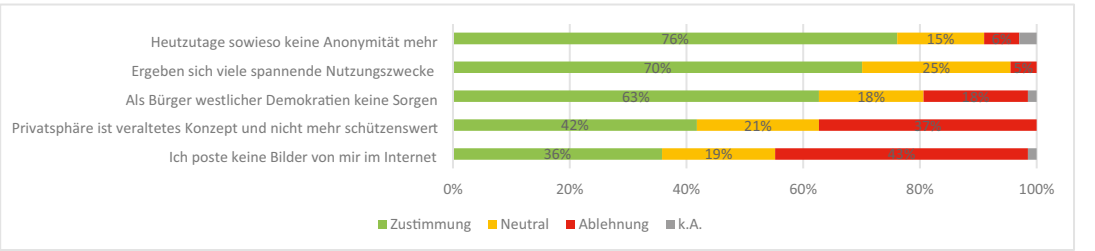

Abbildung 66: Meinung von Befragten, die keine Bedenken äusserten (Mehrfachnennung möglich)

Ausgestaltung des Einsatzes

Abschliessend wurde alle Teilnehmer (N=500) danach gefragt, wie mit der Jedermann-Identifikation umgegangen werden sollte (Abbildung 67). Eine grosse Mehrheit von 82 % vertrat die Ansicht, dass Identitätsdaten durch entsprechendes Technikdesign seitens der Onlinedienste sicher gegen eine Zweckentfremdung geschützt werden müssten. 65 % der Befragten waren der Ansicht, dass es technisch unmöglich gemacht werden sollte, Personen ohne deren Kenntnis zu fotografieren, und 64 % sprachen sich für ein Verbot des verdeckten Einsatzes von Gesichtserkennung mittels Datenbrillen aus.

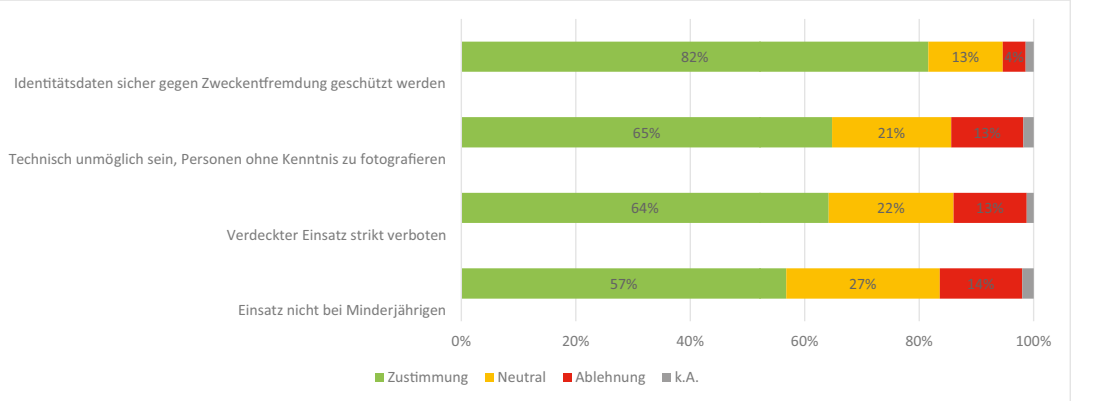

Abbildung 67: Wünsche zur Ausgestaltung der Jedermann-Identifikation (Mehrfachnennung möglich)

5.3. Zusammenfassung der Ergebnisse und Schlussfolgerungen

5.3.1. Zentrale Ergebnisse der Bevölkerungsumfrage

Smarte Lautsprecher

Ein smarter Lautsprecher befindet sich in 37 % aller Schweizer Haushalte. Die grosse Mehrheit dieser Haushalte (72 %) nutzt den Lautsprecher erst seit einem Jahr. Während sich 14 % der übrigen Haushalte die Anschaffung eines smarten Lautsprechers vorstellen können, wird der Gedanke der Anschaffung von 41 % der übrigen Haushalte abgelehnt. Zugleich sind sich 39 % hins. dieser Frage noch unsicher (Abbildung 68).

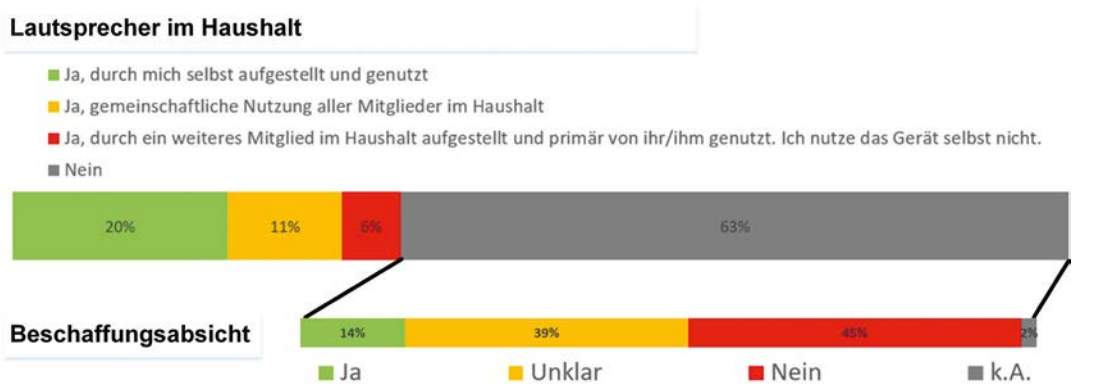

Abbildung 68: Beschaffungsabsicht eines smarten Lautsprechers

Gesichts- und Spracherkennung durch polizeiliche Stellen

Die knappe Mehrheit der befragten Schweizerinnen und Schweizer (33 %) hat keine klare Meinung hins. des Einsatzes der Gesichts- und Spracherkennung durch polizeiliche Stellen. Für den Einsatz votieren 33 % der Befragten, während 31 % der Befragten dafür sind, den Einsatz bis auf Weiteres zu verbieten (Abbildung 69).

Gesichtserkennungstechnologie in Sportstadien

Die knappe absolute Mehrheit der Befragten (50 %) sprach sich für den Einsatz von Gesichtserkennungstechnologien in Sportstadien aus. Abgelehnt wurde der Einsatz von 19 % und eine unklare Meinung äusserten 27 % der Befragten (Abbildung 69).

Aufmerksamkeitsanalyse mittels Gesichtserkennung an Schulen

Die absolute Mehrheit der Befragten (56 %) befürwortete ein grundsätzliches Verbot der Aufmerksamkeitsanalyse an Schweizer Schulen. 28 % gaben an, keine klare Meinung zu haben und 13 % befürworteten den Einsatz. Die Gesichtserkennung in Schulen zu Sicherheitszwecken in Form einer Einlasskontrolle und Identifikation nicht autorisierter Personen wurde von einer absoluten Mehrheit von 60 % der Befragten befürwortet. Die gleichzeitige Nutzung der Gesichtserkennung zur Anwesenheitskontrolle wurde hingegen von einer relativen Mehrheit von 46 % abgelehnt. Eine relative Mehrheit von 45 % befürwortete zudem den Einsatz von Gesichtserkennung zum Zwecke der Verhinderung von Betrug bei (Online-) Prüfungen (Abbildung 69).

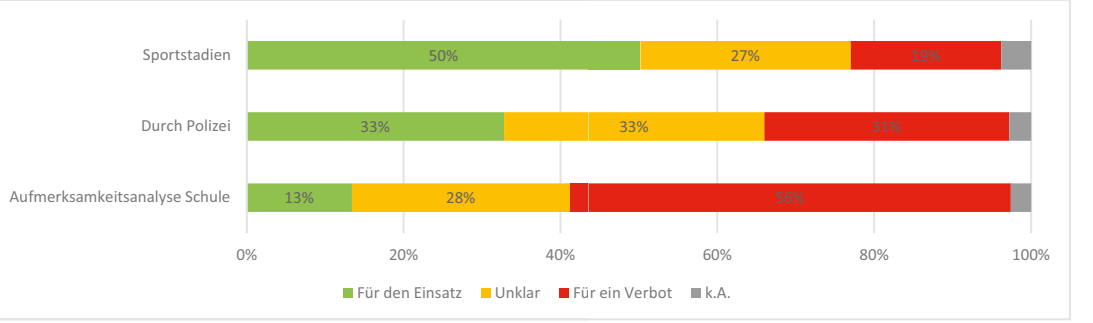

Abbildung 69: Erwünschtheit des Technologieeinsatzes in Sportstadien, durch polizeiliche Stellen und in Schulen

Authentifizierung via Stimme bei Telefonbanking

Hinsichtlich des Einsatzes von Stimmerkennungstechnologien zum Zwecke der Authentifizierung von Anrufenden beim Telefonbanking gab eine relative Mehrheit von 37 % der Befragten an, grosse Bedenken zu haben. Während 33 % der Befragten angaben, keine klare Meinung zum Thema zu haben, äusserten 24 % der Befragten, dass sie keine Bedenken hätten (Abbildung 70).

Erkennung physischer Krankheiten

Die knappe Mehrheit der Befragten (37 %) gab an, keine klare Meinung zum Einsatz von Stimm- und Gesichtserkennungstechnologien zur Erkennung physischer Krankheiten zu

haben. Zugleich äusserten 37 % grosse Bedenken, während 23 % keine Bedenken äusserten (Abbildung 70).

Erkennung psychischer Krankheiten

Auch beim Thema der Erkennung psychischer Krankheiten unter Einsatz von Stimm-, Sprach- und Gesichtserkennungstechnologien äusserte sich eine Mehrheit von 38 % der Befragten dahin gehend, keine klare Meinung zu haben. 36 % der Befragten gaben zugleich an, grosse Bedenken zu haben, und 23 % äusserten keine Bedenken (Abbildung 70).

Jedermann-Identifikation

Die knappe absolute Mehrheit der Befragten (51 %) äusserte grosse Bedenken hins. der Jedermann-Identifikation. 31 % gaben an, geteilter Meinung zu sein, während sich 13 % dahin gehend äusserten, keine Bedenken zu haben (Abbildung 70).

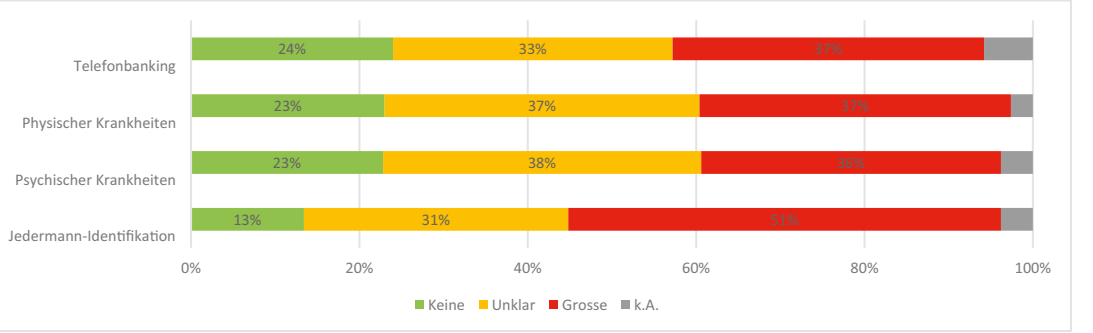

Abbildung 70: Bedenken hins. des Technologieeinsatzes beim Telefonbanking, zur Erkennung physischer und psychischer Krankheiten und der Jedermann-Identifikation

5.3.2. Vor- und Nachteile aus Bevölkerungssicht

Über alle in der Bevölkerungsumfrage abgefragten Anwendungsgebiete hinweg zeigten sich übergreifende Argumente, die aus Sicht der Befragten für und gegen die Technologien sprachen oder die zu einer geteilten Meinung bei ihnen führten.

Übergreifende Argumente für den Einsatz aus Sicht der Befürworter und Unbesorgten

Über verschiedene Anwendungsgebiete hinweg nannte eine Mehrheit der Befürworter und Unbesorgten (bei mehr als zwei Anwendungsgebieten) übergreifende Argumente für einen Technologieeinsatz. Hierzu zählen (in absteigender Häufigkeit der Nennung):

- Sicherheitsgewinn (4)
- Vertrauen in die für den Technologieeinsatz Verantwortlichen (4)
- Erwartete Effizienzsteigerungen (4)
- Glaube, dass mittels Technikeinsatz weniger diskriminiert würde als durch Menschen (3)
- Verbesserung der Gesundheit (2)
- Neugier an neuen Technologien (2)

- Mehr Komfort im Alltag (2)
- Abschreckungswirkung (2)

Als vorteilhaft wurden insb. die möglichen **Sicherheitsgewinne** hervorgehoben. Bei der polizeilichen Gesichtserkennung, der Authentifizierung via Stimme, der Stadionüberwachung und der Gesichtserkennung im Schulkontext waren Sicherheitsaspekte entscheidend im Hinblick auf eine positive Bewertung. Bei der polizeilichen Gesichtserkennung und der Stadionüberwachung standen die Verbrechensbekämpfung und -aufklärung im Vordergrund, bei Ersterer ausserdem insb. das Auffinden vermisster Menschen. Die Authentifizierung via Stimme empfanden die Befragten als eine sicherere Authentifizierungsmethode und im Schulkontext wurde die Gesichtserkennung als eine wirksame Methode zur Einlasskontrolle zur Verhinderung des Eintritts nicht autorisierter Personen angesehen.

Ein weiteres Argument ist das **Vertrauen in die für den Einsatz Verantwortlichen**. Sowohl bei der polizeilichen Gesichtserkennung als auch der Stadionüberwachung gab eine Mehrheit der Befürworter an, den Behörden bzw. Stadionbetreibern zu vertrauen, die Technologie gewissenhaft und verantwortungsvoll einzusetzen. Das Vertrauen in das Gesundheitssystem drückte sich bei den Themen zur Erkennung psychischer und physischer Krankheiten darin aus, dass die Befragten der Meinung waren, die automatische Analyse würde lediglich als Unterstützung für den Arzt dienen und nicht selbstständig Diagnosen stellen.[186]

Die Erwartung, dass der **Technologieeinsatz zu Effizienzsteigerungen führen** werde, wurde ebenfalls über vier Anwendungsgebiete hinweg als ein wichtiger Vorteil genannt. Bei der Stadionüberwachung gab eine grosse Mehrheit der Befürworter an, dass unerwünschte Personen durch die Technologie besser und schneller erkannt werden könnten. Bei der Erkennung physischer und psychischer Krankheiten wurde hingegen auf die Potenziale hins. Kosteneinsparung und der Freihaltung medizinischer Kapazitäten verwiesen. In ähnlicher Weise stimmte eine Mehrheit der Befürworter der Aufmerksamkeitsanalyse dem Argument zu, dass die Überwachung der Aufmerksamkeit durch die Technologie zu erhöhten Lehrkapazitäten bei den Lehrkräften führen könnte, sodass diese sich auf wichtigeres, wie die Betreuung der Schülerinnen und Schüler, konzentrieren könnten.

Mehrfach geäussert wurde auch die Überzeugung, dass computergestützte Verfahren **Diskriminierung verringern** würden. Sowohl bei der polizeilichen Gesichtserkennung als auch der Stadionüberwachung zeigte sich eine Mehrheit der Befragten überzeugt, dass Menschen in stärkerem Masse zu Diskriminierung neigten und der Technologieeinsatz Diskriminierung reduzieren könnte.

Bei der Erkennung physischer und psychischer Krankheiten wurde die Hoffnung geäussert, dass der Technologieeinsatz zur **Verbesserung der Gesundheit** führen werde.

Die **Neugier an der Nutzung neuer Technologien** waren ein wichtiges Argument in den Bereichen smarte Lautsprecher und Jedermann-Identifikation. Bei Letzterer lag der Fokus insb. auf der Möglichkeit spannender neuer Nutzungszwecke.

---

[186] Im Hinblick auf die Jedermann-Identifikation äusserte eine grosse Mehrheit der Befürworter dieses Einsatzes kein direktes Vertrauen in die für den Technologieeinsatz Verantwortlichen. Allerdings wurde die fehlende Sorge vor einem Missbrauch der Technologie mit dem Vertrauen in das demokratische System begründet.

Mehr Komfort im Alltag wurde als Argument bei smarten Lautsprechern und der Authentifizierung mittels Stimme genannt.

Die erwartete **abschreckende Wirkung** des Technologieeinsatzes auf mögliche Gewalt- und Straftäter wurde bei der polizeilichen Gesichtserkennung und der Stadionüberwachung als ein Vorteil geäussert.

Übergreifende Argumente gegen den Einsatz aus Sicht der Gegner und Besorgten

Zu den aus Sicht einer Mehrheit der Gegner und Besorgten (bei mehr als zwei Anwendungsgebieten) unterstützten Argumenten, die gegen einen Technologieeinsatz sprechen, zählen (in absteigender Häufigkeit der Nennung):

- Datenschutzbedenken (8)
- Sorge vor der Unzuverlässigkeit der Technologie (6)
- Allgemeine Sorge vor Missbrauch der Technologie (5)
- Wahrgenommener fehlender Mehrwert des Einsatzzweckes (3)
- Unsicherheit im Hinblick auf die Frage der Verantwortungsübernahme im Falle einer Fehldiagnose (2)

Bei allen Anwendungsgebieten gab eine grosse Mehrheit der Gegner und Besorgten an, diese Meinung aufgrund von **Datenschutzbedenken** zu vertreten. Dies reichte von der Furcht davor, dass smarte Lautsprecher alles Gesprochene mithören und aufzeichnen, über Sorgen vor einer anlasslosen Massenüberwachung der Bevölkerung im Falle polizeilicher Gesichtserkennung, der Angst vor einem unzureichenden Schutz der erhobenen Daten bei der Authentifizierung mittels Stimme und der Erkennung von physischen und psychischen Krankheiten bis hin zur befürchteten vollständigen Erosion der Anonymität im Falle der Jedermann-Identifikation. Grosse Übereinstimmung gab es auch hins. der Befürchtungen, dass personenbezogene Daten in die Hände von Unbefugten, insb. Kriminellen gelangen könnten (bei den Anwendungsfällen Jedermann-Identifikation und Authentifizierung mittels Stimme).

Eine weitere, über fast alle Anwendungsgebiete geteilte Sorge betrifft die befürchtete **Unzuverlässigkeit der Technologie** und daraus möglicherweise resultierende Konsequenzen. Bei der polizeilichen Gesichtserkennung und der Stadionüberwachung wurde befürchtet, dass eine Fehlerkennung Konsequenzen für unschuldige Menschen nach sich ziehen könnte. Bei der Authentifizierung mittels Stimme gab eine grosse Mehrheit an, dass sie sich angesichts der schwankenden Sprachqualität bei Telefonaten nicht vorstellen könnten, dass die Technologie fehlerfrei funktioniert. Bei der Erkennung physischer und psychischer Krankheiten äusserten viele Befragte die Meinung, dass die Technologie gegenwärtig nicht zuverlässig genug funktioniere und daher eine Fehldiagnose zu befürchten sei.

Verknüpft mit Datenschutzbedenken, stimmte auch ein grosser Teil der Gegner bzw. Besorgten darin überein, dass sie eine **allgemeine Sorge vor dem Missbrauch der Technologie** hätten. Das verbindende Element hierbei ist v.a. die Furcht vor einer Zweckentfremdung der Daten: bei der Stadionüberwachung etwa, dass die Daten für Zwecke der personalisierten Werbung verwendet werden könnten. Im Falle von Gesundheitsdaten wurde insb. eine Weitergabe dieser an Krankenkassen befürchtet und entsprechend abgelehnt. Bei der

Jedermann-Identifikation stand schliesslich die Sorge vor einer Zunahme von Stalking im Vordergrund.

Bei drei Anwendungen bzw. Anwendungsgebieten wurde seitens der Befragten, die sich besorgt über den Einsatz oder dagegen aussprachen, angegeben, dass sie **keinen Mehrwert des Einsatzzweckes** erkennen können. Dies betrifft smarte Lautsprecher, die Authentifizierung mittels Stimme und die Stadionüberwachung. Bei beiden Letzteren kommt hinzu, dass bestehende Verfahren zur Authentifizierung bzw. zur Erkennung von Straftätern als ausreichend angesehen wurden.

In den Kontexten der Erkennung physischer und psychischer Krankheiten wurden zudem Bedenken im Hinblick auf die Frage der **Verantwortungsübernahme im Falle einer Fehldiagnose** geäussert.

Übergreifende Gründe für bestehende Unsicherheiten

Auch bei den Personen, die im Hinblick auf die Für- oder Gegensprache eine unklare Meinung äusserten, zeigten sich quer über alle Anwendungsgebiete hinweg (bei mehr als zwei Anwendungsgebieten) übergreifende Gründe dafür. Zu diesen zählen insb. (in absteigender Häufigkeit der Nennung):

- Unzureichende Beschäftigung mit dem Thema (8)
- Unsicherheiten hins. der zu erwartenden Sicherheits- und Privatheitskonsequenzen (7)
- Es komme auf die konkreten Einsatzzwecke und die konkrete Ausgestaltung des jeweiligen Einsatzes an (6)
- Unsicherheiten hins. des Missbrauchspotenzials (6)
- Unsicherheiten hins. der Zuverlässigkeit und Zweckmässigkeit (6)
- Unsicherheiten im Hinblick auf den Mehrwert durch Technikeinsatz (4)
- Zweifel darüber, ob eine unwissentliche Diagnosestellung erfolgen könnte (2)

In allen Anwendungsgebieten äusserten diese Befragten mehrheitlich, dass sie sich noch **nicht ausreichend mit dem Thema befasst** hätten. Zudem wurde bei sieben Anwendungsgebieten auf **Unsicherheiten im Zusammenhang mit den möglichen Sicherheits- und Privatheitskonsequenzen** verwiesen. Die Befragten gaben ausserdem im Hinblick auf je sechs Anwendungen und Anwendungsgebiete an, dass es **auf die konkreten Einsatzzwecke der Technologie und die konkrete Ausgestaltung des jeweiligen Einsatzes ankomme**, sie **unsicher über das Missbrauchspotenzial** seien und auch **unsicher über die Zuverlässigkeit und Zweckmässigkeit** der Einsätze seien. Bei der Erkennung physischer und psychischer Krankheiten, der Aufmerksamkeitsanalyse sowie der Jedermann-Identifikation äusserten die Befragten ausserdem auch **Unklarheit im Hinblick auf den mit einem Einsatz verbundenen Mehrwert**. Unsicherheiten im Hinblick auf das Vertrauen in die Verantwortlichen im Kontext der Erkennung physischer und psychischer Krankheiten äusserten sich darin, dass eine unwissentliche Diagnosestellung befürchtet wurde.

5.3.3. Ergriffene und infrage kommende Selbstschutzmassnahmen

Bei den drei Anwendungsgebieten, in denen die Teilnehmenden nach Selbstschutzmassnahmen gefragt wurden, zeigt sich die grösste Überschneidung im Hinblick auf die **Massnahme der Verhaltensänderung**. Im Kontext smarter Lautsprecher gaben die Befragten an, aufzupassen, was man in der Nähe des Lautsprechers sage, und dort weniger über private Themen zu sprechen. Bei der Gesichtserkennung durch die Polizei und der Stadionüberwachung gaben die Gegner des Einsatzes mehrheitlich an, dass sie derart überwachte Orte in Zukunft meiden würden.

Bei der polizeilichen Gesichtserkennung und der Stadionüberwachung gab eine Mehrheit der Befragten ausserdem an, **Beschwerde gegen den Technologieeinsatz** einzureichen (2) und **andere Menschen vor der Technologie warnen** zu wollen (2). Bei smarten Lautsprechern gab eine Mehrheit an, **technische und organisatorische Vorkehrungen** zu treffen. Eine relative Mehrheit der Gegner der polizeilichen Gesichtserkennung gab zudem an, **gegen den Technologieeinsatz demonstrieren** zu wollen.

5.3.4. Zusammenfassung der wesentlichen Empfehlungen

Deutliche Gemeinsamkeiten über alle Anwendungsgebiete hinweg gab es auch im Hinblick auf die Empfehlungen. Hier zeigte sich, dass sich die Befragten bei allen Anwendungsbereichen mehr Transparenz wünschen; Transparenz einerseits im Hinblick auf die Sichtbarkeit des jeweiligen Technikeinsatzes und andererseits Information bezüglich über Art, Durchführung und Zweck des Einsatzes. Im Hinblick auf sechs Anwendungsbereiche äusserte eine Mehrheit der Befragten den Wunsch, dass der **Technologieeinsatz durch unabhängige Experten begleitet und evaluiert** wird. Bei vier Anwendungsgebieten wurde von einer Mehrheit der Befragten eine **Widerspruchsmöglichkeit** gegen den Technologieeinsatz befürwortet, ohne dass daraus unerwünschte Konsequenzen für die Betroffenen resultieren. Teilweise wurde mit dem Wunsch nach einer Widerspruchsmöglichkeit auch die Nutzung von Alternativen genannt (bspw. die Authentifizierung mittels Angabe von Wohnort, Geburtsdatum oder PIN). Bei je drei Anwendungsgebieten sprachen sich die Befragten zudem für Restriktionen hins. der Speicherdauer, gegen eine Weitergabe ihrer persönlichen Daten sowie dafür aus, dass menschliche Entscheidungen nicht vollständig durch auf Stimm-, Sprach- und Gesichtserkennung basierende Verfahren ersetzt werden sollten.

Zudem wurden Empfehlungen für spezifische Anwendungen geäussert, etwa dass Gesichts- und Spracherkennung durch polizeiliche Stellen nur auf Basis einer konkreten gesetzlichen Grundlage, nach richterlicher Genehmigung und ausschliesslich durch autorisiertes Personal erfolgen sollte. Im Hinblick auf die Jedermann-Identifikation wurde befürwortet, dass Stalking und Identitätsdiebstahl durch technische Vorkehrungen seitens der Dienstanbieter angegangen werden sollten.

Die Empfehlungen im Überblick (in absteigender Häufigkeit der Nennung):

- Transparente Kommunikation über jeden Einsatz (8)
- Begleitung und Evaluation des Einsatzes durch unabhängige Experten (6)
- Widerspruchsmöglichkeit, ohne Konsequenzen befürchten zu müssen (4)

- Restriktionen hins. Datenspeicherdauer (3)
- Keine Weitergabe der Daten (3)
- Einsatz zur Entscheidungsunterstützung sollte menschliche Kontrolle nicht ersetzen (3)
- Einsatz nur nach Vorabtestung auf Wirksamkeit und Zertifizierung (2)
- Einsatz auf Basis konkreter gesetzlicher Grundlage (1)
- Einsatz nach richterlicher Genehmigung (1)
- Datenzugriff ausschliesslich durch autorisiertes Personal (1)
- Technische Vorkehrungen zum Schutz vor Missbrauch (1)

5.4. Zwischenfazit

Die Ergebnisse der Bevölkerungsumfrage deuten auf Präferenzen der Bevölkerung hinsichtlich der Für- und Gegensprache der jeweiligen Technologien hin. Die knappe absolute Mehrheit der Befragten spricht sich für die Stadionüberwachung sowie gegen den Einsatz von Aufmerksamkeitsanalysen in der Schule aus. Ausserdem äussert die absolute Mehrheit der Befragten grosse Bedenken hins. der Jedermann-Identifikation. Zur Authentifizierung mittels Stimme äussert eine relative Mehrheit grosse Bedenken und eine relative Mehrheit der Befragten lehnt die Anschaffung von smarten Lautsprechern ab.

Bei den übrigen Anwendungsgebieten (Gesichts- und Spracherkennung durch polizeiliche Stellen, Erkennung physischer und psychischer Krankheiten) überwiegt der Anteil der Unentschlossenen. Dieser Anteil bringt auch Implikationen für jene Bereiche mit sich, in denen sich knappe relative Mehrheiten finden. So bleibt ein Kippen der öffentlichen Meinung auch in den Anwendungsbereichen Authentifizierung mittels Stimme und bei smarten Lautsprechern noch im Bereich des Möglichen. Der hohe Anteil der Befragten mit unklarer Meinung und deren Gründe für Unentschlossenheit verweisen auf einen grossen Aufklärungsbedarf.

Die Betrachtung der verbundenen Vor- und Nachteile liefert klare Ergebnisse: Als Vorteil erkennen die Befragten den erwarteten Gewinn an Sicherheit, Gesundheit und Komfort. Sie erwarten aber auch Effizienzsteigerungen, sodass freiwerdende personelle oder finanzielle Ressourcen anderweitig sinnvoll eingesetzt werden können. Eine nicht unerhebliche Rolle für die positive Bewertung spielen auch die Wahrnehmung, dass der Einsatz der Technologien menschengemachte Diskriminierung reduzieren könnte und dass es sich bei den für den Einsatz verantwortlichen Akteuren um vertrauenswürdige Stellen handele.

Aufseiten der Nachteile überwiegen die Furcht vor einem Missbrauch personenbezogener Daten und daraus resultierender Gefahren, wie dem Abgreifen von Kontozugangsdaten oder der anlasslosen Massenüberwachung. Daneben befürchten die Befragten, dass Stimm-, Sprach- und Gesichtserkennungstechnologien nicht zuverlässig genug funktionieren könnten und dadurch unerwünschte Folgen nach sich ziehen könnten, wie etwa eine Kontrolle oder Verhaftung infolge einer Fehlerkennung oder eine fatale Fehldiagnose bei medizinischen Anwendungen. Schliesslich sind viele Befragte darüber besorgt, dass die Technologien auch gegen ihre Interessen (unerwünschte Werbung, Preisdiskriminierung etc.) verwendet werden können.

Dass die befragten Einsatzgegner im Hinblick auf die drei Anwendungsbereiche, in denen nach Schutzmassnahmen gefragt wurde, mehrheitlich ankündigten, im Falle eines Technologieeinsatzes ihr Verhalten zu ändern, indem sie künftig Orte meiden und aufpassen würden, was sie sagen, sollte Anlass zur Sorge geben.

Im Bereich der Empfehlungen offenbart sich der Ruf nach mehr **Transparenz** und nach mehr Information. Dass sich so viele Befragte eine **Begleitung und Evaluation der Technologieeinsätze durch unabhängige Experten** wünschen, verweist ausserdem darauf, dass Transparenz nicht allein durch eine Informationsflut seitens der Einsatzverantwortlichen gestemmt und damit die Vertrauenswürdigkeit des Einsatzes nicht alleine auf diese Weise demonstriert werden kann. Stattdessen sollen anbieterseitige Transparenz und die ständige unabhängige Evaluation von Technologieeinsätzen Hand in Hand gehen. Im Bereich der Erkennung physischer und psychischer Krankheiten wird zusätzlich die Einführung einer **Vorabtestung auf Wirksamkeit und eine Zertifizierungspflicht** befürwortet. Auf ähnliche Weise wünscht sich die Mehrheit der Befragten bei der Gesichts- und Spracherkennung durch polizeiliche Stellen, dass ein **Einsatz auf möglichst konkreten gesetzlichen Grundlagen fussen** und erst **nach richterlicher Genehmigung** ausschliesslich **durch autorisiertes Personal** durchgeführt werden sollte. Im Bereich der Erkennung physischer und psychischer Krankheiten sowie der Aufmerksamkeitsanalyse befürwortet die Mehrheit der Befragten zudem, dass der Technologieeinsatz stets **unter dem Vorbehalt menschlicher Kontrolle** erfolgen sollte. Zugleich solle eine **Widerspruchsmöglichkeit** vorhanden sein, die keine negativen Konsequenzen für die Widersprechenden mit sich bringt. Befürwortet wird schliesslich noch ein restriktiver Umgang mit erhobenen Daten: Demnach soll bei Gesundheitsdaten und Daten von smarten Lautsprechern **keine unzulässige Weitergabe an andere Stellen** erfolgen und die Daten sollten zugleich **nach einer möglichst kurzen Speicherdauer gelöscht** werden müssen.

6. Empfehlungen und Schlussfolgerungen

6.1. Empfehlungen

Dieses Kapitel soll notwendige und wünschenswerte Massnahmen für einen nachhaltigen und verantwortungsbewussten gesellschaftlichen Umgang bei der Anwendung von Stimm-, Sprach- oder Gesichtserkennungstechnologien aufzeigen. Die technischen, rechtlichen und ethischen Erörterungen dieser Studie zeigen Handlungsbedarf. Selbiges lässt sich aus der Bevölkerungsumfrage und den Fokusgruppen ableiten.

Anhand vorgängiger Untersuchungen stellen wir hierfür Empfehlungen vor. Unsere Unterscheidung nach Anwendungen mit hohem Risiko einerseits und Anwendungen mit nicht akzeptablem Risiko andererseits ist angelehnt an die entsprechenden Definitionen im KI-Verordnungsvorschlag der EU-Kommission und den dazu laufenden Debatten. Aufgrund der grossen Spannweite an diskutierten Anwendungsszenarien werden zuerst übergreifende Empfehlungen und anschliessend spezifische, auf die untersuchten Szenarien zugeschnittene, besprochen.

Empfehlung 1.1: Regulierung von Hochrisikoanwendungen

Zielgruppe: Gesetzgeber

Anwendungen der Stimm-, Sprach- und Gesichtserkennung, die insb. in die Bereiche Gesundheit, Strafverfolgung, Finanzen und Kreditvergabe, Versicherung und Schul- und Arbeitsumfeld fallen, bergen besonders hohe Risiken für die Grundrechte und das gesellschaftliche Zusammenleben. Mittels staatlicher Regulierung und Kontrolle sollte sichergestellt werden, dass sie keine nachteiligen Auswirkungen für Betroffene und die Gesellschaft entfalten.

Wie unsere Ausführungen demonstriert haben, gibt es trotz der Bemühungen einzelner Betreiber für eine vertrauenswürdige Verwendung eklatante Probleme beim Einsatz entsprechender Systeme. In verstärkter Weise betroffen ist die Möglichkeit der unabhängig nachprüfbaren Gewährleistung technisch-organisatorischer Zuverlässigkeit. Sofern Evaluationen überhaupt bekannt sind, werden sie innerhalb der betreibenden Stelle oder durch privatwirtschaftliche Anbieter ausgeführt, deren Ergebnisse sich nicht unabhängig überprüfen lassen. Angesichts der drohenden Folgen derartiger Systeme für die Grundrechte und das gesellschaftliche Zusammenleben ist ihre Regulierung und Kontrolle angebracht. Die Regulierung sollte folgende Elemente umfassen:

- Begleitung hoch riskanter Anwendungen durch ein systemisches Risikomanagementsystem (Algorithmic Impact Assessment).
- Insb. die Evaluation der technisch-organisatorischen Zuverlässigkeit von Stimm-, Sprach- und Gesichtserkennungstechnologien seitens einer unabhängigen Stelle sowohl im Vorfeld des Einsatzes als auch im Rahmen eines laufenden Post-Market-Monitorings. Einen Anknüpfungspunkt, was zu evaluieren ist, könnten die laufenden

Diskussionen zur ethischen Beurteilung von KI-Systemen leisten (vgl. z.B. auch den «Ethik-Kriterien-Fragekatalog» in Tabelle 20).

- Menschliche Beaufsichtigung hoch riskanter Anwendungen, die in der Lage ist, Möglichkeiten und Grenzen der Systeme zu erkennen und bei Fehlern gegenzusteuern.
- Dokumentation der Einhaltung der Regulierungsvorgaben mittels eines Qualitätsmanagementsystems.
- Staatliche Überprüfung der Einhaltung der Regulierungsvorgaben (z.B. ein Konformitätsbewertungssystem) sowie freiwillige Zertifizierungsmöglichkeit der Anwendungen durch die Betreiber.
- Gewährleistung der Datensicherheit.

Die entsprechende Regulierung selbst sollte allerdings auf einen angemessenen Zeitraum begrenzt und anschliessend auf Erfolg und Notwendigkeit evaluiert und ggf. zurückgenommen oder an neue Bedürfnisse angepasst werden.

Empfehlung 1.2: Verbot für Anwendungen mit nicht akzeptablem Risiko

Zielgruppe: Gesetzgeber

Über die generelle Einstufung von Anwendungen der Stimm-, Sprach- und Gesichtserkennung (die insb. in die Bereiche Gesundheit, Strafverfolgung, Finanzen und Kreditvergabe, Versicherung und Schul- und Arbeitsumfeld fallen) als Hochrisikoanwendungen hinaus, sollten einige Anwendungen, die ein nicht akzeptables Risiko bergen, verboten werden.

Je nach Anwendungsbereich und Technologie können die Gründe dafür unterschiedlich sein. Emotionserkennung etwa kann aus zwei Gründen ein nicht akzeptables Risiko darstellen. Erstens muss ihre technische Zuverlässigkeit aufgrund der wissenschaftlichen Umstrittenheit der Emotionserkennung zumindest aktuell infrage gestellt werden. Im Falle von Fehlzuordnungen drohen je nach Anwendungsgebiet massive Nachteile für Betroffene. Zweitens drohen erhebliche Nachteile auch dann, wenn die Technik zuverlässig funktioniert. In diesem Falle könnte sie die korrekte Interpretation und damit das «Durchschauen» menschlichen Verhaltens ermöglichen und damit zu einer strukturellen Wissens- und Machtasymmetrie führen, bei der sich die Betroffenen nie im Klaren wären, inwiefern das Gegenüber sie bereits durchschaut hat und welche Konsequenzen daraus für sie drohen.

Zugleich muss in vielen Anwendungsbereichen, etwa im Finanz- und Kreditvergabebereich oder bei Bewerbungsverfahren, die Freiwilligkeit einer Einwilligung, die de lege lata erforderlich wäre, bezweifelt werden. Betroffene würden schlicht vor die Entscheidung gestellt, der Bearbeitung entweder einzuwilligen oder zu befürchten, aufgrund ihrer Nicht-Einwilligung schlechter behandelt zu werden bzw. abgelehnt zu werden.

Folgende Anwendungen, von denen ein eindeutiges Risiko für die Bürgerinnen und Bürger ausgeht, sollten verboten werden:

- Verbot automatisierter staatlicher Echtzeitüberwachung und staatlichen Social-Scorings mittels Stimm-, Sprach- oder Gesichtserkennung.
- Verbot vollständig automatisierter Entscheidungen gestützt auf Stimm-, Sprach- und Gesichtserkennungstechnologien in wichtigen Lebensbereichen (z.B. in den Bereichen

Gesundheit, Strafverfolgung, Finanzen und Kreditvergabe, Versicherungen, Schul- und Arbeitsumfeld). Die Ergebnisse von teilautomatisierten Entscheidungsunterstützungssystemen sollten stattdessen von geschultem Personal kritisch überprüft und freigegeben werden müssen.

- Verbot der Nutzung von Datenbrillen und anderen nicht direkt erkennbaren Technologien, die sich zur Überwachung eignen, wie kleinen Kameras, welche mit Gesichtserkennungstechnologie verknüpft sind, in der Öffentlichkeit.

Für gewisse Anwendungen sollte ein Verbot so lange gelten, wie die genügende technische und organisatorische Zuverlässigkeit und Fairness nicht erwiesen ist (Moratorium):

- Verbot von Emotions- und Krankheitserkennung in wichtigen Lebensbereichen (z.B. in den Bereichen Strafverfolgung, Finanzen und Kreditvergabe, Versicherungen, Schul- und Arbeitsumfeld), so insb. Verbot des Einsatzes von Emotions- und Krankheitserkennung bei der Stimmauthentifizierung und Bewerberauswahl durch Private sowie Verbot der Aufmerksamkeitserkennung an Schulen. Bestimmte erwünschte Anwendungen könnten auch als Hochrisikoanwendung klassifiziert und damit unter Auflagen erlaubt sein.

Empfehlung 1.3: Ausdrückliche gesetzliche Grundlagen für den Einsatz durch öffentliche Stellen

Zielgruppe: Gesetzgeber

Die Verwendung von Stimm-, Sprach- oder Gesichtserkennungstechnologien durch öffentliche Akteure benötigt **rechtsstaatliche Sicherungsmechanismen**. Namentlich ist eine **ausdrückliche und konkrete gesetzliche Grundlage in einem Gesetz im formellen Sinn** notwendig, da besonders schützenswerte Personendaten bearbeitet werden oder ein Profiling stattfindet. Die Notwendigkeit des Einsatzes dieser Technologie sollte **aufgrund des schwerwiegenden Grundrechtseingriffs** nicht leichthin angenommen und bereits auf Stufe der Gesetzgebung genau geprüft werden; im Zweifelsfall ist von der Schaffung einer gesetzlichen Grundlage abzusehen, bis die Notwendigkeit erwiesen wurde; unabhängig von einer später noch durchzuführenden Datenschutzfolgenabschätzung.

Empfehlung 1.4: Angemessene Aus- und Weiterbildung

Zielgruppe: Hochschulen, Institutionen der beruflichen Aus- und Weiterbildung, innerbetriebliche Weiterbildung. Je nach Anwendungsbereich auch weitere Akteure, im Gesundheitsbereich bspw. medizinische Fachgesellschaften, Krankenhäuser, Ärzteverband

Es muss gewährleistet sein, dass Personal, das Anwendungen der Stimm-, Sprach- und Gesichtserkennung bedient, in leitender Position für deren Betrieb verantwortlich ist oder für die Überprüfung und Freigabe von automatisierten Entscheidungen zuständig ist, die wichtige Lebensvollzüge von Menschen betreffen, **in angemessener Weise geschult** wurde.

Die Schulungen sollten mehrere Aspekte umfassen, müssten aber je nach Anwendungsfeld ggf. unterschiedliche Schwerpunkte legen oder um weitere Aspekte ergänzt werden:

- Grundsätzlich kritischer Umgang mit probabilistisch arbeitenden Systemen, die stets auf Basis von Wahrscheinlichkeiten und nicht mit sicheren Aussagen operieren.

- Die Erkennung von und einen kritischen Umgang mit algorithmen- oder datenbasierten Diskriminierungseffekten.
- Sensibilisierung mit dem Ziel der Vermeidung von Diskriminierung, wenn ein System von Menschen bedient wird.
- Ggf. die Vermittlung von grundlegendem technischem Hintergrundwissen im Bereich des maschinellen Lernens und in statistischen Methoden.

Medizinisches Fachpersonal müsste also bspw. in die Lage versetzt werden, die Ergebnisse von Diagnosestellungstools korrekt zu interpretieren und grundlegende Fragen zum Technologieeinsatz auf Nachfrage seitens der Patienten selbst beantworten zu können bzw. bezüglich detaillierter Informationen an weiterführende Informationen zu verweisen (siehe auch Empfehlung 6.1). Polizisten müssten einen kritischen Umgang mit den Ergebnissen der genutzten Gesichtserkennungsanwendungen erlernen, um Diskriminierung zu vermeiden.

Diese Inhalte müssten sowohl in die Grundausbildung als auch in laufende Betriebsprozesse (z.B. über Weiterbildungen) integriert werden.

In den entsprechenden Institutionen bzw. bei den Weiterbildungsmassnahmen sollte zudem über Grenzen der Schulbarkeit offen gesprochen und an Lösungsmöglichkeiten für dieses Problem gearbeitet werden.

An dieser Stelle sei zudem auf die Wichtigkeit der Empfehlung der TA-SWISS-Studie zu *Chancen und Risiken der KI* hingewiesen, wonach Fachleute, welche KI-Systeme entwickeln, implementieren oder über deren Einsatz entscheiden, sich Kenntnisse über rechtliche, ethische und soziale Aspekte der Nutzung von KI aneignen sollen (Christen et al. 2020, S. 298).

Empfehlung 1.5: Handreichungen für Betreiber von Stimm-, Sprach- und Gesichtserkennungsanwendungen

Zielgruppe: Je nach Zuständigkeit Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter (EDÖB), kantonale Datenschutzbeauftragte

Die komplexe Rechtslage schafft nicht nur für die Betroffenen Herausforderungen, sondern auch für die (potenziellen) Betreiber von Stimm-, Sprach- und Gesichtserkennungstechnologien. Daher sollten öffentliche Stellen und Private, darunter insb. KMU, die Stimm-, Sprach- und Gesichtserkennung einsetzen wollen, beim datenschutzrechtskonformen Einsatz unterstützt werden.

Zu diesem Zweck sollten **Leitfäden für einzelne Anwendungsbereiche der Stimm-, Sprach- und Gesichtserkennung** erstellt werden respektive bestehende Leitfäden um diese Thematik ergänzt werden**, inkl. Darstellung des datenschutzrechtlich zulässigen und der absolut einzuhaltenden Grenzen**. Besonders dringlich erscheinen die folgenden Anwendungsbereiche:

- Smarte Lautsprecher und andere Smart-Home-Geräte mit Stimm-, Sprach- und Gesichtserkennungskomponenten
- Banken und andere Finanzdienstleistungen

- Arbeitsumfeld
- Stadien, andere Sportstätten, Grossveranstaltungen
- Krankenversicherung
- Kantonale resp. kommunale Schulbehörden
- Strafverfolgung
- Migration

Die zuständigen Datenschutzbeauftragten sollten, ähnlich wie der Europäische Datenschutzausschuss (EDSA) dies auf EU-Ebene praktiziert, den öffentlichen Diskurs verfolgen und diese Liste, je nach Notwendigkeit, um weitere Anwendungsbereiche erweitern.

Empfehlung 1.6: Unterstützung für Betroffene und Verbesserung der Rechtsdurchsetzung in ihrem Sinne

Zielgruppe: Zivilgesellschaft, Konsumentenschutz, Datenschutzbeauftragte

Von Anwendungen der Stimm-, Sprach- und Gesichtserkennung Betroffene sollten bei der Wahrnehmung ihrer Betroffenenrechte und ggf. auch der Nutzung der Anwendungen unterstützt werden.

Betreiber von Stimm-, Sprach- und Gesichtserkennungsanwendungen unterliegen zwar vielfältigen Informationspflichten gegenüber Betroffenen. Doch zeigt die Forschung, dass diese Informationen, wenn sie denn zur Verfügung gestellt werden, übermässig juristisch formuliert sind und es den Betroffenen nicht ermöglichen, in nützlicher Zeit zu verstehen, welche Daten erhoben werden und was damit geschieht. Das Ziel einer informierten Entscheidung wird damit verfehlt. Wenngleich eine transparente Information in erster Linie Pflicht der Datenbearbeiter ist, können auch vertrauenswürdige Dritte wie zivilgesellschaftliche Organisationen, Konsumentenschutz- oder Datenschutzbeauftragte eine wichtige Rolle dabei spielen, die betreiber- bzw. anbieterseitig bereitgestellten umfassenden Informationen kontextspezifisch und auf spezifische Betroffenengruppen zugeschnitten auf besser verständliche Weise zu kommunizieren. Die Schaffung eines Verbandsklagerechts würde die Wahrung der Betroffenenrechte ebenfalls stärken. Aktuell ist die gerichtliche Einforderung der Einhaltung dieser Rechte gänzlich den Einzelnen überlassen, welche häufig nicht über die nötigen Sachkenntnisse und finanziellen und zeitlichen Ressourcen für derartige Verfahren verfügen. Ein Verbandsklagerecht würde es erlauben, dass in der Sache spezialisierte und kompetente Organisationen Verfahren zur Wahrung der Betroffenenrechte anstreben können.

Empfehlung 1.7: Gesellschaftliche Debatte über Vor- und Nachteile von Stimm-, Sprach- und Gesichtserkennungstechnologien und den Umgang damit

Zielgruppe: Politik, Medien, (Zivil-)Gesellschaft

Aufgrund der potenziell vielfältigen unerwünschten Auswirkungen von Stimm-, Sprach und Gesichtserkennung sollte eine gesellschaftliche Debatte über die ethischen und gesellschaftlichen Herausforderungen geführt werden. Im Falle privater Anwendungen kann dies bspw. die informierte Entscheidung für oder gegen die Nutzung der Anwendung betreffen. Im Falle öffentlicher Anwendungen kann dies das demokratische Votum für oder gegen den

Einsatz beeinflussen. Aufgrund der im Rahmen der Erhebung der Bevölkerungsmeinung zutage getretenen Wissenslücken sollten der Bevölkerung zudem verlässliche Informationen zur Beurteilung der mit den Anwendungsmöglichkeiten dieser Technologien verbundenen Vor- und Nachteile bereitgestellt werden.

Zu diskutierende Fragen betreffen insb.:

1. **Welche strukturellen Veränderungen** drohen der Gesellschaft durch die Verbreitung von Stimm-, Sprach und Gesichtserkennungstechnologien und -anwendungen? Droht bspw. die **schleichende Einführung einer staatlichen Massenüberwachung**? Wie könnte diese ggf. verhindert werden? Sollte es eine Grenze für staatliche Überwachung geben?
2. **Wie sollten soziale Normen im Zeitalter einer sich zunehmend digitalisierten Gesellschaft (neu) ausgestaltet werden?** Wie sinnvoll ist es, z.B. smarte Lautsprecher zu verwenden, wenn andere Menschen da sind, und wie müsste mit unterschiedlichen Sensibilitäten bzgl. des Datenschutzes umgegangenen werden? Ist es sozial akzeptabel, mittels Jedermann-Identifikation andere Menschen zu identifizieren? Welche Nutzungsweise von z.B. Datenbrillen wäre eine erwünschte?
3. **Inwiefern kann die individuelle Selbstbestimmung durch neue Technologien beeinträchtigt werden?** Wie soll bspw. mit den möglicherweise schwerwiegenden Auswirkungen personalisierter Werbung umgegangen werden, die auf Basis von Erkenntnissen über individuelle Manipulationsmöglichkeiten aus der Emotionserkennung möglich wird?

Empfehlung 1.8: Bereitstellung ausreichender Ressourcen

Zielgruppe: Politik, Medieninstitutionen, Stiftungen

Vertrauenswürdige Dritte (z.B. Datenschutzbeauftragte, Konsumentenschutzorganisationen, weitere zivilgesellschaftliche Organisationen) wie auch Medien sollten mit ausreichenden Ressourcen ausgestattet bzw. gefördert werden, um ihren Unterstützungsauftrag für Betroffene und die Gesellschaft wahrnehmen zu können.

Neben finanziellen Ressourcen kann sich dies auch auf den Ausbau der Expertise in Form von personellen Kapazitäten beziehen, z.B. die Einstellung von Fachinformatikern.

Im Folgenden werden die anwendungsfeldbezogenen Empfehlungen besprochen.

Tabelle 18: Handlungsempfehlungen

Nr.	Anwendungsfeldspezifische Handlungsempfehlungen	Zielgruppe
2	**Smarte Lautsprecher**	
2.1	**Verbesserung der Transparenz** Da bei Sprach-Interfaces biometrische Daten (Stimme) bearbeitet werden, muss insb. hier die **Transparenz verbessert** werden. Es muss sichergestellt sein, dass die Information über Bearbeitungszwecke und umfang nicht nur im Zeitpunkt der Erstellung eines Nutzerkontos (der Person, die tatsächlich und bewusst mit dem Lautsprecher interagiert) erfolgt, sondern auch diejenigen Personen erreicht, die den smarten Lautsprecher ohne Registrierung nutzen (z.B. Gäste). Zwar gibt es in den meisten smarten Lautsprechern aktuell bereits eine Sprechererkennung und somit die Nutzung mehrerer Nutzerkonten auf einem Gerät. Jedoch schliesst dies eine vorherige Identifizierung ein. Bei Gästen im Haushalt wäre es somit technisch nötig, zuerst kurze Sprachaufnahmen der Person zu erstellen und zu analysieren, damit der Lautsprecher künftig Aussagen dieser Person ignoriert. Hinweise über die Bearbeitung der Daten sollen nicht nur einmalig, sondern frequent und bei Änderungen der technischen Bearbeitung erfolgen und können sowohl über das Sprachinterface als auch über die mit dem Lautsprecher verbundenen Smartphone-Apps erfolgen.	Hersteller/ Anbieter
2.2	**Einholen ausdrücklicher Einwilligung** Es müssen weitreichende Diskussionen und Forschungsarbeiten geleistet werden, um sicherzustellen, dass sowohl Nutzende eines Lautsprechers als auch alle anderen ggf. anwesenden Personen der Bearbeitung ausdrücklich für alle Anwendungsfälle einwilligen. Ebenfalls muss auf die Möglichkeit der Bearbeitung besonders schützenswerter Personendaten und des Profilings hingewiesen werden. Zum aktuellen Zeitpunkt scheint dies weder rechtmässig noch technisch valide möglich zu sein. Die Einwilligung muss sowohl für die direkt nutzbaren Funktionen des Lautsprechers erfolgen als auch für jeweilige nachträglich installierte Zusatzprogramme («Skills»).	Hersteller/ Anbieter
2.3	**Die Ausübung der Betroffenenrechte** (Auskunftsrecht, Recht auf Löschung etc.) sollte direkt via Sprachinterface ermöglicht werden.	Hersteller/ Anbieter

Nr.	Anwendungsfeldspezifische Handlungsempfehlungen	Zielgruppe
2.4	Die Bearbeitung von Daten auf den Geräten selbst kann zu einem Mehr an Datenschutz und individueller Datensouveränität führen und ist daher grundsätzlich zu befürworten. Hersteller und Skill-Anbieter sollten daher verpflichtet werden, **möglichst viele Daten direkt auf dem Gerät zu bearbeiten** (On device, Edge Computing). Allerdings stellen sich dabei verschiedene Herausforderungen: Es ist z.B. unwahrscheinlich, dass aussereuropäische Anbieter dieser Pflicht Folge leisten würden. Zudem müsste eine derartige Verpflichtung sorgfältig gestaltet sein, sodass wünschenswertes Cloud-Computing nicht unbeabsichtigterweise verboten wird.	Gesetzgeber
2.5	Ausbildung sozialer Normen und Verbote/Erlaubnisse, inwiefern private Gastgeber ihre Gäste auf das Vorhandensein eines smarten Lautsprechers aufmerksam machen müssen.	Gesellschaftliche Debatte
3.	Gesichts- und Spracherkennung durch polizeiliche Stellen im öffentlichen Bereich	
3.1	**Verbot von Echtzeitüberwachung und staatlichen Social-Scorings** Eine totale oder sehr weitreichende und flächendeckende staatliche **(Echtzeit-)Überwachung** oder ein staatliches **Social Scoring-System**, das Menschen aufgrund ihres Verhaltens profiliert, würde den absolut geschützten Kernbereich mehrerer Grundrechte betreffen und wäre in jedem Fall **unzulässig**, sollte aber zusätzlich gesetzlich klar ausgeschlossen werden.	Gesetzgeber
3.2	Gesichtserkennung im öffentlichen Raum benötigt **zwingend eine Grundlage in einem Gesetz im formellen Sinn** und allenfalls eine Präzisierung auf Verordnungsstufe. Die zu regelnden Punkte wären: • Klare Definition des Ziels und Zwecks des Einsatzes der Gesichtserkennungstechnologie • Art und Umfang der Datenbearbeitung, insb. Zeitpunkt, Ort und überwachte Personen • Beteiligte Behörden bzw. Zugriffsberechtigte und verantwortliche Stelle • Kategorien der bearbeiteten Daten • Regelung der Aufbewahrung und Löschung der Daten • Gewährleistung der Rechte der Betroffenen • Verhältnismässigkeit (z.B. zeitlich begrenzt und nur bei akuten Bedrohungssituationen) Technische Umsetzung des Überwachungssystems, minimale Zuverlässigkeit und Genauigkeit des Algorithmus, Rückverfolgbarkeit des Prozesses, Sicherheitsmassnahmen und die Systemverantwortlichen, Privacy by Design (technische Zugriffsschranken, technisch unwiderlegbare Protokollierung des Zugriffs), regelmässige Evaluierung durch unabhängige Experten, um die Zuverlässigkeit der Systeme zu gewährleisten. Sofern möglich, sollten entsprechende Berichte der allgemeinen Öffentlichkeit zur Verfügung gestellt werden.	Gesetzgeber, Rechtsanwender und Datenschutzbehörden, Wissenschaft, IT-Security-Unternehmen, …

Nr.	Anwendungsfeldspezifische Handlungsempfehlungen	Zielgruppe
3.3	Auch bei bestehender gesetzlicher Grundlage darf die Gesichtserkennung im öffentlichen Raum angesichts der Schwere des damit einhergehenden Eingriffs in mehrere Grundrechte nicht leichthin angeordnet werden. **Die öffentlichen Interessen an der Überwachung, deren Verhältnismässigkeit, insb. das Vorliegen milderer Mittel** (z.B. herkömmliche Videoaufnahmen, menschliche Super Recognizer etc.) und damit die Erforderlichkeit, **sollten im Einzelfall sorgfältig geprüft, abgewogen und dokumentiert werden**. Auch sollte der Einsatz von Gesichtserkennung während des Einsatzes von unabhängiger Stelle auf deren technisch-organisatorische Zuverlässigkeit hin evaluiert werden.	Rechtsanwender (z.B. Polizei, Gemeinde)
4.	Authentifizierung via Stimme	
4.1	**Die Stimme sollte nicht als (alleiniger) Authentifizierungsfaktor verwendet werden.** Bei der Stimme handelt es sich um ein unveränderliches biometrisches Merkmal einer Person. Das heisst: Einmal abhandengekommen, kann ein Stimmabdruck nicht mehr verändert werden. Schon heute existieren Möglichkeiten der computerbasierten originalgetreuen Imitation einer Stimme in beinahe Echtzeit. Zudem wird es mit fortschreitenden technologischen Möglichkeiten, die etwa Deepfakes mit sich bringen, noch einfacher werden, Stimmen zu imitieren. Wenn auf die Stimmbiometrie als alleiniger Authentifizierungsfaktor gesetzt würde, könnten Angreifer Zugang zu Bankkonten, sensiblen Daten usw. erhalten. Entwickler von Stimmauthentifizierungstechnologien forschen an Möglichkeiten der Authentifizierung einer Stimme unter Einbezug zusätzlicher Merkmale, sodass das alleinige Imitieren erkannt werden können soll. Dies ist zu begrüssen, doch stellt sich dabei die Problematik, dass im Regelfall keine Evaluation derartiger Systeme seitens unabhängiger Stellen erfolgt, sodass deren Zuverlässigkeit schwer von aussen bewertet werden kann. Sofern diesbezügliche Zweifel aus dem Weg geräumt wären, bspw. durch eine unabhängige Evaluation bzw. eine entsprechende wissenschaftliche Debatte, würde sich auch die obige Empfehlung entsprechend relativieren.	Private Betreiber
4.2	Die **Verwendung von Stimmauthentifizierung durch Behörden bedarf einer formellgesetzlichen Grundlage**. Mindestens zu regeln wären: klare Definition des Ziels und Zwecks der Stimmauthentifizierung, Art und Umfang der Datenbearbeitung, beteiligte Behörden bzw. Zugriffsberechtigte, Kategorien der bearbeiteten Daten, Aufbewahrung und Löschung der Daten, Betroffenenrechte. Betroffene müssen die Möglichkeit haben, die behördliche Dienstleistung auch ohne Verwendung von Stimmauthentifizierung zu nutzen. Eine denkbare, verhältnismässige Variante wäre insb. eine optionale Anwendung, die interessierten Bürgerinnen und Bürgern freiwillig offenstehen würde.	Gesetzgeber

Nr.	Anwendungsfeldspezifische Handlungsempfehlungen	Zielgruppe
4.3	Die **Analyse der Stimme von Bankkunden auf Emotionen oder Krankheiten sollte verboten** werden. Sonst würde eine zu grosse Wissens- und Machtasymmetrie zwischen Anrufendem und der Bank bzw. dem Bankmitarbeitenden entstehen (vgl. auch 1.2).	Gesetzgeber
5.	Gewaltprävention und -aufklärung in Sportstadien	
5.1	**Aufnahme einer gesetzlichen Bestimmung** Die Verwendung von Gesichtserkennungstechnologie in Stadien zum Zwecke der Einlasskontrolle sowie dem flächendeckenden Filmen aller Stadiongäste zur Ermöglichung einer nachträglichen Identifikation von Gewalttätern sollte im **Bundesgesetz über Massnahmen zur Wahrung der Inneren Sicherheit (BWIS)** zentral reguliert werden. Die dabei zu überprüfenden und zu regelnden Punkte entsprechen denen aus Punkt 3.2. Da es sich bei Stadionbetreibern um private Stellen handelt, wäre zusätzlich zu regeln, dass der Technologieeinsatz von unabhängigen Instanzen (z.B. wissenschaftlichen Experten oder Datenschutzbehörden) beaufsichtigt und evaluiert werden muss. Der blossen Konsultation eines eigenen Datenschutzberaters, der zwar fachlich und weisungsunabhängig seine Tätigkeit ausführen sollte, ist kritisch gegenüberzustehen.	Gesetzgeber
5.2	**Einsatz milderer Mittel erforderlich** Der Einsatz von Gesichtserkennungstechnologien in Sportstadien dürfte sich in vielen Fällen als nicht zumutbar und nicht erforderlich erweisen, da mildere Mittel (z.B. personalisierte Tickets) eingesetzt werden könnten. Zumindest sollte Betroffenen ohne Aufpreis eine alternative Zugangslösung (z.B. Einlasskontrolle mittels eines Ausweises) angeboten werden.	Stadionbetreiber
5.3	Wird Gesichtserkennungstechnologie im Stadion eingesetzt, muss beim Kauf des Tickets und innerhalb der im Stadion überwachten Orte (also etwa beim Betreten des Stadions und auf den Rängen) **transparent und gut erkennbar** darauf hingewiesen werden. Dies sollte beim Onlineticketkauf durch das aktive Bestätigen (Opt-In) eines Hinweises (muss gelesen und akzeptiert werden) und zusätzlich in den AGB des Sportvereins erfolgen.	Stadionbetreiber
6	Erkennung von physischen oder psychischen Krankheiten	
6.1	Wird Stimm-, Sprach- oder Gesichtserkennungstechnologie zur Diagnosestellung durch medizinisches Fachpersonal verwendet, muss den **Patienten zu jeder Zeit klar sein**, **dass ein solches Diagnoseinstrument verwendet und welche Daten bearbeitet werden**. Grundlegende Fragen des Einsatzes, etwa weshalb klassische Diagnosemethoden durch stimm-, sprach- und gesichtserkennungsbasierte Technologien ersetzt werden, sollten durch das Fachpersonal beantwortet werden können.	Medizinisches Fachpersonal (weitere, z.B. medizinische Fachgesellschaften, Krankenhäuser, Ärzteverband?)

Nr.	Anwendungsfeldspezifische Handlungsempfehlungen	Zielgruppe
	Auch wenn es dem Fachpersonal nicht zumutbar scheint, sich detailliert mit der Anwendung auseinanderzusetzen, sollten ausserdem weiterführende Informationen zum Technologieeinsatz bereitgestellt werden. Dies könnte bspw. in Form von herstellerseitigen Informationsbroschüren erfolgen, die dem Patienten ausgehändigt werden, oder indem bei Fragen ein direkter Ansprechpartner bei der Herstellerfirma aufgezeigt wird.	
6.2	**Kontrolle** Gewährleistung einer verbesserten Marktüberwachung, entweder durch Zertifizierung bzw. intensivierte Kontrollen aufseiten der Marktüberwacher oder durch eine Intensivierung der Meldungen durch Dritte.	Swissmedic, ggf. auch Patienten- und Konsumentenschutzorganisationen
6.3	Eine Verwendung von Diagnosesoftware durch Krankenversicherer zu Zwecken der Prämienanpassung muss für die Versicherten resp. Versicherungsinteressierten erkennbar sein. Da gleich mehrere Datenschutzgrundsätze nicht erfüllt sind (Verhältnismässigkeit, Datenminimierung, Treu und Glauben) ist eine Rechtfertigung notwendig; denkbar ist lediglich eine (ausdrückliche) Einwilligung der Betroffenen. Ein Vertragsabschluss unter gleichen Bedingungen muss aber auch möglich sein, wenn die betroffene Person nicht in die Verwendung der Diagnosesoftware einwilligt.	Krankenversicherer
6.4	Der Einsatz von Diagnosesoftware durch Versicherer sollte **im Versicherungsvertragsgesetz (VVG) geregelt werden**.	Gesetzgeber
6.5	Selbstdiagnose-Apps sollten als **Medizinalprodukt** im Heilmittelgesetz (HMG) und in der Medizinprodukteverordnung (MepV) reguliert werden. Es sollten Kriterien festgelegt werden, wann eine medizinische Zweckbestimmung bei Selbstdiagnose-Apps vorliegt und ob eine Überprüfung durch eine akkreditierte Stelle durchzuführen ist.	Gesetzgeber
6.6	Die Frage, ab welchem Punkt der technologischen Entwicklung medizinisches Fachpersonal Diagnosesoftware zur (Teil-)Automatisierung ihrer Tätigkeit nutzen dürfte, sollte frühzeitig diskutiert werden.	Medizinische Fachgesellschaften, Krankenhäuser, Ärzteverband, Patientenorganisationen
6.7	Zusätzlich zur (ggf. auch nach Einstufung als Medizinprodukt stattfindenden) Überprüfung der Sicherheit und Leistungsfähigkeit von (Selbst-)Diagnosesoftware durch den Hersteller sollten auch **weitere Organisationen diese regelmässig auf Sicherheit und Zuverlässigkeit überprüfen** und ihre Erkenntnisse der Öffentlichkeit zugänglich machen. Diesen Organisationen sollten ausreichend Mittel von staatlicher Seite zur Verfügung gestellt werden.	Patientenorganisationen, Konsumentenschutzorganisationen, Wissenschaft

Nr.	Anwendungsfeldspezifische Handlungsempfehlungen	Zielgruppe
7.	Emotionserkennung	
7.1	Die Verwendung von Emotionserkennungstechnologien durch **Behörden** (z.B. Schulen, Migrationsbehörden) ist nicht verhältnismässig und grundsätzlich rechtswidrig. Ihr Einsatz sollte in den jeweiligen Fachgesetzen zusätzlich ausdrücklich verboten werden.	Gesetzgeber
7.2	**Moratorium für Emotionserkennung durch Private** Da sie derzeit als wenig zuverlässig gilt, aber auch Probleme mit sich bringen könnte, würde sie zuverlässig funktionieren, sollte auch die Emotionserkennung durch Private in wichtigen Lebensbereichen (z.B. in den Bereichen Gesundheit, Strafverfolgung, Kreditvergabe, Versicherungen, Schul- und Arbeitsumfeld, etwa Bewerbungsverfahren) verboten werden, bis ihre Zuverlässigkeit und Fairness in ausreichendem Masse erwiesen wurde. Bis der Gesetzgeber diesbezügliche Regeln auferlegt hat, sollten die Betreiber von stimm-, sprach- und gesichtserkennungsbasierten Emotionserkennungssystemen freiwillig auf den Einsatz in den o.g. Lebensbereichen verzichten. Dort, wo sie doch zur Anwendung kommt, erfordert die Datenbearbeitung mittels Emotionserkennungstechnologie durch Private die aktive Einholung der Einwilligung der Betroffenen, weil sie eine **Persönlichkeitsverletzung darstellt**. Ein allfällig geplanter Einsatz von **Emotionserkennungs- und Diagnosesoftware in Bewerbungsverfahren** ist dem Eidgenössischen Datenschutz- und Öffentlichkeitsbeauftragter (EDÖB) respektive dem eigenen weisungsunabhängigen Datenschutzberater vorzulegen. Eine verdeckte Verwendung von Diagnosesoftware ohne Wissen der Person ist rechtswidrig und kann mit zivilrechtlichen Rechtsmitteln angefochten werden.	Gesetzgeber, Betreiber
8.	Aufmerksamkeitsanalyse	
8.1	Die Verwendung von Aufmerksamkeitserkennung an öffentlichen Schulen hat derzeit keine rechtliche Grundlage und wäre auch weder verhältnismässig noch mit dem besonderen Schutz von Kindern und Jugendlichen sowie dem Kindeswohl vereinbar und somit rechtswidrig. Die **Rechtswidrigkeit** des Einsatzes von Aufmerksamkeitserkennung ist den Schulen proaktiv **zu kommunizieren**.	Kantonale und kommunale Schulbehörden, Gesetzgeber

Nr.	Anwendungsfeldspezifische Handlungsempfehlungen	Zielgruppe
8.2	Die Verwendung von Aufmerksamkeitserkennung in Privatschulen bedürfte einer ausdrücklichen Einwilligung der Schülerinnen und Schüler sowie ihrer Erziehungsberechtigten. Jedoch ist aufgrund der Machtasymmetrie und Missbrauchsgefahr die Freiwilligkeit der Einwilligung nicht leichthin anzunehmen und bedürfte namentlich einer besonders transparenten vorgängigen Aufklärung. Zudem bestehen praktische Schwierigkeiten, wenn nicht alle Kinder einer Klasse resp. deren Eltern eingewilligt haben. **Angesichts des Schadens- bzw. Missbrauchspotenzials sollte der Gesetzgeber den Einsatz von Aufmerksamkeitsanalyse in Privatschulen grundsätzlich verbieten. Bis der Gesetzgeber diesbezüglich tätig wurde, sollten Privatschulen freiwillig auf die Verwendung von Aufmerksamkeitserkennung verzichten.**	Gesetzgeber, Privatschulen
9.	«Jedermann-Identifikation» durch Private	
9.1	Die ungefragte Beschaffung von Gesichtsdaten zwecks Weiterbearbeitung unter Einsatz automatischer Gesichtserkennungstechnologien ist in der Regel klar persönlichkeitsverletzend, sie kann nicht gerechtfertigt werden und mangels Erkennbarkeit der Bearbeitung und der dafür Verantwortlichen ist für Betroffene kaum Rechtsschutz vorhanden. Weil keine milderen Regulierungsmassnahmen in Betracht fallen, sollte die Nutzung von Datenbrillen und anderen nicht direkt erkennbaren Technologien, die sich zur Überwachung eignen, wie kleinen Kameras, welche mit Gesichtserkennungstechnologie verknüpft sind, in der Öffentlichkeit verboten werden.	Gesetzgeber
9.2	**Ausbildung sozialer Normen und Verbote/Erlaubnisse**, in welchen Situationen und Kontexten das Tragen einer Datenbrille erlaubt wäre.	Gesellschaftliche Debatte
9.3	Die **Umsetzung von IT-Sicherheitsvorkehrungen bei Datenbrillen**: Insb. der Verzicht auf offene Schnittstellen, die die Ausführung von Erweiterungen oder eigenem Code erlauben, aber auch regelmässiges Schliessen von Sicherheitslücken durch Updates. Ggf. geschlossene Ökosysteme, die nur das Installieren geprüfter Apps erlauben. Auf diese Weise würde zwar keine absolute Sicherheit, aber zumindest eine Hürde für einen Grossteil der Nutzenden geschaffen, Datenbrillen um Gesichtserkennungsfunktionalitäten zu erweitern.	Hersteller/Anbieter

6.2. Schlussfolgerungen

Anwendungen der Stimm-, Sprach- und Gesichtserkennungstechnologien können viele Vorteile für die Betreiber bzw. Nutzerinnen und Nutzer mit sich bringen, darunter insb. Effizienzsteigerungen sowie einen Gewinn an Sicherheit, Gesundheit und Komfort. Gleichzeitig greifen sie häufig tief in die Grundrechte ein und werfen neue gesellschaftliche Fragen auf oder verstärken bestehende Probleme.

Die Analyse der technologischen Reife von Stimm-, Sprach- und Gesichtserkennungstechnologien hat aufgezeigt, dass gegenwärtig noch zahlreiche technische Herausforderungen vorhanden sind, die einen zuverlässigen Einsatz behindern. Zu erwartende Verbesserungen bei den Erkennungsalgorithmen und der Sensorik werden die technische Verlässlichkeit der Systeme in den nächsten Jahren allerdings zunehmend erhöhen. Der Einsatz von Gesichtserkennung für Authentifizierungszwecke bspw. hat schon heute sehr hohe Trefferraten, weil die Erkennung unter kontrollierten Bedingungen stattfindet. Die Erkennung des Gesichts eines sich bewegenden Menschen aus der Distanz oder einer in schlechter Qualität übermittelten Stimme ist hingegen deutlich herausfordernder. Der Einsatz von Emotionserkennung krankt darüber hinaus an grundsätzlichen konzeptionellen Problemen. Generell gilt: Eine hundertprozentige Zuverlässigkeit, also eine stets korrekte Erkennung ohne Fehler, wird es auch künftig in keinem Anwendungsbereich geben können.

Die Untersuchung des Rechtsrahmens mit Blick auf die Anwendungsfelder zeigte, dass diese durch den Rechtsrahmen an sich gut erfasst werden. Bei der Anwendung von Stimm-, Sprach- oder Gesichtserkennungstechnologien werden in aller Regel biometrische Daten bearbeitet, die gemäss dem Datenschutzgesetz (DSG) besonders schützenswert sind. Somit ergibt sich für Anwender resp. Betreiber die Pflicht, ein höheres Datenschutzniveau zu gewährleisten – unabhängig davon, ob es sich dabei um öffentliche oder private Akteure handelt. Da in den Anwendungsfeldern jedoch von den Anwendern resp. Betreibern vielfach nicht in rechtskonformer Weise agiert wird, sind Verbesserungen im Bereich der Einhaltung der rechtlichen Vorgaben und der Rechtsdurchsetzung im Falle einer Persönlichkeitsverletzung geboten. Zudem drängt sich eine vermehrte staatliche Aufklärung über die teilweise komplexe Rechtslage auf.

Da der Einsatz durch private Akteure viele Risiken für die Betroffenen birgt, entsteht aus grundrechtlicher Perspektive eine staatliche Schutzpflicht vor Grundrechtsverletzungen durch Private. Diese Schutzpflicht sorgt dafür, dass der Staat die Verwendung dieser Technologie durch Private regulieren sollte.

Die ethischen Erörterungen haben hingegen gezeigt, dass selbst ein rechtskonformer Einsatz neue Herausforderungen mit sich bringen kann. Hier sind insb. die Öffentlichkeit und die Politik gefordert, Herausforderungen wie das wachsende Machtgefälle zwischen Individuen und Staat bzw. Wirtschaft oder die Folgen der zunehmenden Konvergenz digitaler Geräte und Daten frühzeitig zu erkennen und an Lösungen für eine faire und verantwortungsbewusste Gestaltung der sich zunehmend digitalisierenden Gesellschaft zu arbeiten. Gerade im gesellschaftlichen Miteinander schliesst das die Bewahrung wünschenswerter bestehender Normen und ggf. auch die Ausbildung neuer sozialer Normen ein, um zu einer verantwortungsbewussten Nutzung von neuen Technologien zu gelangen. So führt die zunehmende Genauigkeit von Stimm-, Sprach- und Gesichtserkennung bspw. dazu, dass nicht nur die Besitzer z.B. eines smarten Lautsprechers von dessen Sensoren erfasst werden, sondern zunehmend auch andere Menschen in deren Umgebung. Eine verantwortungsbewusste Nutzung müsste somit auch in sozialer Verantwortung gegenüber Betroffenen im persönlichen Bereich münden, indem bspw. deren Einwilligung eingeholt oder Geräte ausgeschaltet werden. Gleichzeitig versprechen Technologien wie die Emotionserkennung die Gewinnung sehr weitgehender Erkenntnisse über die Gedanken- und Gefühlswelt der Betroffenen. Unabhängig von der Genauigkeit der Erkennung drohen bei

ihrem Einsatz schwerwiegende Eingriffe in die Selbstbestimmung. Sollte Emotionserkennung technisch zuverlässig funktionieren, könnten die Ergebnisse zur Verhaltenssteuerung jeglicher Form, bspw. zur Manipulation der Bevölkerung mittels personalisierter und auf die Emotionen eines jeden Einzelnen abgestimmter politischer Werbung, verwendet werden.

Im Rahmen der Analyse der öffentlichen Wahrnehmung wurde klar, dass die Bürgerinnen und Bürger die Vorteile der Technologien durchaus erkennen und zu schätzen wissen. Zugleich wird den Technologien und deren Betreibern misstraut und es treibt die Menschen die Sorge um, dass sie unzuverlässig funktionieren, auf intransparente Weise und für missbräuchliche Zwecke eingesetzt werden könnten. Die Untersuchung der Bevölkerungsmeinung verdeutlichte aber auch erhebliche Wissenslücken und einen damit zusammenhängenden Informationsbedarf der Bevölkerung. Dies spiegelte sich auch in der Befürwortung bzw. Ablehnung der unterschiedlichen Anwendungen wider: Die Jedermann-Identifikation und die Aufmerksamkeitsanalyse in Schulen wurden klar abgelehnt und der Einsatz von Gesichtserkennung in Sportstadien klar befürwortet. In allen anderen Anwendungsfeldern ist hingegen der hohe Anteil der Unentschlossenen auffällig und unterstreicht die Wichtigkeit der Forcierung einer anhaltenden öffentlichen Debatte zu den diskutierten Themen. Wichtig zu beachten in der öffentlichen und politischen Debatte ist auch: Eine Anwendung kann durchaus gesellschaftlich akzeptiert und trotzdem nicht rechtens sein und/oder gesellschaftliche Schäden nach sich ziehen. Und eine Minderheit kann negativ von einer Anwendung betroffen sein, weil die Mehrheit Risiken nicht kennt oder für akzeptabel hält, weil sie selbst nicht betroffen ist.

Mit einer Reihe an Empfehlungen formulieren wir Möglichkeiten, wie Entscheidungstragende aus Politik, Wirtschaft, Datenschutzbehörden, Medien und Zivilgesellschaft den Herausforderungen der Stimm-, Sprach- und Gesichtserkennungstechnologien begegnen könnten. Auf Basis der Erkenntnisse der vorliegenden Studie scheint insb. ein Verbot in drei Bereichen und ein Moratorium in einem Bereich angebracht:

- Verbot automatisierter flächendenkender staatlicher Echtzeitüberwachung und staatlichen Social-Scorings mittels Stimm-, Sprach- oder Gesichtserkennung.
- Verbot vollständig automatisierter Entscheidungen gestützt auf Stimm-, Sprach- und Gesichtserkennungstechnologien in wichtigen Lebensbereichen (z.B. in den Bereichen Gesundheit, Strafverfolgung, Finanzen und Kreditvergabe, Versicherungen, Schul- und Arbeitsumfeld). Die Ergebnisse von teilautomatisierten Entscheidungsunterstützungssystemen sollten stattdessen von geschultem Personal kritisch überprüft und freigegeben werden müssen.
- Verbot der Verknüpfung von Datenbrillen und anderen nicht direkt erkennbaren Technologien, wie kleinen Kameras, mit Gesichtserkennungstechnologie oder der Nutzung entsprechender Geräte in der Öffentlichkeit.
- Moratorium für Emotions- und Krankheitserkennung in wichtigen Lebensbereichen (z.B. in den Bereichen Strafverfolgung, Finanzen und Kreditvergabe, Versicherungen, Schul- und Arbeitsumfeld), solange die genügende technische und organisatorische Zuverlässigkeit und Fairness nicht erwiesen sind, so insb. Verbot des Einsatzes von Emotions- und Krankheitserkennung bei der Stimmauthentifizierung und Bewerberauswahl durch Private sowie Verbot der Aufmerksamkeitserkennung an Schulen. Bestimmte erwünsch-

te Anwendungen könnten auch als Hochrisikoanwendung klassifiziert und damit unter Auflagen erlaubt sein.

Zudem weisen wir darauf hin, dass biometrische Merkmale und insb. die Stimme nicht als (alleiniger) Authentifizierungsfaktor verwendet werden sollten. Denn anders als veränderbare Authentifizierungsmerkmale (wie Passwörter) können körperliche Merkmale nicht geändert werden, wenn sie einmal korrumpiert wurden.

Ansonsten unterteilen sich die Empfehlungen in anwendungsfeldspezifische sowie allgemeine, anwendungsfeldübergreifende Empfehlungen. Letztere sind:

- **Die Regulierung von Hochrisikoanwendungen**, insb. in den Bereichen Gesundheit, Strafverfolgung, Kreditvergabe, Versicherungen, Schul- und Arbeitsumfeld, ist geboten, weil sie ein hohes Risiko für Grundrechte und das gesellschaftliche Zusammenleben bergen.
- **Konkrete gesetzliche Grundlagen für den Einsatz durch öffentliche Stellen** sind erforderlich, da besonders schützenswerte Personendaten bearbeitet werden oder ein Profiling stattfindet.
- **Verbesserung der Rechtsdurchsetzung** durch gesetzliche Anpassungen, etwa der Einführung einer Möglichkeit der Verbandsklage.
- **Angemessene Aus- und Weiterbildung** des Personals, das Anwendungen der Stimm-, Sprach- und Gesichtserkennung bedient, für deren Betrieb verantwortlich oder für die Überprüfung und Freigabe von automatisierten Entscheidungen zuständig ist, sodass ein kritischer und verantwortungsbewusster Umgang mit den Technologien möglich ist.
- **Handreichungen für Betreiber von Stimm-, Sprach- und Gesichtserkennungsanwendungen** zur Unterstützung beim datenschutzrechtskonformen Einsatz ihrer Systeme.
- **Unterstützung für Betroffene**, um sie bei der Wahrnehmung ihrer Betroffenenrechte und ggf. auch der Nutzung von Anwendungen der Stimm, Sprach- und Gesichtserkennung zu unterstützen.
- **Gesellschaftliche Debatte über Vor- und Nachteile** von und den Umgang mit Stimm-, Sprach- und Gesichtserkennungstechnologien.
- **Bereitstellung ausreichender Ressourcen** für vertrauenswürdige Dritte und Medienschaffende.

Digitaltechnologien bewegen sich insgesamt in Richtung einer zunehmend umfassenden Konvergenz verschiedener digitaler Anwendungen, die mit der physischen Welt interagieren. Dies betrifft auch zahlreiche relevante Entwicklungen, für die kein Platz in dieser Studie war, die aber eng mit dem vorliegenden Thema verknüpft sind: Eye Tracking, Objekt- und Bewegungserkennung, Augmented Reality oder die Debatten rund um das sog. «Metaverse». Alle diese und weitere Technologien werden sukzessive miteinander verschmelzen, und Datenbearbeitungen und -flüsse werden noch allgegenwärtiger sein und sich auf noch mehr Bereiche des menschlichen Lebens ausweiten.

Die vorliegende Studie mag einen Anstoss für die öffentliche und politische Debatte geben und zu einem verantwortungsbewussten Einsatz von Stimm-, Sprach- und Gesichtserken-

nungstechnologien beitragen. In dieser Debatte ist es wichtig, auch über die Notwendigkeit von zu ziehenden Grenzen zu sprechen – unseres Erachtens überall dort, wo ein ungerechtfertigter Eingriff oder gar eine Schädigung von Grundrechten der Fall wäre oder Nutzungszwecke von der Gesellschaft abgelehnt werden.

Literatur

Aas, Katja Franko; Gundhus, Helene Oppen; Lomell, Heidi Mork (Hg.) (2009): Technologies of inSecurity. The surveillance of everyday life. Abingdon England: Routledge-Cavendish. Online verfügbar unter https://www.taylorfrancis.com/books/9781134040360.

Ablon, Lillian; Bogart, Andy (2017): Zero Days, Thousands of Nights: The Life and Times of Zero-Day Vulnerabilities and Their Exploits. Santa Monica, CA: RAND Corporation. Online verfügbar unter https://www.rand.org/pubs/research_reports/RR1751.html.

Access Now (2018): National digital identity programmes: what's next? Online verfügbar unter https://www.accessnow.org/national-digital-identity-programmes-whats-next/, zuletzt geprüft am 18.02.2022.

ACLU (2001): ACLU Calls for Public Hearings on Tampa's. In: *American Civil Liberties Union*. Online verfügbar unter https://www.aclu.org/press-releases/aclu-calls-public-hearings-tampas, zuletzt geprüft am 11.03.2021.

Affectiva (2021): Humanizing Technology. Online verfügbar unter https://www.affectiva.com/, zuletzt geprüft am 31.03.2021.

Aichouni, Ahmed Baha Eddine; Kamaruddin, Afiqah; Burhan, Mohamad (2019): Review Paper On Ethics Regarding Biometric Technology.

Akinbi, Alex; Berry, Thomas (2020): Forensic Investigation of Google Assistant. In: *SN COMPUT. SCI.* 1 (5). DOI: 10.1007/s42979-020-00285-x.

Albanesius, Chloe (2021): Amazon's Echo Lineup: What's the Difference? Hg. v. PCMag. Online verfügbar unter https://uk.pcmag.com/smart-home/131384/amazons-echo-lineup-whats-the-difference, zuletzt aktualisiert am 19.03.2021, zuletzt geprüft am 19.03.2021.

AlgorithmWatch CH; Amnesty International; Digitale Gesellschaft (2021): Gesichtserkennung stoppen. Online verfügbar unter https://www.gesichtserkennung-stoppen.ch/, zuletzt aktualisiert am 01.11.2021, zuletzt geprüft am 29.04.2022.

Alhazmi, Hamoud; Imran, Ahmed; Abu Alsheikh, Mohammad (2022): How Do Socio-Demographic Patterns Define Digital Privacy Divide? In: *IEEE Access* 10, S. 11296–11307. DOI: 10.1109/ACCESS.2022.3144436.

Ali, Ahmed; Renals, Steve (2018): Word Error Rate Estimation for Speech Recognition: e-WER. In: Iryna Gurevych und Yusuke Miyao (Hg.): Proceedings of the 56th ACL. Melbourne, Australia: ACL, S. 20–24.

Aloufi, Ranya; Haddadi, Hamed; Boyle, David (2019): Emotionless: Privacy-Preserving Speech Analysis for Voice Assistants. Online verfügbar unter http://arxiv.org/pdf/1908.03632v1.

Amann, Julia; Blasimme, Alessandro; Vayena, Effy; Frey, Dietmar; Madai, Vince I. (2020): Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. In: *BMC medical informatics and decision making* 20 (1), S. 310. DOI: 10.1186/s12911-020-01332-6.

Amazon (2018): How Alexa keeps getting smarter. Online verfügbar unter https://www.aboutamazon.com/devices/how-alexa-keeps-getting-smarter, zuletzt aktualisiert am 22.04.2021, zuletzt geprüft am 08.02.2022.

Amazon Science (2020): How we taught Alexa to correct her own defects. Online verfügbar unter https://www.amazon.science/blog/how-we-taught-alexa-to-correct-her-own-defects, zuletzt aktualisiert am 28.01.2020, zuletzt geprüft am 08.02.2022.

Ammari, Tawfiq; Kaye, Jofish; Tsai, Janice Y.; Bentley, Frank (2019): Music, Search, and IoT. In: *ACM Trans. Comput.-Hum. Interact.* 26 (3), S. 1–28. DOI: 10.1145/3311956.

Andrejevic, Mark; Selwyn, Neil (2020): Facial recognition technology in schools: critical questions and concerns. In: *Learning, Media and Technology* 45 (2), S. 115–128. DOI: 10.1080/17439884.2020.1686014.

Anthes, Emily (2020): Alexa, do I have COVID-19? (7827). Online verfügbar unter https://www.spektrum.de/news/kuenstliche-intelligenz-unterscheidet-stimme-von-gesunden-und-kranken/1777593, zuletzt aktualisiert am 24.11.2020, zuletzt geprüft am 19.04.2021.

AnyVision (Hg.) (2021): AI-driven computer vision for a safer world. Online verfügbar unter https://www.anyvision.co/, zuletzt geprüft am 22.04.2021.

Aptex AG (2021): AptexStadien & Veranstaltungen – Videoüberwachung, Zutritt – Aptex. Aptex AG (blog). Online verfügbar unter https://www.aptex.ch/einsatzbereiche/stadien-veranstaltungen/, zuletzt aktualisiert am 06.10.2021, zuletzt geprüft am 15.06.2021.

Aronson, A. E.; Bless, D. (2011): Clinical Voice Disorders: Thieme. Online verfügbar unter https://books.google.de/books?id=kmugBjlqGBkC.

Ars Technica (2019): AI can diagnose some genetic disorders using photos of faces. Online verfügbar unter https://arstechnica.com/science/2019/01/ai-facial-recognition-can-be-used-to-make-diagnoses-not-just-id-stalkers/, zuletzt aktualisiert am 15.09.2021, zuletzt geprüft am 15.09.2021.

Art. 29 WP (2014): Opinion 8/2014 on Recent Developments on the Internet of Things. Article 29 Working Party.

Arthur, Charles (2013a): Google Glass security failings may threaten owner's privacy. In: *The Guardian*. Online verfügbar unter http://www.theguardian.com/technology/2013/may/01/google-glass-security-privacy-risk, zuletzt geprüft am 04.05.2021.

Arthur, Charles (2013b): Google Glass: is it a threat to our privacy? In: *The Guardian*. Online verfügbar unter http://www.theguardian.com/technology/2013/mar/06/google-glass-threat-to-our-privacy, zuletzt geprüft am 23.04.2021.

ARTICLE 19 (Hg.) (2021): Emotional Entanglement: China's emotion recognition market and its implications for human rights. Online verfügbar unter https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf.

AWS (2018): Alexa Privacy and Data Handling Overview (20180720). Online verfügbar unter https://d1.awsstatic.com/product-marketing/A4B/White%20Paper%20-%20Alexa%20Privacy%20and%20Data%20Handling%20Overview.pdf, zuletzt geprüft am 23.09.2021.

Bacchini, Fabio; Lorusso, Ludovica (2019): Race, again: how face recognition technology reinforces racial discrimination. In: *Journal of Information, Communication and Ethics in Society* 17 (3), S. 321–335. DOI: 10.1108/JICES-05-2018-0050.

Baker, Janet M.; Baker, James K. (1989): Dragon. In: Unknown (Hg.): Proceedings of the workshop on Speech and Natural Language – HLT '89. the workshop. Philadelphia, Pennsylvania, 2/21/1989 – 2/23/1989. Morristown, NJ, USA: Association for Computational Linguistics, S. 143.

Baltrusaitis, Justinas (2019): Top 10 Countries and Cities by Number of CCTV Cameras. Hg. v. PreciseSecurity. Online verfügbar unter https://www.precisesecurity.com/articles/Top-10-Countries-by-Number-of-CCTV-Cameras/, zuletzt aktualisiert am 20.06.2020, zuletzt geprüft am 24.08.2021.

Barazzetti, Gaia; Bühler, Nolwenn; Audétat, Marc; Kaufmann, Alain (2021): Making personalized medicine ethical: a critical examination of the new promises of 'personalized health' in Switzerland. In: *Science and Public Policy* 48 (6), S. 818–828. DOI: 10.1093/scipol/scab051.

Barrett, Lisa Feldman (2017): How Emotions Are Made. The Secret Life of the Brain. Boston: Houghton Mifflin Harcourt. Online verfügbar unter https://ebookcentral.proquest.com/lib/gbv/detail.action?docID=4707479.

Barrett, Lisa Feldman; Adolphs, Ralph; Marsella, Stacy; Martinez, Aleix M.; Pollak, Seth D. (2019): Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. In: *PSPI* 20 (1), S. 1–68. DOI: 10.1177/1529100619832930.

Bärtschi, Simon; Widmer, Joel (2006): Hooligan-Erkennung funktioniert. In: *Sonntagszeitung*, S. 7.

Baumann, Max (2012): Gentest für jedermann und das Recht. In: *Jusletter* (5).

BBC News (2018): 2,000 wrongly matched with possible criminals at Champions League. In: *BBC News*. Online verfügbar unter https://www.bbc.com/news/uk-wales-south-west-wales-44007872, zuletzt geprüft am 16.06.2021.

BdWi (2006): Fußball-WM – datenschutzverträglich? Online verfügbar unter https://www.bdwi.de/forum/archiv/uebersicht/136482.htm?pic=0, zuletzt geprüft am 24.02.2022.

Beck, Rafael; Kistler, Roger; Lanz, Karin (2018): Internes Crowdsourcing zum Trainieren automatischer Spracherkennung in Schweizerdeutsch. Hg. v. HSR Hochschule für Technik Rapperswil. Online verfügbar unter https://eprints.hsr.ch/670/, zuletzt geprüft am 14.12.2020.

Behring, Maria von (2019): Sicherheitslücke: Biometrische Daten von Millionen Menschen offen im Netz. netzpolitik.org. Online verfügbar unter https://netzpolitik.org/2019/sicherheitsluecke-biometrische-daten-von-millionen-menschen-offen-im-netz/, zuletzt aktualisiert am 15.08.2019, zuletzt geprüft am 15.02.2022.

Belser, Eva Maria; Epiney, Astrid; Waldmann, Bernhard (2011): Datenschutzrecht. Grundlagen und öffentliches Recht. Bern: Stämpfli (Stämpflis juristische Lehrbücher).

Belser, Eva Maria; Molinari, Eva (2015): Kommentar zu Art. 7 BV. In: Bernhard Waldmann, Eva Maria Belser und Astrid Epiney (Hg.): Schweizerische Bundesverfassung (BV), Basler Kommentar. Basel.

Bergen, Jennifer (2013): The evolution of 3D: Hands on with Canon's new MREAL System. In: *Digital Trends*. Online verfügbar unter https://www.digitaltrends.com/computing/hands-on-with-canons-new-mreal-system-for-mixed-reality/, zuletzt geprüft am 04.05.2021.

Berle, Ian (2020): Face Recognition Technology. Cham: Springer International Publishing.

Bertelsmann Stiftung; VDE (2020): From Principles to Practice.

Beuth, Patrick (2013): Datenbrille: Verbotszonen für Google Glass. In: *Die Zeit*. Online verfügbar unter https://www.zeit.de/digital/mobil/2013-05/google-glass-verboten, zuletzt geprüft am 04.05.2021.

Beuth, Patrick (2017): iPhone X: Was Apple aus einem Gesicht macht. In: *Zeit ONLINE*, 13.09.2017. Online verfügbar unter https://www.zeit.de/digital/mobil/2017-09/iphone-x-face-id-gesichtserkennung-hacker, zuletzt geprüft am 07.12.2020.

Beuth, Patrick (2019): Amazon Echo und Google Home: Berliner Hacker machen Smart Speaker zu Wanzen. Online verfügbar unter https://www.spiegel.de/netzwelt/gadgets/amazon-echo-und-google-home-apps-machen-smart-speaker-zu-wanzen-a-1292367.html, zuletzt aktualisiert am 20.10.2019, zuletzt geprüft am 24.08.2021.

Beuth, Patrick (2021): Clearview AI: Wie kommt man aus der Biometrie-Datenbank heraus? In: *SPIEGEL*, 07.06.2021. Online verfügbar unter https://www.spiegel.de/netzwelt/web/clearview-ai-wie-kommt-man-aus-der-biometrie-datenbank-heraus-a-10b14385-0002-0001-0000-000177779166, zuletzt geprüft am 10.09.2021.

Biaggini, Giovanni (2017): OFV-Kommentar zu Art. 11. In: Giovanni Biaggini (Hg.): BV. Kommentar: Bundesverfassung der Schweizerischen Eidgenossenschaft. 2., überarbeitete und erweiterte Auflage. Zürich: Orell Füssli Verlag.

Big Brother Awards 2006 (2006): Schlittschuhclub Bern (SCB) und Unisys (Schweiz) AG: Gesichtsbiometrische Kontrollen im Hockeystadion. Online verfügbar unter https://archiv.bigbrotherawards.ch/2006/nomination/nominees/5244-SCB-Unisys.pdf, zuletzt geprüft am 10.03.2021.

Big Brother Watch (2018): FACE Off: The lawless growth of facial recognition in UK policing. Big Brother Watch. Online verfügbar unter https://bigbrotherwatch.org.uk/campaigns/stop-facial-recognition/report/, zuletzt aktualisiert am 16.06.2021, zuletzt geprüft am 16.06.2021.

Bilton, Nick (2012): Google to Sell Heads-Up Display Glasses by Year's End. In: *Bits Blog*. Online verfügbar unter https://bits.blogs.nytimes.com/2012/02/21/google-to-sell-terminator-style-glasses-by-years-end/, zuletzt geprüft am 04.05.2021.

Bischof, Sarah: Datenschutz und Berufsgeheimnis im ambulanten Leistungsbereich. Übermittlung von Rechnungsdaten an die sozialen Krankenversicherer, insbesondere nach TARMED.

Bitirim, Yiltan; Bitirim, Selin; Celik Ertugrul, Duygu; Toygar, Onsen (2020): An Evaluation of Reverse Image Search Performance of Google. In: IEEE COMPSAC. Madrid, Spain: IEEE, S. 1368–1372.

BKA (2007): Das Forschungsprojekt Foto-Fahndung. Abschlussbericht. Online verfügbar unter https://www.bka.de/DE/UnsereAufgaben/Forschung/ForschungsprojekteUndErgebnisse/Foto-Fahndung/fotoFahndung.html, zuletzt geprüft am 30.03.2022.

Bledsoe, Woodrow Wilson (1966): The model method in facial recognition. In: *Panoramic Research Inc.* 15 (47), S. 2.

BMW AG (2021): Amazon Alexa & BMW ConnectedDrive. Online verfügbar unter https://www.bmw.de/de/topics/service-zubehoer/bmw-connecteddrive/amazon-alexa.html, zuletzt geprüft am 19.03.2021.

Bösel, Bernd (2019): Affective Computing. In: Kevin Liggieri und Oliver Müller (Hg.): Mensch-Maschine-Interaktion. Handbuch zu Geschichte – Kultur – Ethik. Stuttgart: J.B. Metzler, S. 223–225.

Bosker, Bianca (2013): SIRI RISING: The Inside Story Of Siri's Origins – And Why She Could Overshadow The iPhone. In: *Huffpost*, 22.01.2013. Online verfügbar unter https://www.huffpost.com/entry/siri-do-engine-apple-iphone_n_2499165, zuletzt geprüft am 08.12.2020.

Bouhlal, M.; Aarika, K.; Abdelouahid, R. Ait; Elfilali, S.; Benlahmar, E. (2020): Emotions recognition as innovative tool for improving students' performance and learning approaches. In: *Procedia Computer Science* 175, S. 597–602. DOI: 10.1016/j.procs.2020.07.086.

Braun Binder, Nadja; Burri, Thomas; Lohmann, Melinda Florina; Simmler, Monika; Thouvenin, Florent; Vokinger, Kerstin Noelle (2021): Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht. In: *Jusletter*. Online verfügbar unter https://www.alexandria.unisg.ch/263381/.

Braun Binder, Nadja; Kunz, Eliane; Obrecht, Liliane (2022): Maschinelle Gesichtserkennung im öffentlichen Raum. In: *sg*. DOI: 10.21257/sg.204.

Brennan, Grace (2020): Emotion Analytics Used in AI Recruitment Tools Are Not Only Unethical But Incorrect. Hg. v. The Sociable. Online verfügbar unter https://sociable.co/technology/emotion-analytics-ai-recruitment-tools-incorrect/, zuletzt geprüft am 31.03.2021.

Brien, Jörn (2018): Echo, Home und Homepod: Schon jeder 10. Deutsche benutzt smarte Lautsprecher. In: *t3n Magazin*. Online verfügbar unter https://t3n.de/news/smarte-lautsprecher-deutschland-1133808/, zuletzt geprüft am 05.03.2021.

Brown, Dalvin (2019): Team of 'white hat' hackers found bugs in Amazon Echo and Galaxy S10. In: *USA Today*, 11.10.2019. Online verfügbar unter https://eu.usatoday.com/story/tech/2019/11/10/team-hackers-found-vulnerabilities-amazon-echo-galaxy-s-10/2555266001/, zuletzt geprüft am 24.06.2021.

Bruhn, Manfred; Burmann, Christoph; Kirchgeorg, Manfred (Hg.) (2020): Marketing Weiterdenken. Zukunftspfade für eine marktorientierte Unternehmensführung. 2. Aufl. Wiesbaden: Springer Gabler.

BSV (2022): Kinderrechte. Online verfügbar unter https://www.bsv.admin.ch/bsv/de/home/sozialpolitische-themen/kinder-und-jugendfragen/kinderrechte.html, zuletzt aktualisiert am 18.02.2022, zuletzt geprüft am 18.02.2022.

Büchler, Andrea; Michel, Margot (2014): Medizin – Mensch – Recht. Eine Einführung in das Medizinrecht der Schweiz. 2. Auflage. Zürich/Basel/Genf.

Bugeja, Joseph (Hg.) (2021): On Privacy and Security in Smart Connected Homes. Malmö University. Online verfügbar unter https://www.researchgate.net/publication/349297209.

Bühlmann, Lukas; Schüepp, Michael (2020): EDÖB zu Clearview App: massenhafte Beschaffung allgemein zugänglicher Gesichtsdaten ist «kaum» datenschutzkonform. Hg. v. MLL AG. Online verfügbar unter https://www.mll-news.com/edoeb-zu-clearview-app-

massenhafte-beschaffung-allgemein-zugaenglicher-gesichtsdaten-ist-kaum-datenschutzkonform/, zuletzt aktualisiert am 13.05.2020, zuletzt geprüft am 30.08.2021.

Bundesamt für Statistik (2019): Internetnutzung in den Haushalten im Jahr 2019. Online verfügbar unter https://www.bfs.admin.ch/bfs/de/home/statistiken/kultur-medien-informationsgesellschaft-sport/erhebungen/omn2019.gnpdetail.2019-0047.html, zuletzt geprüft am 30.04.2021.

Bundeskanzlei, Schweizerische (2017): Botschaft zum Bundesgesetz über die Totalrevision des Bundesgesetzes über den Datenschutz und die Änderung weiterer Erlasse zum Datenschutz (BBl). Online verfügbar unter https://fedlex.data.admin.ch/filestore/fedlex.data.admin.ch/eli/fga/2017/2057/de/pdf-a/fedlex-data-admin-ch-eli-fga-2017-2057-de-pdf-a.pdf, zuletzt geprüft am 13.06.2021.

Bundesrat: Stellungnahme des Bundesrats vom 11.08.2021 zur Interpellation Glättli (21.3580): Regulierung der Gesichtserkennung im öffentlichen Raum. Online verfügbar unter https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20213580, zuletzt geprüft am 27.05.2022.

Bundesrat (2018): Aktionsplan Digitale Schweiz. Hg. v. Bundesamt für Kommunikation BAKOM. Online verfügbar unter https://www.bakom.admin.ch/bakom/de/home/digital-und-internet/strategie-digitale-schweiz.html.

Buolamwini, Joy; Gebru, Timnit (2018): Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In: Conference on Fairness, Accountability and Transparency: PMLR, S. 77–91. Online verfügbar unter http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.

Burch, Felix; Rosch, Benjamin (2015): Verdecktes Filmen von Fussballfans: Deshalb ist «Focus One» so umstritten. Hg. v. watson. Online verfügbar unter https://www.watson.ch/schweiz/hooligans/333412437-verdecktes-filmen-von-fussballfans-deshalb-ist-focus-one-so-umstritten, zuletzt geprüft am 10.09.2021.

Burkhart, Michael; Huesman-Koecke, Sevilay (2021): Künstliche Intelligenz in der Gesundheitswirtschaft. Wie KI zu einer besseren und günstigeren Gesundheitsversorgung beitragen kann. Hg. v. PricewaterhouseCoopers. Online verfügbar unter https://www.pwc.de/de/gesundheitswesen-und-pharma/wie-kuenstliche-intelligenz-das-gesundheitssystem-revolutioniert.html, zuletzt geprüft am 19.04.2021.

Business Wire (2001): Birmingham City Centre CCTV Installs Visionics' FaceIt; Same Crime Fighting Face Recognition Technology Already in Use in London Borough of Newham. In: *www.allbusiness.com*. Online verfügbar unter https://web.archive.org/web/20081220212810/https://www.allbusiness.com/government/government-bodies-offices-regional/6111139-1.html, zuletzt geprüft am 09.03.2021.

Cacioppo, J. T.; Tassinary, L. G. (1990): Inferring psychological significance from physiological signals. In: *The American psychologist* 45 (1), S. 16–28. DOI: 10.1037//0003-066X.45.1.16.

Campbell, Zach; Jones, Chris (2022): Kitchen Appliance Maker Wants to Revolutionize Video Surveillance. In: *The Intercept*, 02.11.2022. Online verfügbar unter https://theintercept.com/2022/02/11/surveillance-video-ai-bosch-azena/, zuletzt geprüft am 31.03.2022.

Canedy, Dana (2001): Tampa Scans the Faces in Its Crowds for Criminals. In: *The New York Times*. Online verfügbar unter https://www.nytimes.com/2001/07/04/us/tampa-scans-the-faces-in-its-crowds-for-criminals.html, zuletzt geprüft am 11.03.2021.

Castelluccia, Claude; Le Métayer, Daniel (2020): Impact Analysis of Facial Recognition: Towards a Rigorous Methodology. Online verfügbar unter https://hal.inria.fr/hal-02480647/document, zuletzt geprüft am 15.02.2022.

Cavazos, Jacqueline G.; Phillips, P. Jonathon; Castillo, Carlos D.; OrToole, Alice J. (2020): Accuracy comparison across face recognition algorithms: Where are we on measuring race bias? In: *IEEE Trans. Biom. Behav. Identity Sci.*, S. 1. DOI: 10.1109/TBIOM.2020.3027269.

Chadha, Neha; Gangwar, R. C.; Bedi, Rajeev (2015): Current Challenges and Application of Speech Recognition Process using Natural Language Processing: A Survey. In: *IJCA* 131 (11), S. 28–31. DOI: 10.5120/ijca2015907471.

Chambers, Lauren (2020): Five things you need to know about a gov't study on face surveillance. Hg. v. ACLU of Massachusetts. Online verfügbar unter https://privacysos.org/blog/five-fast-facts-from-the-federal-study-of-demographic-bias-in-facial-recognition/, zuletzt geprüft am 15.04.2021.

Chazan, Guy (2019): German security agencies want access to home devices. In: *Financial Times*. Online verfügbar unter https://www.ft.com/content/ad765972-87a2-11e9-a028-86cea8523dc2, zuletzt geprüft am 15.06.2021.

Chen, Xiang; Qing, Linbo; He, Xiaohai; Su, Jie; Peng, Yonghong (2018): From Eyes to Face Synthesis: a New Approach for Human-Centered Smart Surveillance. In: *IEEE Access* 6, S. 14567–14575. DOI: 10.1109/ACCESS.2018.2803787.

Chin, Josh (2018): Chinese Police Add Facial-Recognition Glasses to Surveillance Arsenal. In: *The Wall Street Journal*. Online verfügbar unter https://www.wsj.com/articles/chinese-police-go-robocop-with-facial-recognition-glasses-1518004353, zuletzt geprüft am 15.06.2021.

Chiusi, Fabian (2020): In Italy, an appetite for face recognition in football stadiums – AlgorithmWatch. Online verfügbar unter https://algorithmwatch.org/en/italy-stadium-face-recognition, zuletzt aktualisiert am 15.06.2021, zuletzt geprüft am 16.06.2021.

Choi, Woo Yong; Song, Kyu Ye; Lee, Chan Woo (2018): Convolutional Attention Networks for Multimodal Emotion Recognition from Speech and Text Data. In: Amir Zadeh, Paul Pu Liang, Louis-Philippe Morency, Soujanya Poria, Erik Cambria und Stefan Scherer (Hg.): Proceedings of Challenge-HML. Melbourne, Australia. Stroudsburg, PA, USA: ACL, S. 28–34.

Christen, Markus; Mader, Clemens; Čas, Johann; Abou-Chadi, Tarik; Bernstein, Abraham; Braun Binder, Nadja et al. (2020): Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz. Zürich: vdf Hochschulverlag AG an der ETH Zürich.

Christiaan008 (2014): 30C3: Glass Hacks (EN). Online verfügbar unter https://www.youtube.com/watch?v=wlSy1Y7Vkos, zuletzt geprüft am 23.04.2021.

Chung, Hyunji; Park, Jungheum; Lee, Sangjin (2017): Digital forensic approaches for Amazon Alexa ecosystem. In: *Digital Investigation* 22, S15-S25. DOI: 10.1016/j.diin.2017.06.010.

Ciresan, D.; Meier, U.; Schmidhuber, J. (2012): Multi-column deep neural networks for image classification. In: IEEE CVPR. Piscataway, NJ: IEEE, S. 3642–3649.

Clark, Teri (2007): The Complete Personal Finance Handbook. Step-by-Step Instructions to Take Control of Your Financial Future. Ocala: Atlantic Publishing Group.

Cmm360 (2018): PostFinance halbiert die Authentifizierungsdauer mit NICE Real-Time Authentication. Online verfügbar unter https://www.cmm360.ch/postfinance-mit-nice-real-time-authentication, zuletzt aktualisiert am 18.02.2022, zuletzt geprüft am 18.02.2022.

CoE (2021): Consultative committee of the convention for the protection of individuals with regard to automatic processing of personal data: Guidelines on Facial Recognition. T-PD(2020)03rev4 (T-PD(2020)03rev4). Online verfügbar unter https://rm.coe.int/guidelines-on-facial-recognition/1680a134f3, zuletzt aktualisiert am 2021.

Colantonio, Sara; Germanese, Danila; Moroni, Davide; Giorgi, Daniela; Pascali, Mariantonietta; Righi, Marco et al. (2015): Semeoticons – reading the face code of cardio-metabolic risk. In: IWCIM. Prague, Czech Republic. Piscataway, NJ: IEEE, S. 1–5.

Corcoran, Cheryl M.; Carrillo, Facundo; Fernández-Slezak, Diego; Bedi, Gillinder; Klim, Casimir; Javitt, Daniel C. et al. (2018): Prediction of psychosis across protocols and risk cohorts using automated language analysis. In: *WPA* 17 (1), S. 67–75. DOI: 10.1002/wps.20491.

Council of Europe (Hg.) (1981): Explanatory Report to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (108). Online verfügbar unter https://rm.coe.int/16800ca434, zuletzt geprüft am 15.02.2022.

Cowie, R.; Douglas-Cowie, E.; Tsapatsoulis, N.; Votsis, G.; Kollias, S.; Fellenz, W.; Taylor, J. G. (2001): Emotion recognition in human-computer interaction. In: *IEEE Signal Process. Mag.* 18 (1), S. 32–80. DOI: 10.1109/79.911197.

Crawford, Kate (2021): Time to regulate AI that interprets human emotions. In: *Nature* 592 (7853), S. 167. DOI: 10.1038/d41586-021-00868-5.

Crawford, Kate; Dobbe, Roel; Dryer, Theodora; Fried, Genevieve; Green, Ben; Kaziunas, Elizabeth et al. (2019): AINOW Report. Hg. v. AI Now Institute. NewYork. Online verfügbar unter https://ainowinstitute.org/AI_Now_2019_Report.pdf.

Cui, Liqing; Li, Shun; Zhu, Tingshao (2015): Emotion detection from natural walking.

Cummins, Nicholas; Ren, Zhao; Mallol-Ragolta, Adria; Schuller, Björn (2020): Machine learning in digital health, recent trends, and ongoing challenges. In: Debmalya Barh (Hg.): Artificial intelligence in precision health. From concept to applications. London: Elsevier, S. 121–148.

Cummins, Nicholas; Scherer, Stefan; Krajewski, Jarek; Schnieder, Sebastian; Epps, Julien; Quatieri, Thomas F. (2015): A review of depression and suicide risk assessment using speech analysis. In: *Speech Communication* 71, S. 10–49. DOI: 10.1016/j.specom.2015.03.004.

D'Mello, Sidney K. (2017): Emotional Learning Analytics. In: Charles Lang, George Siemens, Alyssa Wise und Dragan Gasevic (Hg.): Handbook of Learning Analytics. Society for Learning Analytics Research (SoLAR), S. 115–127.

Damavandi, Babak; Kumar, Shankar; Shazeer, Noam; Bruguier, Antoine (2016): NN-Grams: Unifying Neural Network and n-Gram Language Models for Speech Recognition. In: Interspeech, 8-12 Sep 2016: ISCA (Interspeech), S. 3499–3503.

Das, S. K.; Picheny, M. A. (1996): Issues in Practical Large Vocabulary Isolated Word Recognition: The IBM Tangora System. In: Chin-Hui Lee, Frank K. Soong und Kuldip K. Paliwal (Hg.): Automatic Speech and Speaker Recognition. Advanced Topics, Bd. 355. Boston, MA: Springer, S. 457–479.

Datenschutzbeauftragter des Kantons Zürich: Tätigkeitsbericht 2001. Zürich.

Datenschutzbeauftragter des Kantons Zürich: Tätigkeitsbericht 2002. Zürich.

Dave, Sheinin; Hume, Mike (2016): When fans get banned for life from sports stadiums. In: *The Washington Post*. Online verfügbar unter https://www.washingtonpost.com/news/sports/wp/2016/10/07/when-fans-get-banned-for-life-from-sports-stadiums/, zuletzt geprüft am 15.06.2021.

Deep Impact (2018): Stadion Überwachung bei Beşiktaş Istanbul. Online verfügbar unter https://www.deep-impact.ch/de/blog/stadion-ueberwachung-bei-besiktas-istanbul, zuletzt geprüft am 20.04.2021.

Deep Impact (2019): Stadion Überwachung bei Beşiktaş Istanbul. Online verfügbar unter https://www.deep-impact.ch/de/blog/stadion-ueberwachung-bei-besiktas-istanbul, zuletzt aktualisiert am 15.06.2021, zuletzt geprüft am 15.06.2021.

Deep Impact (2021a): Ava-X: Gesichtserkennung – jetzt mit Fiebermessung! Online verfügbar unter https://www.deep-impact.ch/de/blog/avax-fiebermesser, zuletzt aktualisiert am 16.06.2021, zuletzt geprüft am 16.06.2021.

Deep Impact (2021b): Sentinel | AVA-X. Online verfügbar unter http://www.ava-x.ai/sentinel/, zuletzt aktualisiert am 20.04.2021, zuletzt geprüft am 20.04.2021.

Dellwo, Volker; French, Peter; He, Lei (2018): Voice Biometrics for Forensic Speaker Recognition Applications. In: Sascha Frühholz und Pascal Belin (Hg.): The Oxford handbook of voice perception. First edition, impression: 1. Oxford: Oxford University Press, S. 777–798.

Dellwo, Volker; Pellegrino, Elisa; He, Lei; Kathiresan, Thayabaran (2019): The dynamics of indexical information in speech: Can recognizability be controlled by the speaker? In: *AUC PHILOLOGICA* (2), S. 57–75. DOI: 10.14712/24646830.2019.18.

Dempsey, P. (2017): The Teardown: Google Home personal assistant. In: *Engineering & Technology* 12 (3), S. 80–81. DOI: 10.1049/et.2017.0330.

Deng, Jun; Cummins, Nicholas; Schmitt, Maximilian; Qian, Kun; Ringeval, Fabien; Schuller, Björn (2017): Speech-based Diagnosis of Autism Spectrum Condition by Generative Adversarial Network Representations. In: Patty Kostkova und A. Special Interest Group on Knowledge DiscoveryC.M. in Data (Hg.): Proceedings of the 2017 ICDH. London, United Kingdom: ACM, S. 53–57.

Desoi, Monika (2018): Intelligente Videoüberwachung. Rechtliche Bewertung und rechtsgemäße Gestaltung. Dissertation, Wiesbaden.

Dewan, M. Ali Akber; Murshed, Mahbub; Lin, Fuhua (2019): Engagement detection in online learning: a review. In: *Smart Learn. Environ.* 6 (1). DOI: 10.1186/s40561-018-0080-z.

Dhall, Abhinav; Goecke, Roland; Ghosh, Shreya; Joshi, Jyoti; Hoey, Jesse; Gedeon, Tom (2017): From individual to group-level emotion recognition: EmotiW 5.0. In: Edward Lank und A. Special Interest Group on Computer-HumanC.M. Interaction (Hg.): Proceedings of the 19th ACM International Conference on Multimodal Interaction. ICMI '17: INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION. Glasgow UK, 13 11 2017 - 17 11 2017. [Place of publication not identified]: ACM, S. 524–528.

Diggelmann, Oliver (2015): Kommentar zu Art. 13 BV. In: Bernhard Waldmann, Eva Maria Belser und Astrid Epiney (Hg.): Schweizerische Bundesverfassung (BV), Basler Kommentar. Basel.

Doffman, Zak (2018): Why Facial Recognition In Schools Seems To Be An Aimless Recipe For Disaster. Hg. v. Forbes. Online verfügbar unter https://www.forbes.com/sites/zakdoffman/2018/11/07/why-facial-recognition-in-schools-seems-to-be-an-aimless-recipe-for-disaster/, zuletzt geprüft am 07.04.2021.

doitvoluntarily (2018): The Future Of The Workplace: Using Facial Recognition To Scan For Smiles. Hg. v. steemit. Online verfügbar unter https://steemit.com/news/@doitvoluntarily/the-future-of-the-workplace-using-facial-recognition-to-scan-for-smiles, zuletzt aktualisiert am 26.06.2018, zuletzt geprüft am 29.04.2022.

Donahue, Chris; McAuley, Julian; Puckette, Miller (2018): Adversarial Audio Synthesis. Online verfügbar unter http://arxiv.org/pdf/1802.04208v3.

Donath, Andreas (2013): Datenschutz: Gesichtserkennung mit Google Glass verboten – Golem.de. In: *Golem.de*. Online verfügbar unter https://www.golem.de/news/datenschutz-gesichtserkennung-mit-google-glass-verboten-1306-99556.html, zuletzt geprüft am 04.05.2021.

Donath, Andreas (2014): Datenbrille: Google Glass erscheint nicht mehr dieses Jahr – Golem.de. In: *Golem.de*. Online verfügbar unter https://www.golem.de/news/datenbrille-google-glass-erscheint-nicht-mehr-dieses-jahr-1411-110578.html, zuletzt geprüft am 04.05.2021.

dpa (2015): «Glass»: Google stoppt den Verkauf seiner Datenbrille. In: *FAZ.net*. Online verfügbar unter https://www.faz.net/aktuell/wirtschaft/netzwirtschaft/google/google-glass-neuanfang-fuer-die-datenbrille-13372678.html, zuletzt geprüft am 04.05.2021.

Dpa, Andreas Landwehr (2018): China schafft digitales Punktesystem für den «besseren» Menschen. In: *Heise Online*. Online verfügbar unter https://www.heise.de/newsticker/meldung/China-schafft-digitales-Punktesystem-fuer-den-besseren-Menschen-3983746.html, zuletzt geprüft am 15.06.2021.

Dubois, Daniel J.; Kolcun, Roman; Mandalari, Anna Maria; Paracha, Muhammad Talha; Choffnes, David; Haddadi, Hamed (2020): When Speakers Are All Ears: Characterizing Misactivations of IoT Smart Speakers. In: *Proceedings on Privacy Enhancing Technologies* (4), S. 255–276. DOI: 10.2478/popets-2020-0072.

EDÖB (Hg.) (2005): Einsatz von Biometrie beim Check-In und Boarding im Rahmen des Pilotprojektes «Secure Check» der Swissport International AG und Checkport Schweiz AG am Flughafen Zürich-Kloten. Schlussbericht. Bern. Online verfügbar unter https://www.edoeb.admin.ch/dam/edoeb/de/dokumente/2006/01/schlussbericht_checkport.pdf.download.pdf/schlussbericht_checkport.pdf, zuletzt geprüft am 17.05.2022.

EDÖB (2006): Drogentests in der Lehre. Online verfügbar unter https://www.edoeb.admin.ch/edoeb/de/home/datenschutz/arbeitsbereich/drogentests-in-der-lehre.html, zuletzt aktualisiert am 30.08.2021, zuletzt geprüft am 30.08.2021.

EDÖB (2009): 16. Tätigkeitsbericht 2008/2009. Online verfügbar unter https://www.edoeb.admin.ch/edoeb/de/home/dokumentation/taetigkeitsberichte/aeltere-berichte/16--taetigkeitsbericht-2008-2009.html, zuletzt geprüft am 27.05.2022.

EDÖB (2015): Pay as you drive (PAYD): Erläuterungen zum Einsatz von Black Boxes in Motorfahrzeugen. Online verfügbar unter https://www.edoeb.admin.ch/edoeb/de/home/datenschutz/versicherungen/fahrzeugversicherungen/pay-as-you-drive--payd---erlaeuterungen-zum-einsatz-von-black-bo.html, zuletzt geprüft am 27.09.2021.

EDÖB (2016): Erläuterungen zum Einsatz von Fitnesstrackern im Versicherungsbereich. Online verfügbar unter https://www.edoeb.admin.ch/edoeb/de/home/datenschutz/gesundheit/kranken--und-unfallversicherungen/erlaeuterungen-zum-einsatz-von-fitness-trackern-im-versicherungsb.html, zuletzt aktualisiert am 27.09.2021, zuletzt geprüft am 27.09.2021.

EDÖB (2017): Erläuterungen zu Stimmerkennungsverfahren. Online verfügbar unter https://www.edoeb.admin.ch/edoeb/de/home/datenschutz/technologien/biometrie/erlaeuterungen-zu-stimmerkennungsverfahren.html, zuletzt geprüft am 18.02.2022.

EDÖB (2020a): Privacy Shield CH-USA bietet nach Auffassung des EDÖB kein adäquates Datenschutzniveau. Online verfügbar unter https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-80318.html, zuletzt aktualisiert am 15.06.2021, zuletzt geprüft am 15.06.2021.

EDÖB (2020b): Stellungnahme zur Applikation «Clearview». Online verfügbar unter https://www.edoeb.admin.ch/edoeb/de/home/aktuell/aktuell_news.html#215418784, zuletzt aktualisiert am 11.02.2020, zuletzt geprüft am 30.08.2021.

EDÖB (2022): Personentracking. Online verfügbar unter https://www.edoeb.admin.ch/edoeb/de/home/datenschutz/technologien/personentracking.html, zuletzt aktualisiert am 27.05.2022.

EDPB (2021a): Guidelines 02/2021 on Virtual Voice Assistants Version 1.0. Online verfügbar unter https://edpb.europa.eu/system/files/2021-03/edpb_guidelines_022021_virtual_voice_assistants_adopted-public-consultation_en.pdf, zuletzt geprüft am 15.06.2021.

EDPB (2021b): Guidelines 02/2021 on Virtual Voice Assistants. Hg. v. European Data Protection Board. Online verfügbar unter https://edpb.europa.eu/our-work-tools/documents/public-consultations/2021/guidelines-022021-virtual-voice-assistants_de, zuletzt geprüft am 24.06.2021.

EDPB-EDPS (Hg.) (2021): Joint Opinion 5/2021 on the proposal for a Regulation of the Euro-pean Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Online verfügbar unter https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf, zuletzt geprüft am 29.08.2021.

EDPS (2015): Opinion 4/2015 – Towards a New Digital Ethics: Data, Dignity and Technology. Brussels: European Data Protection Supervisor. Online verfügbar unter https://

edps.europa.eu/sites/edp/files/publication/15-09-11_data_ethics_en.pdf, zuletzt geprüft am 10.09.2021.

EDPS (2018): EDPS Opinion on online manipulation and personal data. Online verfügbar unter https://edps.europa.eu/sites/edp/files/publication/18-03-19_online_manipulation_en.pdf, zuletzt geprüft am 30.03.2022.

EDPS (2019): Technology report No 1. Smart glasses and data protection. Online verfügbar unter https://edps.europa.eu/sites/edp/files/publication/19-01-18_edps-tech-report-1-smart_glasses_en.pdf.

EDRi (2020): Ban biometric mass surveillance! Online verfügbar unter https://edri.org/our-work/blog-ban-biometric-mass-surveillance/, zuletzt aktualisiert am 26.11.2021, zuletzt geprüft am 15.02.2022.

Edu, Jide S.; Such, Jose M.; Suarez-Tangil, Guillermo (2021): Smart Home Personal Assistants. In: *ACM Comput. Surv.* 53 (6), S. 1–36. DOI: 10.1145/3412383.

educa.ch (Hg.) (2009): Datenschutz. Sicherer Umgang mit Personendaten, Schule und ICT. Online verfügbar unter https://biblio.educa.ch/sites/default/files/guide_datenschutz.pdf, zuletzt geprüft am 29.08.2021.

Eick, Volker (2011): Lack of Legacy? Shadows of Surveillance after the 2006 FIFA World Cup in Germany. In: *Urban Studies* 48 (15), S. 3329–3345. DOI: 10.1177/0042098011422389.

Ekman, Paul (1978): Facial action coding system. Palo Alto, CA: Consulting Psychologists Press. Online verfügbar unter http://worldcatlibraries.org/wcpa/oclc/605256401.

Ekman, Paul; Mania, Hubert; Havener, Thorsten (2011): Ich weiss, dass du lügst. Was Gesichter verraten. Reinbek bei Hamburg: Rowohlt Verlag.

Ekman, Paul; O'Sullivan, Maureen (1991): Who can catch a liar? In: *The American psychologist* 46 (9), S. 913–920. DOI: 10.1037/0003-066X.46.9.913.

Elfenbein, Hillary Anger; Ambady, Nalini (2002): On the universality and cultural specificity of emotion recognition: a meta-analysis. In: *Psychological bulletin* 128 (2), S. 203–235. DOI: 10.1037/0033-2909.128.2.203.

El-Seoud, Samir Abou; Ahmed, Samaa A. (2019): IQ and EQ Enhancement for People with Mental Illness. In: Michael E. Auer und Thrasyvoulos Tsiatsos (Hg.): The Challenges of the Digital Transformation in Education. Proceedings of the 21st International Conference on Interactive Collaborative Learning (ICL2018) – Volume 2. Cham, 2019. Cham: Springer International Publishing (Advances in Intelligent Systems and Computing, 917), S. 197–209.

Emmenegger, Susan (2019): Biometrische Daten im Bankkundenverkehr am Beispiel der Stimmauthentifizierung. In: Susan Emmenegger (Hg.): Banken und Datenschutz. SBT 2019 – Schweizerische Bankrechtstagung 2019.

Emmenegger, Susan; Reber, Martina (2019): Biometrische Daten im Bankkundenverkehr am Beispiel der Stimmauthentifizierung. In: Susan Emmenegger (Hg.): Banken und Datenschutz. SBT 2019 – Schweizerische Bankrechtstagung 2019.

Epiney, Astrid (2015): Kommentar zu Art. 36 BV. In: Bernhard Waldmann, Eva Maria Belser und Astrid Epiney (Hg.): Schweizerische Bundesverfassung (BV), Basler Kommentar. Basel.

Epiney, Astrid (2016): Staatliche Überwachung versus Rechtsstaat: Wege aus dem Dilemma? Online verfügbar unter https://doc.rero.ch/record/308954/files/Aufsatz_163.pdf, zuletzt geprüft am 15.06.2021.

Epiney, Astrid; Nüesch, Daniela (2015): § 3 Prinzipien der Datenbearbeitung durch Privatpersonen und Behörden. In: Nicolas Passadelis (Hg.): Datenschutzrecht. Beraten in Privatwirtschaft und öffentlicher Verwaltung. Basel: Helbing Lichtenhahn (Handbücher für die Anwaltspraxis).

Erziehungsdirektion des Kantons Bern, Amt für Kindergarten, Volksschule und Beratung (Hg.): Datenschutz in den Schulen des Kantons Bern. Leitfaden (Nachschlagewerk). Online verfügbar unter https://www.erz.be.ch/dam/documents/ERZ/AKVB/de/09_Schulleitungen_Lehrpersonen/sl_lp_Unterlagen_datenschutz_leitfaden_d.pdf, zuletzt geprüft am 29.08.2021.

Europäische Kommission (01.04.2021): Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz und zur Änderung. 52021PC0206. Online verfügbar unter https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A52021PC0206, zuletzt geprüft am 29.08.2021.

European Commission (Hg.) (1994): Speech Understanding and Dialogue. SUNDIAL Project. Publication Office/CORDIS. Online verfügbar unter https://cordis.europa.eu/project/id/2218, zuletzt aktualisiert am 26.11.2020, zuletzt geprüft am 08.12.2020.

European Commission (2021): Proposal for a Regulation of the European Parliament and of the Council lay-ing down harmonised rules on artificial intelligence (artificial intelligence act) an demanding certain union legislative acts. Online verfügbar unter https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206, zuletzt aktualisiert am 15.06.2021, zuletzt geprüft am 15.06.2021.

Evers-Wölk, Michaela; Oertel, Britta; Sonk, Matthias (2018): Gesundheits-Apps. Innovationsanalyse. Unter Mitarbeit von Mattis Jacobs.

Face2Gene (2021): Face2Gene. Online verfügbar unter https://www.face2gene.com/, zuletzt geprüft am 15.09.2021.

Fasel, B.; Luettin, Juergen (2003): Automatic facial expression analysis: a survey. In: *Pattern Recognition* 36 (1), S. 259–275. DOI: 10.1016/S0031-3203(02)00052-3.

Fedotov, Dmitrii; Matsuda, Yuki; Minker, Wolfgang (2019): From Smart to Personal Environment: Integrating Emotion Recognition into Smart Houses. In: IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops 2019). Kyoto, Japan, 11–15 March 2019. Piscataway, NJ: IEEE, S. 943–948.

Fehr, Reto (2019): Psychische Gesundheit in der Schweiz – ein Überblick in Zahlen. Online verfügbar unter https://www.watson.ch/schweiz/wissen/131610675-psychische-krankheiten-in-der-schweiz-diese-zahlen-musst-du-kennen, zuletzt geprüft am 23.04.2021.

Feldstein, Steven (2019): The Global Expansion of AI Surveillance. Online verfügbar unter https://carnegieendowment.org/files/WP-Feldstein-AISurveillance_final1.pdf, zuletzt geprüft am 15.06.2021.

Fichter, Adrienne (2019): Jedes Gesicht in Sekunden identifiziert. Hg. v. Republik. Online verfügbar unter https://www.republik.ch/2019/10/29/360-ueberwachung-made-in-turkey-jedes-gesicht-in-sekunden-identifiziert, zuletzt geprüft am 24.08.2021.

Fight for the Future (2021): Open Letter: banning government use of facial recognition surveillance is not enough, we must ban corporate and private use as well. Online verfügbar unter https://www.fightforthefuture.org/, zuletzt aktualisiert am 14.06.2021, zuletzt geprüft am 15.06.2021.

Fiske, Amelia; Henningsen, Peter; Buyx, Alena (2019): Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. In: *Journal of medical Internet research* 21 (5). DOI: 10.2196/13216.

Fjeld, Jessica; Achten, Nele; Hilligoss, Hannah; Nagy, Adam; Srikumar, Madhulika (2020): Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. In: *SSRN Journal*. DOI: 10.2139/ssrn.3518482.

Floridi, Luciano; Taddeo, Mariarosaria (2016): What is data ethics? In: *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 374 (2083). DOI: 10.1098/rsta.2016.0360.

Foltynova, Kristyna (2021): We See You! How Russia Has Expanded Its Video-Surveillance System. Online verfügbar unter https://www.rferl.org/a/russia-video-surveillance/31052482.html, zuletzt geprüft am 15.06.2021.

Forensic Focus (Hg.) (2019): Griffeye Analyze DI Pro. Online verfügbar unter https://www.forensicfocus.com/reviews/griffeye-analyze-di-pro/, zuletzt aktualisiert am 24.05.2019, zuletzt geprüft am 22.04.2021.

FRA (2019): Facial recognition technology: fundamental rights considerations n the context of law enforcement. Online verfügbar unter https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper-1_en.pdf, zuletzt geprüft am 15.06.2021.

Fraser, Kathleen C.; Meltzer, Jed A.; Rudzicz, Frank (2016): Linguistic Features Identify Alzheimer's Disease in Narrative Speech. In: *Journal of Alzheimer's disease : JAD* 49 (2), S. 407–422. DOI: 10.3233/JAD-150520.

Fraunhofer IIS (2014): Weltpremiere mit Durchblick: Fraunhofer IIS zeigt erste App zur Emotionserkennung auf Google Glass. Presseinformation. Fraunhofer-Institut für Integrierte Schaltungen IIS. Online verfügbar unter https://www.iis.fraunhofer.de/de/pr/2014/20140827_BS_Shore_Google_Glass.html, zuletzt geprüft am 23.04.2021.

Freitas-Magalhães, A. (2018): Facial Action Coding System--Manual of Scientific Codification of the Human Face. Porto, Portugal: FEELab Science Books.

Frid, Alex; Kantor, Ariel; Svechin, Dimitri; Manevitz, Larry M. (2016): Diagnosis of Parkinson's disease from continuous speech using deep convolutional networks without manual selection of features. In: IEEE International Conference on the Science of Electrical Engineering (ICSEE). 16–18 Nov. 2016. Eilat, Israel. Piscataway, NJ: IEEE, S. 1–4.

Friedewald, Michael; Bieker, Felix; Obersteller, Hannah; Nebel, Maxi; Martin, Nicholas; Rost, Martin; Hansen, Marit (2017): DATENSCHUTZ-FOLGENABSCHÄTZUNG. Ein Werkzeug für einen besseren Datenschutz. Hg. v. Fraunhofer-Institut für System-und Innovationsforschung ISI. Karlsruhe. Online verfügbar unter https://www.forum-privatheit.

de/wp-content/uploads/Forum_Privatheit_White_Paper_DSFA-3.pdf, zuletzt geprüft am 10.09.2021.

Frost & Sullivan (2019): Global Mega Trends to 2030 Futurecasting Key Themes that will Shape Our Future Lives. Global 360º Research Team at Frost & Sullivan (K1D4-MT).

Frost & Sullivan (Hg.) (2021): Innovations And Growth Opportunities In Advanced Medical Devices, Medical Imaging, And Digital Health Solutions. Online verfügbar unter https://store.frost.com/innovations-and-growth-opportunities-in-advanced-medical-devices-medical-imaging-and-digital-health-solutions.html, zuletzt aktualisiert am 18.02.2022, zuletzt geprüft am 18.02.2022.

Fuchs, Katharina Anna (2014): Emotionserkennung und Empathie. Eine multimethodale psychologische Studie am Beispiel von Psychopathie und sozialer Ängstlichkeit. Wiesbaden: Springer VS.

Fuchs, Philippe; Giovanettoni, Marco (2013): Apps als Medizinprodukte – und die Folgen davon. In: *Jusletter*.

Fulterer, Ruth (2021): RAY-Ban Stories: Die smarte Brille von Facebook ist eine Kamera. In: *NZZ*, 14.09.2021.

futurezone (2020): Russland nutzt Gesichtserkennung zur Corona-Bekämpfung. In: *futurezone.at*. Online verfügbar unter https://futurezone.at/netzpolitik/russland-nutzt-gesichtserkennung-zur-corona-bekaempfung/400804994, zuletzt geprüft am 15.06.2021.

Gan, Nectar (2020): China is installing surveillance cameras outside people's front doors … and sometimes inside their homes – CNN. Hg. v. CNN. Online verfügbar unter https://edition.cnn.com/2020/04/27/asia/cctv-cameras-china-hnk-intl/index.html, zuletzt aktualisiert am 15.06.2021, zuletzt geprüft am 15.06.2021.

Gao, Shangbin; Ye, Liang (2019): A Physical and Verbal Bullying Detecting Algorithm Based on K-NN for School Bullying Prevention. In: Shuai Han, Liang Ye und Weixiao Meng (Hg.): Artificial Intelligence for Communications and Networks. First EAI International Conference, AICON 2019, Harbin, China, May 25–26, 2019, Proceedings, Part II. Cham: Springer (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering), S. 150–157.

Gates, Kelly (2011): Our biometric future. Facial recognition technology and the culture of surveillance. New York: New York University Press (Critical cultural communication). Online verfügbar unter http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&AN=1021018.

GDS (2020): Strategie Digitale Schweiz – Arbeitsgruppe Künstliche Intelligenz. Geschäftsstelle Digitale Schweiz GDS. Online verfügbar unter https://www.digitaldialog.swiss/de/arbeitsgruppe-kunstliche-intelligenz, zuletzt aktualisiert am 10.09.2021, zuletzt geprüft am 10.09.2021.

Geminn, Christian (2021): Die Regulierung Künstlicher Intelligenz: Anmerkungen zum Entwurf eines Artificial Intelligence Act. In: *Zeitschrift Datenschutz (ZD)* (7), S. 354–359.

Geneve Aeroport (Hg.) (2019): Genève Aéroport tests automated border control. Online verfügbar unter https://www.gva.ch/en/Site/Geneve-Aeroport/News/2020-2016/Geneve-Aeroport-teste-controle-automatise, zuletzt aktualisiert am 24.08.2021, zuletzt geprüft am 24.08.2021.

Gershgorn, Dave (2020): This Simple Facial Recognition Search Engine Can Track You Down Across the Internet. In: *Medium*. Online verfügbar unter https://onezero.medium.com/this-simple-facial-recognition-search-engine-can-track-you-down-across-the-internet-518c7129e454, zuletzt geprüft am 12.05.2021.

Geuter, Ulfried (2015): Körperpsychotherapie. Grundriss einer Theorie für die klinische Praxis. Berlin: Springer (Psychotherapie). Online verfügbar unter https://books.google.de/books?id=pkDzBgAAQBAJ.

Ghassemi, Marzyeh; Oakden-Rayner, Luke; Beam, Andrew L. (2021): The false hope of current approaches to explainable artificial intelligence in health care. In: *The Lancet Digital Health* 3 (11), e745-e750. DOI: 10.1016/S2589-7500(21)00208-9.

Gilliom, John (2011): A Response to Bennett's 'In Defence of Privacy'. In: *SS* 8 (4), S. 500–504. DOI: 10.24908/ss.v8i4.4186.

Gisolf, Floris; Geradts, Zeno; Worring, Marcel (2020): Analysing large and complex image collections during a safety investigation. Online verfügbar unter https://isasi.org/documents/library/technical-papers/2019/tues/parallel/5.%20analysing%20large%20and%20complex%20image%20collections%20during%20a%20safety%20investigation.pdf.

Glaus, Christian (2019): Luzerner Polizei setzt auf künstliche Intelligenz. In: *Luzerner Zeitung*. Online verfügbar unter https://www.luzernerzeitung.ch/zentralschweiz/luzern/luzerner-polizei-setzt-auf-kuenstliche-intelligenz-ld.1128325, zuletzt geprüft am 15.06.2021.

Glenn, Meghan; Strassel, Stephanie; Lee, Haejoong; Maeda, Kazuaki; Zakhary, Ramez; Li, Xuansong (2010): Transcription Methods for Consistency, Volume and Efficiency. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). Online verfügbar unter http://www.lrec-conf.org/proceedings/lrec2010/pdf/849_Paper.pdf.

Godino-Llorente, J. I.; Shattuck-Hufnagel, S.; Choi, J. Y.; Moro-Velázquez, L.; Gómez-García, J. A. (2017): Towards the identification of Idiopathic Parkinson's Disease from the speech. New articulatory kinetic biomarkers. In: *PLOS ONE* 12 (12), e0189583. DOI: 10.1371/journal.pone.0189583.

Golla, Sebastian J. (2015): Arzt, Patient und Assistenzsystem. In: *InTeR*, S. 194–197. Online verfügbar unter https://online.ruw.de/suche/inter/Arzt-Patient-und-Assistenzsystem-75231cd9bb3054ff33a91f0b9c7673df?crefresh=1, zuletzt geprüft am 30.08.2021.

Gong, Ya-Yun; Tang, Xiao-Yu; Liu, Si-Rui (2020): Research on Evaluation Method of Primary School Science Teaching Based on Students' Emotion. In: International Conference on Artificial Intelligence and Education. ICAIE 2020 : proceedings : Tianjin, China, 26–28 June 2020. Los Alamitos, CA: IEEE Computer Society, Conference Publishing Services, S. 392–397.

Goode, Alan (2018): Biometrics for banking: best practices and barriers to adoption. In: *Biometric Technology Today* 2018 (10), S. 5–7. DOI: 10.1016/S0969-4765(18)30156-5.

Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S. et al. (2014): Generative adversarial nets. In: Advances in Neural Information Processing Systems (3), S. 2672–2680. Online verfügbar unter https://www.scopus.com/inward/

record.uri?eid=2-s2.0-84937849144&partnerID=40&md5=6441b2a288c5fdded2adbcc8b21e092c.

Goodfellow, Ian J.; Erhan, Dumitru; Carrier, Pierre Luc; Courville, Aaron; Mirza, Mehdi; Hamner, Ben et al. (2013): Challenges in Representation Learning: A Report on Three Machine Learning Contests. In: Minho Lee, Akira Hirose, Zeng-Guang Hou und Rhee Man Kil (Hg.): Neural information processing. 20th International Conference, ICONIP 2013, Daegu, Korea, November 3-7, 2013; proceedings, Bd. 8228. Berlin: Springer (Lecture notes in computer science, 8228), S. 117–124.

Google Inc. (Hg.) (2021a): Adjust how sensitive Google Assistant is to «Hey Google». Online verfügbar unter https://support.google.com/assistant/answer/9712065?hl=en, zuletzt geprüft am 10.03.2021.

Google Inc. (2021b): Assistant. Datenschutz als Priorität. Online verfügbar unter https://safety.google/intl/de_de/assistant/, zuletzt aktualisiert am 04.03.2021, zuletzt geprüft am 11.03.2021.

Google Inc. (Hg.) (2021c): Discover our data center locations. Online verfügbar unter https://www.google.com/about/datacenters/locations/, zuletzt aktualisiert am 12.02.2021, zuletzt geprüft am 11.03.2021.

Google Inc. (Hg.) (2021d): FAQs about accounts for the Nest app – Android – Google Nest Help. Online verfügbar unter https://support.google.com/googlenest/answer/9297676?co=GENIE.Platform%3DAndroid&hl=en#zippy=%2Cdo-i-need-to-set-up-an-account-to-use-my-nest-energy-safety-and-security-products, zuletzt geprüft am 24.03.2021.

Google Inc. (Hg.) (2021e): Google Nest-Hilfe: Datensicherheit und Datenschutz auf Geräten, die mit Google Assistant kompatibel sind. Online verfügbar unter https://support.google.com/googlenest/answer/7072285?hl=de, zuletzt geprüft am 10.03.2021.

Google Inc. (Hg.) (2021f): Manage audio recordings in your Web & App Activity. Online verfügbar unter https://support.google.com/accounts/answer/6030020?co=GENIE.Platform%3DAndroid&hl=en, zuletzt geprüft am 10.03.2021.

Google Inc. (2022a): Audioaufnahmen in Ihren Web- & App-Aktivitäten verwalten. Online verfügbar unter https://support.google.com/websearch/answer/6030020, zuletzt aktualisiert am 08.02.2022, zuletzt geprüft am 08.02.2022.

Google Inc. (2022b): Einsatz der Mustererkennung durch Google – Datenschutzerklärung & Nutzungsbedingungen. Online verfügbar unter https://policies.google.com/technologies/pattern-recognition?hl=de&gl=de, zuletzt aktualisiert am 29.04.2022, zuletzt geprüft am 29.04.2022.

Gordon, Clara-Ann (2016): Daten aus Selbstvermessung. In: *Digma*, S. 70–75.

Gordon, Tuula; Holland, Janet; Lahelma, Elina; Campling, Jo (2000): Making Spaces: Citizenship and Difference in Schools. Citizenship and difference in schools. London: Palgrave Macmillan UK.

Grafsgaard, Joseph; Wiggins, Joseph B.; Boyer, Kristy Elizabeth; Wiebe, Eric N.; Lester, James (2013): Automatically recognizing facial expression: Predicting engagement and frustration. In: Educational Data Mining 2013.

Green, Ben; Chen, Yiling (2019): Disparate Interactions. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. Atlanta, GA, USA. New York, NY, USA: ACM, S. 90–99.

Griffeye (Hg.) (2021a): Analyze DI Pro. Online verfügbar unter https://www.griffeye.com/the-platform/analyze-di/, zuletzt geprüft am 22.04.2021.

Griffeye (Hg.) (2021b): GRIFFEYE INTELLIGENCE DATABASE. Online verfügbar unter https://www.griffeye.com/the-platform/griffeye-intelligence-database/.

Grossenbacher, Timo; Michel, Felix (2020): Automatische Gesichtserkennung – So einfach ist es, eine Überwachungsmaschine zu bauen. Hg. v. SRF. Online verfügbar unter https://www.srf.ch/news/schweiz/automatische-gesichtserkennung-so-einfach-ist-es-eine-ueberwachungsmaschine-zu-bauen, zuletzt aktualisiert am 07.02.2020, zuletzt geprüft am 31.03.2022.

Grother, Patrick; Ngan, Mei; Hanaoka, Kayee (2019): Face recognition vendor test part 3. Hg. v. NIST. Gaithersburg, MD.

Grundrechte.ch (2015): Polizei will Fussball-Fans mit Kameras überwachen – grundrechte.ch. In: *www.grundrechte.ch*. Online verfügbar unter https://grundrechte.ch/polizei-will-fussball-fans-mit-kameras-ueberwachen.html, zuletzt geprüft am 11.03.2021.

Guillén-Gámez, Francisco D.; García-Magariño, Iván (2014): Facial Authentication before and after Applying the Smowl Tool in Moodle. In: Sigeru Omatu, Hugues Bersini, Juan M. Corchado, Sara Rodríguez, Paweł Pawlewski und Edgardo Bucciarelli (Hg.): Distributed Computing and Artificial Intelligence, 11th International Conference, Bd. 290. Cham: Springer International Publishing (Advances in Intelligent Systems and Computing), S. 173–180.

Gunther, M.; Hu, P.; Herrmann, C.; Chan, C. H.; Jiang, M.; Yang, S. et al. (2017): Unconstrained Face Detection and Open-Set Face Recognition Challenge. In: IEEE International Joint Conference on Biometrics – IJCB 2017. October 1st–4th, Denver, CO. Institute of Electrical and Electronics Engineers; IJCB. Piscataway, NJ: IEEE, S. 697–706.

Gurovich, Yaron; Hanani, Yair; Bar, Omri; Nadav, Guy; Fleischer, Nicole; Gelbman, Dekel et al. (2019): Identifying facial phenotypes of genetic disorders using deep learning. In: *Nat Med* 25 (1), S. 60–64. DOI: 10.1038/s41591-018-0279-0.

Hagendorff, Thilo (2020): The Ethics of AI Ethics: An Evaluation of Guidelines. In: *Minds & Machines* 30 (1), S. 99–120. DOI: 10.1007/s11023-020-09517-8.

Hagerty, Alexa; Albert, Alexandra (2021): AI is increasingly being used to identify emotions – here's what's at stake. Online verfügbar unter https://theconversation.com/ai-is-increasingly-being-used-to-identify-emotions-heres-whats-at-stake-158809, zuletzt geprüft am 17.05.2021.

Hahn, T.; Nierenberg, A. A.; Whitfield-Gabrieli, S. (2017): Predictive analytics in mental health: applications, guidelines, challenges and perspectives. In: *Mol Psychiatry* 22 (1), S. 37–43. DOI: 10.1038/mp.2016.201.

Halbauer, I.; Klarmann, M. (2019): How Voice Retailers Can Predict Customer Mood and How They Can Use That Information. Working Paper.

Hale, Benjamin (2005): Identity crisis: Face recognition technology and freedom of the will. In: *Ethics, Place & Environment* 8 (2), S. 141–158. DOI: 10.1080/13668790500237047.

Han, Tian; Zhang, Jincheng; Zhang, Zhu; Sun, Guobing; Ye, Liang; Ferdinando, Hany et al. (2018): Emotion recognition and school violence detection from children speech. In: *J Wireless Com Network* 2018 (1), S. 1–10. DOI: 10.1186/s13638-018-1253-8.

Handelszeitung (2008): Spätes Umdenken. In: *Handelszeitung*, S. 28.

Hans, Julian (2017): Wie Russland Demonstranten identifiziert. In: *Süddeutsche Zeitung*, 12.07.2017. Online verfügbar unter https://www.sueddeutsche.de/politik/gesichtserkennung-wie-russland-demonstranten-identifiziert-1.3582647, zuletzt geprüft am 10.09.2021.

Hardesty, Larry (2019): Alexa's ASRU papers concentrate on extracting high-value training data. Hg. v. Amazon Science. Online verfügbar unter https://www.amazon.science/blog/alexas-asru-papers-concentrate-on-extracting-high-value-training-data, zuletzt aktualisiert am 02.01.2020, zuletzt geprüft am 08.02.2022.

Hartzog, Woodrow (2018): Facial Recognition Is the Perfect Tool for Oppression. Online verfügbar unter https://medium.com/s/story/facial-recognition-is-the-perfect-tool-for-oppression-bc2a08f0fe66, zuletzt geprüft am 07.04.2021.

Harwell, Drew; Dou, Eva (2020): Huawei tested AI software that could recognize Uighur minorities and alert police, report says. In: *The Washington Post*. Online verfügbar unter https://www.washingtonpost.com/technology/2020/12/08/huawei-tested-ai-software-that-could-recognize-uighur-minorities-alert-police-report-says/, zuletzt geprüft am 15.06.2021.

Heilweil, Rebecca (2020): The dystopian tech that companies are selling to help schools reopen sooner. Hg. v. Vox Media. Online verfügbar unter https://www.vox.com/recode/2020/8/14/21365300/artificial-intelligence-ai-school-reopening-technology-covid-19, zuletzt geprüft am 23.03.2021.

Held, Cornelius: Intelligente Videoüberwachung. Dissertation.

Herbig, Daniel (2020): Augmented Reality: Google kauft North. In: *Heise Online*. Online verfügbar unter https://www.heise.de/news/Augmented-Reality-Google-kauft-North-4800970.html, zuletzt geprüft am 05.05.2021.

Herrmann, Sebastian (2017): Emotionale Tonlage. In: *Süddeutsche Zeitung*, 15.10.2017. Online verfügbar unter https://www.sueddeutsche.de/wissen/psychologie-emotionale-tonlage-1.3707730.

Hertig, Maya (2015a): Kommentar zu Art. 16 BV. In: Bernhard Waldmann, Eva Maria Belser und Astrid Epiney (Hg.): Schweizerische Bundesverfassung (BV), Basler Kommentar. Basel.

Hertig, Maya (2015b): Kommentar zu Art. 17 BV. In: Bernhard Waldmann, Eva Maria Belser und Astrid Epiney (Hg.): Schweizerische Bundesverfassung (BV), Basler Kommentar. Basel.

Herzkammer (2021): An der Stimme Krankheiten erkennen. Hg. v. CSU-Fraktion im Bayerischen Landtag. Online verfügbar unter https://www.herzkammer.bayern/12/zentral/der-stimme-krankheiten-erkennen, zuletzt geprüft am 19.04.2021.

Herzog, Walter (2015): Emanzipation in Schule und Familie. In: *Psychotherapie-Wissenschaft* 5 (2), S. 110–117. Online verfügbar unter https://a-jour.info/index.php/psywis/article/view/281.

Hess, Ursula; Kleck, Robert E. (1990): Differentiating emotion elicited and deliberate emotional facial expressions. In: *Eur. J. Soc. Psychol.* 20 (5), S. 369–385. DOI: 10.1002/ejsp.2420200502.

Hill, Kashmir (2020): The Secretive Company That Might End Privacy as We Know It. In: *The New York Times*. Online verfügbar unter https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html, zuletzt geprüft am 15.06.2021.

Hill, Kashmir (2021): Facial Recognition: What Happens When We're Tracked Everywhere We Go? In: *The New York Times*. Online verfügbar unter https://www.nytimes.com/interactive/2021/03/18/magazine/facial-recognition-clearview-ai.html, zuletzt geprüft am 15.06.2021.

HireVue (2021): HireVue Delivers Game-Based Assessments for Measuring Job-related Emotional Intelligence. Online verfügbar unter https://www.hirevue.com/press-release/hirevue-delivers-game-based-assessments-for-measuring-job-related-emotional-intelligence, zuletzt geprüft am 31.03.2021.

Ho, Daniel E.; Black, Emily; Agrawala, Maneesh; Fei-Fei, Li (2020): Evaluating Facial Recognition Technology: A Protocol for Performance Assessment in New Domains. Hg. v. Stanford Institute for Human-Centered Artificial Intelligence. Online verfügbar unter https://law.stanford.edu/publications/evaluating-facial-recognition-technology-a-protocol-for-performance-assessment-in-new-domains/, zuletzt geprüft am 11.06.2021.

Hochreiter, S.; Schmidhuber, J. (1997): Long short-term memory. In: *Neural computation* 9 (8), S. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

Hodgson, Camilla (2019): AI lie detector developed for airport security. Hg. v. Financial Times. Online verfügbar unter https://www.ft.com/content/c9997e24-b211-11e9-bec9-fdcab53d6959, zuletzt geprüft am 31.03.2021.

Hoffman, Samantha (2017): Programming China:The Communist Party's autonomic approach to managing state security. Online verfügbar unter https://merics.org/sites/default/files/2020-05/Programming%20China.pdf, zuletzt geprüft am 15.06.2021.

Holmlund, Terje B.; Chandler, Chelsea; Foltz, Peter W.; Cohen, Alex S.; Cheng, Jian; Bernstein, Jared C. et al. (2020): Applying speech technologies to assess verbal memory in patients with serious mental illness. In: *npj Digit. Med.* 3 (1), S. 1–8. DOI: 10.1038/s41746-020-0241-7.

Hoppe, Sabrina; Loetscher, Tobias; Morey, Stephanie A.; Bulling, Andreas (2018): Eye Movements During Everyday Behavior Predict Personality Traits. In: *Frontiers in human neuroscience* 12, S. 105. DOI: 10.3389/fnhum.2018.00105.

Hoppenstedt, Max (2019): Wanzen im Wohnzimmer. In: *Süddeutsche Zeitung*, 21.10.2019. Online verfügbar unter https://www.sueddeutsche.de/digital/amazon-echo-google-home-abhoeren-ueberwachung-hacking-1.4649508, zuletzt geprüft am 07.03.2021.

Hoy, Matthew B. (2018): Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. In: *Medical Reference Services Quarterly* 37 (1), S. 81–88. DOI: 10.1080/02763869.2018.1404391.

Hsieh, Tzung-Chien; Bar-Haim, Aviram; Moosa, Shahida; Ehmke, Nadja; Gripp, Karen W.; Pantel, Jean Tori et al. (2021): GestaltMatcher: Overcoming the limits of rare disease matching using facial phenotypic descriptors. DOI: 10.1101/2020.12.28.20248193.

Huang, Shanshi; Peng, Xiaochen; Jiang, Hongwu; Luo, Yandong; Yu, Shimeng (2020): New Security Challenges on Machine Learning Inference Engine: Chip Cloning and Model Reverse Engineering. Online verfügbar unter http://arxiv.org/pdf/2003.09739v1.

Huang, Zhaocheng; Epps, Julien; Joachim, Dale (2019): Investigation of Speech Landmark Patterns for Depression Detection. In: *IEEE Trans. Affective Comput.*, S. 1. DOI: 10.1109/TAFFC.2019.2944380.

Hutchins, Brett; Andrejevcic, Mark (2021): Olympian Surveillance: Sports Stadiums and the Normalization of Biometric Monitoring. In: *International Journal of Communication* 15 (0), S. 1–20. Online verfügbar unter https://ijoc.org/index.php/ijoc/article/download/16377/3323.

IAB Switzerland (2016): DACH-Studie Mediennutzungsverhalten: Mediennutzung wird zunehmend mobil. In: *IAB Switzerland*. Online verfügbar unter https://www.iab-switzerland.ch/dach-studie-mediennutzungsverhalten-mediennutzung-wird-zunehmend-mobil/, zuletzt geprüft am 05.05.2021.

IBM Corporation (2011): IBM100 – Pioneering Speech Recognition. Online verfügbar unter https://www.ibm.com/ibm/history/ibm100/us/en/icons/speechreco/breakthroughs/, zuletzt aktualisiert am 07.03.2012, zuletzt geprüft am 08.12.2020.

Ienca, Marcello; Ignatiadis, Karolina (2020): Artificial Intelligence in Clinical Neuroscience: Methodological and Ethical Challenges. In: *AJOB neuroscience* 11 (2), S. 77–87. DOI: 10.1080/21507740.2020.1740352.

ifixit (Hg.) (2019): Amazon Echo Teardown. Online verfügbar unter https://www.ifixit.com/Teardown/Amazon+Echo+Teardown/33953, zuletzt geprüft am 19.03.2021.

Ihlenfeld, Jens (2011): Datenschützer kündigt rechtliche Schritte gegen Facebook an. Hg. v. Golem.de. Online verfügbar unter https://www.golem.de/1111/87672.html, zuletzt geprüft am 07.12.2020.

Ikea (Hg.) (2021): Häufig gestellte Fragen & Problembehandlung für SYMFONISK. Online verfügbar unter https://www.ikea.com/de/de/customer-service/product-support/symfonisk/wifi-lautsprecher-faq-pube3e23db6, zuletzt geprüft am 16.07.2021.

Iliou, Theodoros; Paschalidis, Georgios (2011): Using an Automated Speech Emotion Recognition Technique to Explore the Impact of Bullying on Pupils Social Life. In: Pantelis Angelidis (Hg.): Proceedings 15th Panhellenic Conference on Informatics (PCI 2011). Castoria, Greece, 30 September – 2 October 2011. Greek Computer Society; Panhellenic Conference on Informatics; PCI. Piscataway, NJ: IEEE, S. 18–22.

Imani, Maryam; Montazer, Gholam Ali (2019): A survey of emotion recognition methods with emphasis on E-Learning environments. In: *Journal of Network and Computer Applications* 147. DOI: 10.1016/j.jnca.2019.102423.

Imbernón Cuadrado, Luis-Eduardo; Manjarrés Riesco, Ángeles; La Paz López, Félix de (2019): FER in Primary School Children for Affective Robot Tutors. In: José Manuel Ferrández Vicente, José Ramón Álvarez-Sánchez, Félix de La Paz López, Javier Toledo Moreo und Hojjat Adeli (Hg.): From Bioinspired Systems and Biomedical Applications to Machine Learning. 8th International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2019, Almería, Spain, June 3–7, 2019, Pro-

ceedings, Part II. Cham: Springer (Theoretical Computer Science and General Issues), S. 461–471.

Information Commissioner (2000): CCTV Code of Practice. Hg. v. The Office of the Data Protection Commissioner. London. Online verfügbar unter https://web.archive.org/web/20060221053829/http:/www.ico.gov.uk/documentUploads/cctvcop1.pdf, zuletzt geprüft am 10.09.2021.

Internet Archive (2020): PimEyes | WebArchive. Online verfügbar unter https://web.archive.org/web/20200405180313if_/https://pimeyes.com/en/, zuletzt geprüft am 09.06.2021.

Inthavisas, K.; Lopresti, D. (2012): Secure speech biometric templates for user authentication. In: *IET Biometrics* 1 (1), S. 46–54. DOI: 10.1049/iet-bmt.2011.0008.

Isler, Michael (2019): Mobile App zur natürlichen Empfängnisverhütung ist ein Medizinprodukt. In: *LSR*.

IT Finanzmagazin (2020): Migros setzt auf Stimmbiometrie von Spitch. Online verfügbar unter https://www.it-finanzmagazin.de/migros-setzt-auf-stimmbiometrie-von-spitch-111101/, zuletzt aktualisiert am 29.04.2021, zuletzt geprüft am 30.04.2021.

IT-I-Ko (Hg.) (2020): Spitch: Sprachbiometrie-Banking ist am sichersten. Online verfügbar unter https://www.itiko.de/artikel/1847811/spitch-sprachbiometrie-banking-ist-am-sichersten.html, zuletzt geprüft am 30.04.2021.

Jack, Rachael E.; Garrod, Oliver G. B.; Yu, Hui; Caldara, Roberto; Schyns, Philippe G. (2012): Facial expressions of emotion are not culturally universal. In: *Proceedings of the National Academy of Sciences of the United States of America* 109 (19), S. 7241–7244. DOI: 10.1073/pnas.1200155109.

Jacobs, Harrison; Zheng, Annie (2018): Alibaba Hema Xiansheng supermarket reveals Whole Foods' Amazon future. In: *Insider*, 21.05.2018. Online verfügbar unter https://www.businessinsider.com/alibaba-hema-xiansheng-supermarket-whole-foods-amazon-future-2018-5, zuletzt geprüft am 29.04.2022.

James, William (1913): The principles of psychology. Henry Holt and Company.

Jantschewski, Patricia (2019): Back to the 90s – Die bizarre Welt der Vintage VR. Online verfügbar unter https://www.aspekteins.com/back-to-the-90s-die-bizarre-welt-der-vintage-vr/, zuletzt geprüft am 04.05.2021.

Jargon, Julie (2020): Back to School? Look Out for Covid-Tracking Surveillance Tech. Hg. v. The Wall Street Journal. Online verfügbar unter https://www.wsj.com/articles/back-to-school-look-out-for-covid-tracking-surveillance-tech-11597150800, zuletzt geprüft am 23.03.2021.

Jeancolas, Laetitia; Petrovska-Delacrétaz, Dijana; Mangone, Graziella; Benkelfat, Badr-Eddine; Corvol, Jean-Christophe; Vidailhet, Marie et al. (2021): X-Vectors: New Quantitative Biomarkers for Early Parkinson's Disease Detection From Speech. In: *Frontiers in neuroinformatics* 15, S. 578369. DOI: 10.3389/fninf.2021.578369.

Jobin, Anna; Ienca, Marcello; Vayena, Effy (2019): The global landscape of AI ethics guidelines. In: *Nat Mach Intell* 1 (9), S. 389–399. DOI: 10.1038/s42256-019-0088-2.

Johnson, Khari (2019): Google Assistant no longer saves voice recordings by default. Hg. v. venturebeat. Online verfügbar unter https://venturebeat.com/2019/09/23/google-assistant-no-longer-saves-voice-recordings-by-default/, zuletzt geprüft am 05.03.2021.

Jones, Rupert (2018): Voice recognition: is it really as secure as it sounds? Hg. v. The Guardian. Online verfügbar unter https://www.theguardian.com/money/2018/sep/22/voice-recognition-is-it-really-as-secure-as-it-sounds, zuletzt geprüft am 30.04.2021.

Jordan, Jochen (2014): Die Psychologie des Lügens. In: *ceg* 67 (67), S. 45–61. DOI: 10.4000/ceg.1656.

Juen, Florian; Huber, Eva Bänninger; Peham, Doris (2012): Geschlechts- und Altersunterschiede in der Emotionserkennung von Kindern und Jugendlichen. In: *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 44 (4), S. 178–191. DOI: 10.1026/0049-8637/a000072.

Kakkirala, Krishna Rao; Chalamala, Srinivasa Rao; Jami, Santosh Kumar (2017): Thermal Infrared Face Recognition: A Review. In: David Al-Dabass (Hg.): UKSim-AMSS 19th International Conference on Modelling and Simulation. Cambridge, 4/5/2017 – 4/7/2017. Piscataway, NJ: IEEE, S. 55–60.

Kaltheuner, Frederike; Obermüller, Nele (2018): Diskriminierende Gesichtserkennung: Ich sehe was, was du nicht bist. Online verfügbar unter https://netzpolitik.org/2018/diskriminierende-gesichtserkennung-ich-sehe-was-was-du-nicht-bist/, zuletzt geprüft am 14.01.2021.

Kanton Zürich (2018): Fünf Ausreise-Schleusen für die automatisierte Passkontrolle am Flughafen Zürich. Online verfügbar unter https://www.zh.ch/de/news-uebersicht/medienmitteilungen/2018/05/1805291m.html, zuletzt aktualisiert am 31.05.2018, zuletzt geprüft am 29.04.2022.

Karaboga, Murat; Masur, Philipp; Matzner, Tobias; Mothes, Cornelia; Nebel, Maxi; Ochs, Carsten et al. (2014): White Paper Selbstdatenschutz. Forum Privatheit und selbstbestimmtes Leben in der digitalen Welt. Hg. v. Fraunhofer-Institut für System- und Innovationsforschung. Karlsruhe. Online verfügbar unter https://www.forum-privatheit.de/forum-privatheit-de/texte/veroeffentlichungen-des-forums/themenpapiere-white-paper/Forum_Privatheit_White_Paper_Selbstdatenschutz_2.Auflage.pdf.

Kaur, Paramjit; Krishan, Kewal; Sharma, Suresh K.; Kanchan, Tanuj (2020): Facial-recognition algorithms: A literature review. In: *Medicine, science, and the law* 60 (2), S. 131–139. DOI: 10.1177/0025802419893168.

Keist, Ramona (2019): Gesichtserkennung im zivilrechtlichen Persönlichkeitsschutz. In: *Jusletter*.

Kelly, Makena (2019): Amazon confirms it holds on to Alexa data even if you delete audio files. Hg. v. The Verge. Online verfügbar unter https://www.theverge.com/2019/7/3/20681423/amazon-alexa-echo-chris-coons-data-transcripts-recording-privacy, zuletzt geprüft am 19.03.2021.

Ketamo, Harri; O'Rourke, Petra (2019): Proceedings of Fake Intelligence Online Summit: Satakunta University of Applied Sciences. Online verfügbar unter http://www.theseus.fi/bitstream/10024/169049/1/2019_D_1_SAMK_Proceedings_FakeIntelligenceOnlineSummit2019.pdf.

Khan, Wasim Ahmed (2020): Functional Reverse Engineering of Machine Tools. Milton: Taylor et Francis Group (Computers in Engineering Design and Manufacturing Ser).

Kim, Young-Bum; Kim, Dongchan; Kumar, Anjishnu; Sarikaya, Ruhi (2018): Efficient Large-Scale Neural Domain Classification with Personalized Attention. In: Iryna Gurevych und Yusuke Miyao (Hg.): Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia. Stroudsburg, PA, USA: Association for Computational Linguistics, S. 2214–2224.

Kinsella, Bret (2020): New Alexa Skill Data Show New U.S. Skills Launched in 2019 Fall to Lowest Level Since 2016. Hg. v. Voicebot.ai. Online verfügbar unter https://voicebot.ai/2020/01/17/new-alexa-skill-data-show-new-u-s-skills-launched-in-2019-fall-to-lowest-level-since-2016/, zuletzt geprüft am 19.03.2021.

Kirby, M.; Sirovich, L. (1990): Application of the Karhunen-Loeve procedure for the characterization of human faces. In: *IEEE Trans. Pattern Anal. Machine Intell.* 12 (1), S. 103–108. DOI: 10.1109/34.41390.

Kire (2020): Staatstrojaner werden in der Schweiz intensiv eingesetzt. Hg. v. Digitale Gesellschaft. Online verfügbar unter https://www.digitale-gesellschaft.ch/2020/01/12/staatstrojaner-werden-in-der-schweiz-intensiv-eingesetzt-unzulaessiger-einsatz/, zuletzt aktualisiert am 12.01.2020, zuletzt geprüft am 31.03.2022.

Klett, Barbara (2017): Digitalisierte Gesundheit – Abgrenzungen und Regulierung. In: *HAVE*, S. 104–113.

Klett, Barbara; Verde, Michel (2016): Medizinprodukt- und haftpflichtrechtliche Aspekte bei Medizinal-Apps. In: *Sicherheit & Recht / Sécurité & Droit* (1). Online verfügbar unter https://www.dike.ch/sr-1-2016.

Kline, Jeffrey A.; Neumann, Dawn; Haug, Melissa A.; Kammer, David J.; Krabill, Virginia A. (2015): Decreased facial expression variability in patients with serious cardiopulmonary disease in the emergency care setting. In: *Emergency medicine journal : EMJ* 32 (1), S. 3–8. DOI: 10.1136/emermed-2014-203602.

Klöß, Sebastian (2020): Die Zukunft der Consumer Technology – 2020. Marktentwicklung, Trends, Mediennutzung, Technologien, Geschäftsmodelle. Hg. v. Bitkom. Online verfügbar unter https://www.bitkom.org/sites/default/files/2020-08/200826_ct_studie_2020_online.pdf.

Knupfer, Gabriel (2020a): Gesichtserkennung: «Seither hatten sie nie mehr ein Geisterspiel». Hg. v. Handelszeitung. Online verfügbar unter https://www.handelszeitung.ch/tech/seither-hatten-sie-nie-mehr-ein-geisterspiel, zuletzt aktualisiert am 16.06.2021, zuletzt geprüft am 16.06.2021.

Knupfer, Gabriel (2020b): Gesichtserkennung: Das Potenzial ist riesig – aber ebenso die Gefahren. Hg. v. Handelszeitung. Online verfügbar unter https://www.handelszeitung.ch/panorama/gesichtserkennung-das-potenzial-ist-riesig-aber-ebenso-die-gefahren, zuletzt aktualisiert am 24.08.2021, zuletzt geprüft am 24.08.2021.

Knupfer, Gabriel (2020c): Interview mit Christian Fehrlin von Deep Impact: «Wir verkaufen nur an demokratische Länder». Hg. v. Handelszeitung. Online verfügbar unter https://www.handelszeitung.ch/tech/seither-hatten-sie-nie-mehr-ein-geisterspiel, zuletzt geprüft am 20.04.2021.

Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, Jong-Seok; Yazdani, A.; Ebrahimi, T. et al. (2012): DEAP: A Database for Emotion Analysis ; Using Physiological Signals. In: *IEEE Trans. Affective Comput.* 3 (1), S. 18–31. DOI: 10.1109/T-AFFC.2011.15.

Koenecke, Allison; Nam, Andrew; Lake, Emily; Nudell, Joe; Quartey, Minnie; Mengesha, Zion et al. (2020): Racial disparities in automated speech recognition. In: *Proceedings of the National Academy of Sciences of the United States of America*, S. 7684–7689. DOI: 10.1073/pnas.1915768117.

Kohse, Petra (2020): «Die Maschine kann Stimmen so gut lesen, wie es ein Mensch tut». Online verfügbar unter https://www.berliner-zeitung.de/zukunft-technologie/die-maschine-kann-stimmen-so-gut-lesen-wie-es-ein-mensch-tut-li.84407, zuletzt aktualisiert am 07.06.2020, zuletzt geprüft am 19.04.2021.

Koops, Bert-Jaap; Newell, Bryce Clayton; Timan, Tjerk; Škorvánek, Ivan; Chokrevski, Tom; Galič, Maša (2016): A Typology of Privacy.

Kraus, Michael W. (2017): Voice-only communication enhances empathic accuracy. In: *The American psychologist* 72 (7), S. 644–654. DOI: 10.1037/amp0000147.

Kremp, Matthias (2019): Siri-Transkription: So werden Sprachbefehle von Menschen ausgewertet. In: *SPIEGEL*, 08.08.2019. Online verfügbar unter https://www.spiegel.de/netzwelt/gadgets/siri-transkription-so-werden-sprachbefehle-von-menschen-ausgewertet-a-1280333.html, zuletzt geprüft am 14.12.2020.

Krempl, Stefan (2020): NIST-Studie: Algorithmen erkennen Gesichter trotz Schutzmasken besser. In: *Heise Online*, 12.04.2020. Online verfügbar unter https://www.heise.de/news/NIST-Studie-Algorithmen-erkennen-Gesichter-trotz-Schutzmasken-besser-4979878.html, zuletzt geprüft am 29.04.2022.

Kreppmeier, Lea (2020): Ein Startup aus dem Silicon Valley arbeitet an Augmented-Reality-Kontaktlinsen. In: *Business Insider*. Online verfügbar unter https://www.businessinsider.de/tech/us-startup-arbeitet-an-augmented-reality-kontaktlinsen/, zuletzt geprüft am 05.05.2021.

Kretschmer, Fabian (2021): China zeigt die High-Tech Schule der Zukunft. In: *Neue Zürcher Zeitung*, 18.06.2021. Online verfügbar unter https://www.nzz.ch/international/china-zeigt-die-high-tech-schule-der-zukunft-ld.1628314, zuletzt geprüft am 16.07.2021.

Kuhn, Johannes (2019): Mein Smartphone weiß, dass ich wütend bin. In: *sueddeutsche*, 27.03.2019. Online verfügbar unter https://www.sueddeutsche.de/digital/smartphone-software-emotionen-simulation-ki-1.4377004, zuletzt geprüft am 06.04.2021.

Kühne, Stefan (2022): Automatisierte Bearbeitung von Personendaten im Strafprozess- und Polizeirecht. In: *Sicherheit & Recht*, 1, S. 13–24.

Kunath Funk, Gabriele; Hofstetter, Reto; Maurer, Markus; Lopetrone, R.; Jörg, D. (2020): Voice First Barometer Schweiz 2019. In: *Farner Consulting AG*. Online verfügbar unter https://www.farner.ch/2020/04/voicefirstbarometer2019/, zuletzt geprüft am 05.03.2021.

Laguarta, Jordi; Hueto, Ferran; Subirana, Brian (2020): COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings. In: *IEEE Open J. Eng. Med. Biol.* 1, S. 275–281. DOI: 10.1109/OJEMB.2020.3026928.

Landolt, Hardy (2017): Medizinalproduktehaftpflicht – ein Überblick. In: *HAVE*, S. 100–102.

Langley, Hugh; Pattison Tuohy, Jennifer (2019): Smart home privacy: What Amazon, Google and Apple do with your data. Online verfügbar unter https://www.the-ambient.com/features/how-amazon-google-apple-use-smart-speaker-data-338, zuletzt geprüft am 05.03.2021.

Latif, Siddique; Qadir, Junaid; Qayyum, Adnan; Usama, Muhammad; Younis, Shahzad (2020): Speech Technology for Healthcare: Opportunities, Challenges, and State of the Art. In: *IEEE reviews in biomedical engineering* 14, S. 342–356. DOI: 10.1109/RBME.2020.3006860.

Lau, Josephine; Zimmerman, Benjamin; Schaub, Florian (2018): Alexa, Are You Listening? In: *Proc. ACM Hum.-Comput. Interact.* 2 (CSCW), S. 1–31. DOI: 10.1145/3274371.

Laufer, Daniel; Meineck, Sebastian (2020): PimEyes – Eine polnische Firma schafft gerade unsere Anonymität ab. In: *netzpolitik.org*. Online verfügbar unter https://netzpolitik.org/2020/gesichter-suchmaschine-pimeyes-schafft-anonymitaet-ab/, zuletzt geprüft am 24.07.2020.

Le, Vuong; Brandt, Jonathan; Lin, Zhe; Bourdev, Lubomir; Huang, Thomas S. (2012): Interactive Facial Feature Localization. In: Andrew Fitzgibbon (Hg.): Computer vision – ECCV 2012. 12th European Conference on Computer Vision, Florence, Italy, October 7 – 13, 2012; proceedings. Berlin/Heidelberg: Springer (Lecture notes in computer science, 7574), S. 679–692.

Ledebur, Michael von (2018): Überwachung in der Badi: Zürcher Sportamt zieht Videoreglement zurück. In: *Neue Zürcher Zeitung*. Online verfügbar unter https://www.nzz.ch/zuerich/ueberwachung-in-der-badi-der-zuercher-stadtrat-zieht-video-reglement-zurueck-ld.1440235?reduced=true, zuletzt geprüft am 15.06.2021.

Lee, Jiyoung; Kim, Seungryong; Kim, Sunok; Park, Jungin; Sohn, Kwanghoon (2019): Context-Aware Emotion Recognition Networks. In: International Conference on Computer Vision. ICCV 2019 : proceedings : 27 October – 2 November 2019, Seoul, Korea. 2019. Piscataway, NJ: IEEE, S. 10142–10151.

Leins-Zurmühle, Sarah (2021): Mobile Applikationen als Medizinprodukte, Qualifikation und Pflichten nach der revidierten Schweizer Medizinprodukteverordnung (3), S. 137–147.

Lewis, Amanda E.; Diamond, John (2015): Despite the best intentions. How racial inequality thrives in good schools. Oxford: Oxford University Press (Transgressing boundaries).

LexisNexis (Hg.) (2021): Nexis Startseite. Online verfügbar unter https://advance.lexis.com, zuletzt geprüft am 15.01.2021.

Li, Shan; Deng, Weihong (2020): Deep Facial Expression Recognition: A Survey. In: *IEEE Trans. Affective Comput.*, S. 1. DOI: 10.1109/TAFFC.2020.2981446.

Li, Shancang; Choo, Kim-Kwang Raymond; Sun, Qindong; Buchanan, William J.; Cao, Jiuxin (2019): IoT Forensics: Amazon Echo as a Use Case. In: *IEEE Internet Things J.* 6 (4), S. 6487–6497. DOI: 10.1109/JIOT.2019.2906946.

Liang, Fan; Das, Vishnupriya; Kostyuk, Nadiya; Hussain, Muzammil M. (2018): Constructing a Data-Driven Society: China's Social Credit System as a State Surveillance Infrastructure. In: *Policy & Internet* 10 (4), S. 415–453. DOI: 10.1002/poi3.183.

Liang, Shenbao (2018): Einwilligung in medizinische Behandlungen. Eine rechtsvergleichende Analyse nach schweizerischem und chinesischem Privatrecht (Arbeiten aus dem Iuristischen Seminar der Universität Freiburg Schweiz, 390).

Liaw, Hongming Leonard; Chiu, Mei-Hung; Chou, Chin-Cheng (2014): Using facial recognition technology in the exploration of student responses to conceptual conflict phenomenon. In: *Chem. Educ. Res. Pract.* 15 (4), S. 824–834. DOI: 10.1039/c4rp00103f.

Liaw, Hongming; Yu, Yuh-Ru; Chou, Chin-Cheng; Chiu, Mei-Hung (2021): Relationships between Facial Expressions, Prior Knowledge, and Multiple Representations: a Case of Conceptual Change for Kinematics Instruction. In: *J Sci Educ Technol* 30 (2), S. 227–238. DOI: 10.1007/s10956-020-09863-3.

Lin, Shen; Li, Zhigang; Fu, Bowen; Chen, Sipeng; Li, Xi; Wang, Yang et al. (2020): Feasibility of using deep learning to detect coronary artery disease based on facial photo. In: *European heart journal* 41 (46), S. 4400–4411. DOI: 10.1093/eurheartj/ehaa640.

Linhart, Lisa-Marie (2019): Bewerbungsgespräch mit KI: Ersetzt Gesichtsanalyse den persönlichen Kontakt? Hg. v. karriere.blog. Online verfügbar unter https://www.karriere.at/blog/automatische-gesichtsanalyse-bewerbungsgespraech.html, zuletzt aktualisiert am 03.10.2019, zuletzt geprüft am 29.04.2022.

Lippmann, Richard P. (1997): Speech recognition by machines and humans. In: *Speech Communication* 22 (1), S. 1–15. DOI: 10.1016/S0167-6393(97)00021-6.

Litman-Navarro, Kevin (2019): We Read 150 Privacy Policies. They Were an Incomprehensible Disaster. Opinion. In: *The New York Times*, 12.06.2019. Online verfügbar unter https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html, zuletzt geprüft am 04.05.2022.

Lobo, Sascha (2013): Googles fahrlässige Glass-Kampagne. In: *Spiegel Online*. Online verfügbar unter https://www.spiegel.de/netzwelt/web/sascha-lobo-googles-fahrlaessige-glass-kampagne-a-898512.html, zuletzt geprüft am 04.05.2021.

Loderer, Kristina; Gentsch, Kornelia; Duffy, Melissa C.; Zhu, Mingjing; Xie, Xiyao; Chavarría, Jason A. et al. (2020): Are concepts of achievement-related emotions universal across cultures? A semantic profiling approach. In: *Cognition & emotion* 34 (7), S. 1480–1488. DOI: 10.1080/02699931.2020.1748577.

Lokalinfo.ch (2018): Videoüberwachung: Bereit für Gesichtserkennung. Hg. v. Lokalinfo.ch. Online verfügbar unter https://www.lokalinfo.ch/news/artikel/videoueberwachung-bereit-fuer-gesichtserkennung/, zuletzt aktualisiert am 15.06.2021, zuletzt geprüft am 15.06.2021.

Lopatto, Elizabeth (2020): Clearview AI CEO says «over 2,400 police agencies» use its facial recognition software. In: *The Verge*. Online verfügbar unter https://www.theverge.com/2020/8/26/21402978/clearview-ai-ceo-interview-2400-police-agencies-facial-recognition, zuletzt geprüft am 15.06.2021.

Luchetta, Simone (2021a): Umstrittene Technologie im Einsatz – So jagen Schweizer Polizisten mit Gesichtserkennung Verbrecher. In: *Tages-Anzeiger*. Online verfügbar unter https://www.tagesanzeiger.ch/so-jagen-schweizer-polizisten-mit-gesichtserkennung-verbrecher-608167461846, zuletzt geprüft am 15.06.2021.

Luchetta, Simone (2021b): Neue Helfer bei der Verbrecherjagd. In: *Tages-Anzeiger*, 17.04.2021. Online verfügbar unter https://www.tagesanzeiger.ch/so-jagen-schweizer-polizisten-mit-gesichtserkennung-verbrecher-608167461846.

Lüchinger, Corinne Widmer (2019): Apps, Algorithmen und Roboter in der Medizin: Haftungsrechtliche Herausforderungen. In: *HAVE*, S. 3–15.

Luxand, Inc (Hg.) (2021): Detect and Recognize Faces and Facial Features with Luxand FaceSDK. Online verfügbar unter https://www.luxand.com/facesdk/, zuletzt geprüft am 22.04.2021.

Lyon, David (2014): Surveillance, Snowden, and Big Data: Capacities, consequences, critique. In: *Big Data & Society* 1 (2), 205395171454186. DOI: 10.1177/2053951714541861.

Ma, Yong; Drewes, Heiko; Butz, Andreas (2021): Fake Moods: Can Users Trick an Emotion-Aware VoiceBot? In: Yoshifumi Kitamura, Aaron Quigley, Katherine Isbister und Takeo Igarashi (Hg.): Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21. Yokohama Japan, 08 05 2021 13 05 2021. New York, NY, USA: ACM, S. 1–4.

Mac, Ryan (2020): Clearview AI's Facial Recognition Tech Is Being Used By The Justice Department, ICE, And The FBI. In: *BuzzFeed News*. Online verfügbar unter https://www.buzzfeednews.com/article/ryanmac/clearview-ai-fbi-ice-global-law-enforcement, zuletzt geprüft am 15.06.2021.

Mach, Vaclav; Kotkova, Barbora; Hromada, Martin (2019): Detection and Recognition of People by Camera – Reliability and Use. In: Branko Katalinic (Hg.): DAAAM International Scientific Book, Bd. 18: DAAAM International Vienna, S. 233–240.

Mäder, Philipp (2005): Euro 08 will Hooligans dank Gesichtserkennung fern halten. In: *Tages-Anzeiger*, S. 7.

Magdin, Martin; Turcani, Milan; Hudec, Lukas (2016): Evaluating the Emotional State of a User Using a Webcam. In: *international journal of interactive multimedia and artificial intelligence* 4 (1), 61-68-0. Online verfügbar unter https://scimatic.org/storage/journals/16568/pdfs.

Maio, Giovanni (2012): Chancen und Grenzen der personalisierten Medizin – eine ethische Betrachtung. In: *GGW* 12 (1), S. 15–19. Online verfügbar unter https://www.wido.de/fileadmin/Dateien/Dokumente/Publikationen_Produkte/GGW/wido_ggw_0112_maio.pdf, zuletzt geprüft am 18.02.2022.

Malik, Khalid Mahmood; Malik, Hafiz; Baumann, Roland (2019): Towards Vulnerability Analysis of Voice-Driven Interfaces and Countermeasures for Replay Attacks. In: Second International Conference on Multimedia Information Processing and Retrieval. Proceedings : 28–30 March 2019 San Jose, California. 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). San Jose, CA, USA, 3/28/2019 – 3/30/2019. Los Alamitos, California, Washington, Tokyo: Conference Publishing Services, IEEE Computer Society, S. 523–528.

Manikonda, Lydia; Deotale, Aditya; Kambhampati, Subbarao (2018): What's up with Privacy? User Preferences and Privacy Concerns in Intelligent Personal Assistants. In: Jason Furman, Gary Marchant, Huw Price und Francesca Rossi (Hg.): Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. AIES '18: AAAI/ACM Con-

ference on AI, Ethics, and Society. New Orleans LA USA, 02 02 2018 03 02 2018. [S.l.]: ACM, S. 229–235.

Mantelero, Alessandro (2018): AI and Big Data: A blueprint for a human rights, social and ethical impact assessment. In: *Computer Law & Security Review* 34 (4), S. 754–772. DOI: 10.1016/j.clsr.2018.05.017.

Maor, Elad; Sara, Jaskanwal D.; Orbelo, Diana M.; Lerman, Lilach O.; Levanon, Yoram; Lerman, Amir (2018): Voice Signal Characteristics Are Independently Associated With Coronary Artery Disease. In: *Mayo Clinic Proceedings* 93 (7), S. 840–847. DOI: 10.1016/j.mayocp.2017.12.025.

MarketWatch (2021): Emotion Detection and Recognition (EDR) Market 2021–2030 Global Industry Growth Opportunities, Share Estimation, Strategy, Benefits, Demand, Manufactures Analysis and Regional Forecast. Online verfügbar unter https://www.marketwatch.com/press-release/emotion-detection-and-recognition-edr-market-2021-2030-global-industry-growth-opportunities-share-estimation-strategy-benefits-demand-manufactures-analysis-and-regional-forecast-2021-08-27?tesla=y, zuletzt geprüft am 10.09.2021.

Marmar, Charles R.; Brown, Adam D.; Qian, Meng; Laska, Eugene; Siegel, Carole; Li, Meng et al. (2019): Speech-based markers for posttraumatic stress disorder in US veterans. In: *Depression and anxiety* 36 (7), S. 607–616. DOI: 10.1002/da.22890.

Marti, Fabio (2021): Amsterdam's Johan Cruijff ArenA Partners With Security & Safety Things to Enhance Fan Experience, Health and Safety. In: *Security & Safety Things*. Online verfügbar unter https://www.parking-net.com/parking-news/security-safety-things/amsterdams-johan-cruijff-arena-enhance-fan-experience-health-and-safety, zuletzt geprüft am 16.06.2021.

Martin, S. (2021): Teaching and Learning Advances on Sensors for IoT. [S.l.]: MDPI AG.

Martinez-Martin, Nicole (2019): What Are Important Ethical Implications of Using Facial Recognition Technology in Health Care? In: *AMA journal of ethics* 21 (2), E180-187. DOI: 10.1001/amajethics.2019.180.

Mathelitsch, Leopold; Verovnik, Ivo (2016): Geistertöne. In: *Physik in unserer Zeit* 47 (2), S. 82–83. DOI: 10.1002/piuz.201601428.

Matzner, Tobias (2016): Grasping the ethics and politics of algorithms. Hg. v. Medium. Online verfügbar unter https://medium.com/@t_matzner/grasping-the-ethics-and-politics-of-algorithms-c2932804fa9d, zuletzt aktualisiert am 18.10.2016, zuletzt geprüft am 29.04.2022.

Maurer-Lambrou, Urs; Honsell, Heinrich (Hg.) (2014): Datenschutzgesetz, Öffentlichkeitsgesetz. 3. Aufl. Basel: Helbing Lichtenhahn (Basler Kommentar).

McLaughlin, Michael; Castro, Daniel (2020): The Critics Were Wrong: NIST Data Shows the Best Facial Recognition Algorithms Are Neither Racist Nor Sexist. Online verfügbar unter https://itif.org/sites/default/files/2020-best-facial-recognition.pdf, zuletzt geprüft am 15.06.2021.

McManus, Ashley (2018): The User Experience of In-Vehicle Emotion Detection: A Nuance Workshop. Online verfügbar unter https://blog.affectiva.com/the-user-experience-of-in-vehicle-emotion-detection-a-nuance-workshop, zuletzt geprüft am 18.02.2022.

Mcschindler (Hg.) (2019): Social Media in der Schweiz 2019: Ergebnisse der 10. Befragung. Online verfügbar unter https://www.mcschindler.com/social-media-in-der-schweiz-2019-ergebnisse-der-10-befragung/, zuletzt geprüft am 05.03.2021.

MDR (Hg.) (2019): Künstliche Intelligenz liest aus Gesichtern Gendefekte. Online verfügbar unter https://www.mdr.de/wissen/mensch-alltag/gesichtserkennung-als-diagnose-hilfe-100.html, zuletzt aktualisiert am 11.03.2019, zuletzt geprüft am 19.04.2021.

Meek, James (2002): Robo cop. In: *The Guardian*. Online verfügbar unter http://www.theguardian.com/uk/2002/jun/13/ukcrime.jamesmeek, zuletzt geprüft am 09.03.2021.

Mehmood, Raja Majid; Lee, Hyo Jong (2017): Towards Building a Computer Aided Education System for Special Students Using Wearable Sensor Technologies. In: *Sensors (Basel, Switzerland)* 17 (2). DOI: 10.3390/s17020317.

Meidert, Ursula; Scheermesser, Mandy; Prieur, Yvonne; Hegyi, Stefan; Stockinger, Kurt; Eyyi, Gabriel et al. (2018): Quantified Self – Schnittstelle zwischen Lifestyle und Medizin. Zürich: vdf Hochschulverlag AG an der ETH Zürich (TA-SWISS, 67). Online verfügbar unter https://vdf.ch/quantified-self-schnittstelle-zwischen-lifestyle-und-medizin-e-book.html.

Meili, Andreas (2018): Kommentar zu Art. 26 ZGB. In: Thomas Geiser und Christiana Fountoulakis (Hg.): Zivilgesetzbuch I Art. 1–456 ZGB. Basler Kommentar. Basel.

Melchior, Laura (2018): Amazon.com liefert nicht mehr in die Schweiz. In: *Internet World Business*. Online verfügbar unter /plattformen/amazon/amazoncom-liefert-in-schweiz-1655296.html, zuletzt geprüft am 05.03.2021.

Menegus, Bryan (2019): Amazon's Defense of Rekognition Tool Undermined by Police Client. Online verfügbar unter https://gizmodo.com/defense-of-amazons-face-recognition-tool-undermined-by-1832238149, zuletzt geprüft am 14.06.2021.

Merialdo, Bernard (1988): Multilevel decoding for Very-Large-Size-Dictionary speech recognition. In: *IBM J. Res. & Dev.* 32 (2), S. 227–237. DOI: 10.1147/rd.322.0227.

Metallinou, Angeliki; Yang, Zhaojun; Lee, Chi-chun; Busso, Carlos; Carnicke, Sharon; Narayanan, Shrikanth (2016): The USC CreativeIT database of multimodal dyadic interactions: from speech and full body motion capture to continuous emotional annotations. In: *Lang Resources & Evaluation* 50 (3), S. 497–521. DOI: 10.1007/s10579-015-9300-0.

Meyer, Roland (2020): Ein aufhaltsamer Aufstieg: Zur Geschichte der automatisierten Gesichtserkennung. In: *0932-5409*.

Meyer-Lindenberg, A. (2018): Künstliche Intelligenz in der Psychiatrie – ein Überblick. In: *Nervenarzt* 89 (8), S. 861–868. DOI: 10.1007/s00115-018-0557-6.

Meyers, Alyssa (2019): Auf Smart Speakern wie Amazons Alexa könnte bald mehr Werbung laufen. In: *Business Insider*, 12.02.2019. Online verfügbar unter https://www.business-insider.de/tech/smart-speaker-amazon-alexa-mehr-werbung-2019-12/, zuletzt geprüft am 30.03.2022.

Michael, B.; Roth, R. (2001): Biometrische Identifikation. Grundlagen, Verfahren, Perspektiven. Wiesbaden: Vieweg+Teubner Verlag (DuD-Fachbeiträge).

Microsoft (2018): Six principles to guide Microsoft's facial recognition work. Online verfügbar unter https://blogs.microsoft.com/on-the-issues/2018/12/17/six-principles-to-guide-mi-

crosofts-facial-recognition-work/, zuletzt aktualisiert am 08.01.2019, zuletzt geprüft am 15.02.2022.

Migros Bank (Hg.) (2021): Stimmbiometrie. Online verfügbar unter https://www.migrosbank.ch/de/stimmbiometrie.html, zuletzt aktualisiert am 30.08.2021, zuletzt geprüft am 30.08.2021.

Miller, Dan; Smallman, Matt; Top, Derek (2020): Intelligent Authentication and Fraud Prevention Intelliview. Solutions for Emerging Security Threats and CX Challenges. Hg. v. Opus Research. Online verfügbar unter https://opusresearch.net/wordpress/pdfreports/2020_IAuth_Intelliview_final_leadup.pdf, zuletzt geprüft am 18.02.2022.

Mitev, Richard; Pazii, Anna; Miettinen, Markus; Enck, William; Sadeghi, Ahmad-Reza (2020): LeakyPick: IoT Audio Spy Detector. In: Annual Computer Security Applications Conference. ACSAC '20. Austin USA. New York, NY, USA: ACM, S. 694–705.

Mittelstadt, Brent Daniel; Floridi, Luciano (2016): The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. In: *Science and engineering ethics* 22 (2), S. 303–341. DOI: 10.1007/s11948-015-9652-2.

Mühlhoff, Rainer (2021): Predictive privacy: towards an applied ethics of data analytics. In: *Ethics Inf Technol* 23 (4), S. 675–690. DOI: 10.1007/s10676-021-09606-x.

Mukhopadhyay, Moutan; Pal, Saurabh; Nayyar, Anand; Pramanik, Pijush Kanti Dutta; Dasgupta, Niloy; Choudhury, Prasenjit (2020): Facial Emotion Detection to Assess Learner's State of Mind in an Online Learning System. In: Proceedings of the 2020 5th International Conference on Intelligent Information Technology. ICIIT 2020. Hanoi Viet Nam, 19 02 2020 22 02 2020. New York, NY, United States: Association for Computing Machinery (ACM Digital Library), S. 107–115.

Müller, Jörg Paul; Schefer, Markus (2008): Grundrechte in der Schweiz. Im Rahmen der Bundesverfassung, der EMRK und der UNO-Pakte. 4. [neubearb.] Aufl. Bern: Stämpfli.

Munteanu, Cosmin; Penn, Gerald (2018): Speech and Hands-free Interaction. In: Regan Mandryk, Mark Hancock, Mark Perry und Anna Cox (Hg.): Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18. Montreal, QC, Canada. New York, NY, USA: ACM, S. 1–4.

Muresan, Remus (2021): Künstliche Intelligenz und Medizinalprodukte: Qualifizierung und Klassifizierung unter der MDR. In: *LSR* (1), S. 17–40.

Mustafa, Mohammed Kyari; Allen, Tony; Appiah, Kofi (2019): A comparative review of dynamic neural networks and hidden Markov model methods for mobile on-device speech recognition. In: *Neural Comput & Applic* 31 (S2), S. 891–899. DOI: 10.1007/s00521-017-3028-2.

Myers, Lisa (2022): An Exploration of Voice Biometrics. Hg. v. SANS Institute. Online verfügbar unter https://www.sans.org/white-papers/1436/, zuletzt aktualisiert am 29.04.2022, zuletzt geprüft am 29.04.2022.

Natatsuka, Atsuko; Iijima, Ryo; Watanabe, Takuya; Akiyama, Mitsuaki; Sakai, Tetsuya; Mori, Tatsuya (2019): Poster: A first look at the privacy risks of voice assistant apps. In: Lorenzo Cavallaro, Johannes Kinder, XiaoFeng Wang und Jonathan Katz (Hg.): CCS'19. Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications

Security, London, United Kingdom, November 11-15, 2019. CCS ‘19. New York, NY: Association for Computing Machinery, S. 2633–2635.

Nayar, Revathy (2017): Towards designing speech technology based assistive interfaces for children’s speech therapy. In: Edward Lank und A. Special Interest Group on Computer-HumanC.M. Interaction (Hg.): Proceedings of the 19th ACM International Conference on Multimodal Interaction. ICMI ‘17. Glasgow UK, 13 11 2017 – 17 11 2017. [Place of publication not identified]: ACM, S. 609–613.

NeatoCode Techniques (2013): First Facial Recognition Hack for Google Glass, zuletzt aktualisiert am 12.05.2013, zuletzt geprüft am 30.03.2022.

Nguyen; Alexander (2020): Leitlinien 3/2019 zur Verarbeitung personenbezogener Daten durch Videogeräte Version 2.0. Online verfügbar unter https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201903_video_devices_de.pdf, zuletzt geprüft am 13.06.2021.

Nguyen, Duong Tung; Le, Long Bao; Bhargava, Vijay (2021): Price-Based Resource Allocation for Edge Computing: A Market Equilibrium Approach. In: *IEEE Trans. Cloud Comput.* 9 (1), S. 302–317. DOI: 10.1109/TCC.2018.2844379.

Nguyen, Thai-Son; Stueker, Sebastian; Waibel, Alex (2020): Super-Human Performance in Online Low-latency Recognition of Conversational Speech. In: *arXiv preprint*. Online verfügbar unter https://arxiv.org/abs/2010.03449.

Nikopoulou, Rozalia; Vernikos, Ioannis; Spyrou, Evaggelos; Mylonas, Phivos (2018): Emotion Recognition from Speech. In: PETRA 2018. The 11th ACM International Conference on PErvasive Technologies Related to Assistive Environments: June 26–29, 2018, Corfu, Greece : conference proceedings. University of Texas at Arlington; National Science Foundation; Association for Computing Machinery; PETRA. New York, NY, USA: ACM (ICPS), S. 104–105.

NIST (2016): Face Recognition Grand Challenge (FRGC). Online verfügbar unter https://www.nist.gov/programs-projects/face-recognition-grand-challenge-frgc, zuletzt aktualisiert am 25.08.2016, zuletzt geprüft am 04.12.2020.

NIST (2017): Face Recognition Technology (FERET). Online verfügbar unter https://www.nist.gov/programs-projects/face-recognition-technology-feret, zuletzt aktualisiert am 13.07.2017, zuletzt geprüft am 04.12.2020.

NIST (2019): Face in Video Evaluation (FIVE). Online verfügbar unter https://www.nist.gov/programs-projects/face-video-evaluation-five, zuletzt aktualisiert am 08.05.2019, zuletzt geprüft am 04.12.2020.

NIST (2020): Face Recognition Vendor Test (FRVT). Online verfügbar unter https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt, zuletzt aktualisiert am 04.12.2020, zuletzt geprüft am 04.12.2020.

NIST (2021): FRVT 1:N Identification. Online verfügbar unter https://pages.nist.gov/frvt/html/frvt1N.html, zuletzt aktualisiert am 05.08.2021, zuletzt geprüft am 24.08.2021.

Nuance (Hg.) (2021): Dragon Spracherkennungssoftware. Online verfügbar unter https://www.nuance.com/de-de/index.html, zuletzt aktualisiert am 14.01.2021, zuletzt geprüft am 14.01.2021.

NZZ (2002): Biometrie ist die Kunst, die Eigenschaften. In: *Neue Zürcher Zeitung*.

NZZ (2009): Kein Pyro, weniger Alkohol, mehr Kontrolle; Koordiniertes Vorgehen im Kampf gegen Gewalt an Sportanlässen. In: *Neue Zürcher Zeitung*, S. 13.

NZZ (2016): Mit Videokameras und Netzen gegen Hooligans. In: *Neue Zürcher Zeitung*. Online verfügbar unter https://www.nzz.ch/schweiz/hooligan-konkordat-mit-videokameras-und-netzen-gegen-hooligans-ld.108923?reduced=true, zuletzt geprüft am 15.06.2021.

Odell, Julian; Mukerjee, Kunal (2007): Architecture, User Interface, and Enabling Technology in Windows Vista's Speech Systems. In: *IEEE Trans. Comput.* 56 (9), S. 1156–1168. DOI: 10.1109/TC.2007.1065.

Oettinger, Renate (2010): Flexibel Geld verwalten: Telefon-Banking: Anruf genügt. Hg. v. computerwoche. Online verfügbar unter https://www.computerwoche.de/a/telefon-banking-anruf-genuegt,1939393, zuletzt geprüft am 30.04.2021.

Ogawa, Atsunori; Hori, Takaaki; Nakamura, Atsushi (2012): Recognition rate estimation based on word alignment network and discriminative error type classification. In: IEEE Spoken Language Technology Workshop (SLT 2012). Miami, Florida, USA, 2 – 5 December 2012. Institute of Electrical and Electronics Engineers. Piscataway, NJ: IEEE, S. 113–118.

Olson, Parmy (2020): Facial Recognition's Next Big Play: the Sports Stadium. In: *Wall Street Journal*. Online verfügbar unter https://www.wsj.com/articles/facial-recognitions-next-big-play-the-sports-stadium-11596290400, zuletzt geprüft am 16.08.2020.

Omega Foundation (2000): Crowd Control Technologies (An appraisal of technologies for political control) Final Study. Working Document. European Parliament Directorate General for Research Directorate A The STOA Programme. Luxembourg (PE 168. 394/Fin. St).

O'Toole, Alice J.; Jonathon Phillips, P.; Jiang, Fang; Ayyad, Janet; Penard, Nils; Abdi, Hervé (2007): Face recognition algorithms surpass humans matching faces over changes in illumination. In: *IEEE Trans. Pattern Anal. Machine Intell.* 29 (9), S. 1642–1646. DOI: 10.1109/TPAMI.2007.1107.

Owen, Glyn (2017): British Cops Will Scan Every Fan's Face at the Champions League Final. In: *VICE*. Online verfügbar unter https://www.vice.com/en/article/d7bwny/british-cops-will-scan-every-fans-face-at-the-champions-league-final, zuletzt geprüft am 16.06.2021.

Paliscope (2021): Paliscope Platform. Online verfügbar unter https://www.paliscope.com/2018/12/10/integrating-new-tools-icac-investigators/, zuletzt geprüft am 09.06.2021.

Panasonic (2021): WV-ASF950 FacePro – Face Recognition Software using Deep Learning Technology, Security Solutions. Online verfügbar unter https://business.panasonic.co.uk/security-solutions/facepro-face-recognition-software-using-deep-learning-technology/wv-asf950, zuletzt geprüft am 20.04.2021.

Pantel, Jean Tori; Hajjir, Nurulhuda; Danyel, Magdalena; Elsner, Jonas; Abad-Perez, Angela Teresa; Hansen, Peter et al. (2020): Efficiency of Computer-Aided Facial Phenotyping (DeepGestalt) in Individuals With and Without a Genetic Syndrome: Diagnostic Accuracy Study. In: *Journal of medical Internet research* 22 (10), e19263. DOI: 10.2196/19263.

Papa, Roberta; Pietruszak, Thomas (2015): § 17 Datenschutz im Personalwesen. In: Nicolas Passadelis (Hg.): Datenschutzrecht. Beraten in Privatwirtschaft und öffentlicher Verwaltung. Basel: Helbing Lichtenhahn (Handbücher für die Anwaltspraxis).

Pareek, Vishakha; Sharma, R. K. (2016): Coronary heart disease detection from voice analysis. In: Siddharth Saxena (Hg.): IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS). March 5–6, 2016, Maulana Azad National Institute of Technology, Bhopal, India. Institute of Electrical and Electronics Engineers. Piscataway, NJ: IEEE, S. 1–6.

Park, Youngseok; Choi, Hyunsang; Cho, Sanghyun; Kim, Young-Gab (2019): Security Analysis of Smart Speaker: Security Attacks and Mitigation. In: *Computers, Materials & Continua* 61 (3), S. 1075–1090. DOI: 10.32604/cmc.2019.08520.

Passadelis, Nicolas (Hg.) (2015): Datenschutzrecht. Beraten in Privatwirtschaft und öffentlicher Verwaltung. Basel: Helbing Lichtenhahn (Handbücher für die Anwaltspraxis).

Pawlaszczyk, D.; Friese, J.; Hummert, C. (2019): Alexa, tell me – A forensic examination of the Amazon Echo Dot 3 rd Generation. In: *ijcse* 7 (11), S. 20–29. DOI: 10.26438/ijcse/v7i11.2029.

Pelc, Karine; Kornreich, Charles; Foisy, Marie-Line; Dan, Bernard (2006): Recognition of emotional facial expressions in attention-deficit hyperactivity disorder. In: *Pediatric neurology* 35 (2), S. 93–97. DOI: 10.1016/j.pediatrneurol.2006.01.014.

Perani, Martina (2018): KI im Schweizer Gesundheitswesen: Chancen und Herausforderungen. Fachbeitrag FFHS. Hg. v. Netzwoche. Online verfügbar unter https://www.netzwoche.ch/news/2018-12-07/ki-im-schweizer-gesundheitswesen-chancen-und-herausforderungen, zuletzt geprüft am 28.04.2021.

Peters, Christina (2018): Augmented Reality: Kontaktlinsen mit Bildschirm «in 10 bis 15 Jahren». In: *Wirtschaftswoche*. Online verfügbar unter https://www.wiwo.de/futureboard/augmented-reality-kontaktlinsen-mit-bildschirm-in-10-bis-15-jahren/23116006.html, zuletzt geprüft am 05.05.2021.

Pfister, Beat; Kaufmann, Tobias (2017): Sprachverarbeitung. Grundlagen und Methoden der Sprachsynthese und Spracherkennung. 2., aktualisierte und erweiterte Auflage. Berlin: Springer Vieweg.

Phillips, P. Jonathon; Yates, Amy N.; Hu, Ying; Hahn, Carina A.; Noyes, Eilidh; Jackson, Kelsey et al. (2018): Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. In: *Proceedings of the National Academy of Sciences of the United States of America* 115 (24), S. 6171–6176. DOI: 10.1073/pnas.1721355115.

Picard, Rosalind W. (2000): Affective computing. First paperback edition. Cambridge, Massachusetts, London: The MIT Press.

Piltch, Avram (2013): Dual-Eye Augmented Reality Goggles Demo Brilliantservice VR Headset. In: *Laptop*. Online verfügbar unter https://web.archive.org/web/20131029185157/http://blog.laptopmag.com/viking-ar-goggles, zuletzt geprüft am 04.05.2021.

PimEyes (2021): PimEyes: Face Recognition Search Engine and Reverse Image Search. Online verfügbar unter https://pimeyes.com/en, zuletzt geprüft am 09.06.2021.

Plaßmann, Wilfried (2016): Mehrfachübertragung – Multiplexverfahren. In: Wilfried Plaßmann und Detlef Schulz (Hg.): Handbuch Elektrotechnik. 7., neu bearbeitete Auflage. Wiesbaden: Springer Vieweg, S. 1257–1260.

Pluta, Werner (2019): Emotionen erkennen: Ein Lächeln macht noch keinen Frohsinn. Hg. v. Golem.de. Online verfügbar unter https://www.golem.de/news/dfki-projekt-empat-ein-laecheln-macht-noch-keinen-frohsinn-1902-139370.html, zuletzt geprüft am 31.03.2021.

Pohle, Jörg (2016): PERSONAL DATA NOT FOUND: Personenbezogene Entscheidungen als überfällige Neuausrichtung im Datenschutz. In: *Datenschutz Nachrichten* 39, S. 14–19.

Pokorny, Florian B.; Schuller, Björn; Marschik, Peter B.; Brueckner, Raymond; Nyström, Pär; Cummins, Nicholas et al. (2017): Earlier Identification of Children with Autism Spectrum Disorder: An Automatic Vocalisation-Based Approach. In: Interspeech, 20–24 August 2017. ISCA: ISCA, S. 309–313.

Portmann, Wolfgang; Rudolph, Roger (2020): Kommentar zu Art. 328b OR. In: Widmer Lüchinger und David Oser (Hg.): Obligationenrecht: OR, I: Art. 1–529 OR. 7. Auflage. Basel: Helbing & Lichtenhahn.

PostFinance (Hg.) (2021): Authentifizierung mit Stimmerkennung. Online verfügbar unter https://www.postfinance.ch/de/support/persoenliche-daten/authentifizierung-stimmerkennung.html, zuletzt geprüft am 30.04.2021.

Pounder, C. N. M. (2008): Nine principles for assessing whether privacy is protected in a surveillance society. In: *IDIS* 1 (1), S. 1–22. DOI: 10.1007/s12394-008-0002-2.

Prainsack, Barbara (2013): Personalisierte Medizin aus Sicht des Patienten – Nutzen oder Überforderung? In: Personalisierte Medizin – der Patient als Nutznießer oder Opfer?, S. 23–33.

Preto, Sara (2019): Emotion-reading algorithms cannot predict intentions via facial expressions. Hg. v. USC. Online verfügbar unter https://news.usc.edu/160360/algorithms-emotions-facial-expressions-predict-intentions/, zuletzt geprüft am 01.04.2021.

Prieur, Yvonne (2017): Im Spannungsfeld zwischen Selbst- und Fremdvermessung. In: *Jusletter*. Online verfügbar unter https://jusletter.weblaw.ch/juslissues/2017/918/im-spannungsfeld-zwi_6fec321652.html__ONCE&login=false.

Raab, Charles D. (2020): Information privacy, impact assessment, and the place of ethics. In: *Computer Law & Security Review* 37, S. 105404. DOI: 10.1016/j.clsr.2020.105404.

Rampini, Corrado (2014): Vorbemerkungen zu Art. 12–15 DSG. In: Urs Maurer-Lambrou und Heinrich Honsell (Hg.): Datenschutzgesetz, Öffentlichkeitsgesetz. 3. Aufl. Basel: Helbing Lichtenhahn (Basler Kommentar).

Rana, Rajib; Latif, Siddique; Gururajan, Raj; Gray, Anthony; Mackenzie, Geraldine; Humphris, Gerald; Dunn, Jeff (2019): Automated screening for distress: A perspective for the future. In: *European journal of cancer care* 28 (4), e13033. DOI: 10.1111/ecc.13033.

Rashida, Richardson (2021): Facial Recognition in the Public Sector: The Policy Landscape. Online verfügbar unter https://www.gmfus.org/publications/facial-recognition-public-sector-policy-landscape, zuletzt aktualisiert am 02.04.2021, zuletzt geprüft am 15.06.2021.

Rauss, Patrick J.; Phillips, Jonathan; Hamilton, Mark K.; DePersia, A. Trent (1997): FERET (Face Recognition Technology) program. In: David H. Schaefer und Elmer F. Williams (Hg.): 25th AIPR Workshop: Emerging Applications of Computer Vision. Washington, DC, Wednesday 16 October 1996: SPIE (SPIE Proceedings), S. 253–263.

Rebiger, Simon (2017): Russische Demonstranten per Gesichtserkennungs-Software identifiziert. Netzpolitik.org. Online verfügbar unter https://netzpolitik.org/2017/russische-demonstranten-per-gesichtserkennungs-software-identifiziert/, zuletzt aktualisiert am 15.06.2021, zuletzt geprüft am 15.06.2021.

Reclaim Your Face (2022): Secretive. Unlawful. Inhumane. Online verfügbar unter https://reclaimyourface.eu/the-problem/, zuletzt aktualisiert am 23.11.2020, zuletzt geprüft am 15.02.2022.

Regierungsrat des Kantons Zürich (2018): Fünf Ausreise-Schleusen für die automatisierte Passkontrolle am Flughafen Zürich | Kanton Zürich. Der Regierungsrat (Medienmitteilung). Online verfügbar unter https://www.zh.ch/de/news-uebersicht/medienmitteilungen/2018/05/1805291m.html, zuletzt aktualisiert am 15.06.2021, zuletzt geprüft am 15.06.2021.

Reisman, Dillon; Schultz, Jason; Crawford, Kate; Whittaker, Meredith (2018): Algorithmic Impact Assessment: A Practical Framework for Public Agency Accountability. Hg. v. AI Now Institute. Online verfügbar unter https://www.ftc.gov/system/files/documents/public_comments/2018/08/ftc-2018-0048-d-0044-155168.pdf, zuletzt geprüft am 10.09.2021.

retorio GmbH (2021): Face, Emotion, And Body Pose Recognition For Sales Recruitment With AI. Online verfügbar unter https://www.retorio.com/blog/face-emotion-and-body-pose-recognition-for-sales-recruitment-with-ai, zuletzt aktualisiert am 25.03.2021, zuletzt geprüft am 31.03.2021.

Reusser, Ruth; Lüscher, Kurt (2014): Kommentar zu Art. 11 BV. In: Bernhard Ehrenzeller (Hg.): Die schweizerische Bundesverfassung. St. Galler Kommentar. 3. Aufl. Zürich: dike; Schulthess.

Reuter, Markus (2017): Surveillance under Surveillance: Weltkarte der Videoüberwachung wächst rasant. Netzpolitik.org. Online verfügbar unter https://netzpolitik.org/2017/surveillance-under-surveillance-weltkarte-der-videoueberwachung-waechst-rasant/, zuletzt aktualisiert am 15.06.2021, zuletzt geprüft am 15.06.2021.

Reuter, Markus (2019a): San Francisco erschwert Anschaffung von Überwachungstechnologien und verbietet Gesichtserkennung. Netzpolitik.org. Online verfügbar unter https://netzpolitik.org/2019/san-francisco-erschwert-anschaffung-von-ueberwachungstechnologien-und-verbietet-gesichtserkennung/, zuletzt aktualisiert am 15.05.2019, zuletzt geprüft am 15.02.2022.

Reuter, Markus (2019b): Gesichtserkennung statt Klassenbuch: Schule in Schweden kassiert Strafe. Netzpolitik.org. Online verfügbar unter https://netzpolitik.org/2019/gesichtserkennung-statt-klassenbuch-schule-in-schweden-kassiert-strafe/, zuletzt aktualisiert am 27.05.2022.

Riccio, Gianluca (2019): Eine KI erstellt eine «Sprachbank» für diejenigen, die ihre Stimme verlieren müssen. Online verfügbar unter https://www.futuroprossimo.it/de/2019/11/unai-crea-una-banca-vocale-che-perdere-la-voce-se-la-perdiamo/, zuletzt geprüft am 19.04.2021.

Richter, Felix (2019): AI Now Report 2019 – Forschungsinstitut warnt vor sozialen Folgen von KI. Online verfügbar unter https://netzpolitik.org/2019/forschungsinstitut-warnt-vor-sozialen-folgen-von-ki/, zuletzt geprüft am 31.03.2021.

Ringeval, Fabien; Schuller, Björn; Valstar, Michel; Cowie, Roddy; Kaya, Heysem; Schmitt, Maximilian et al. (2018): AVEC 2018 Workshop and Challenge. In: Fabien Ringeval (Hg.): Proceedings of the 2018 on AudioVisual Emotion Challenge and Workshop. MM '18: ACM Multimedia Conference. Seoul Republic of Korea, 22 10 2018. New York, NY: ACM (ACM Conferences), S. 3–13.

Ringeval, Fabien; Schuller, Björn; Valstar, Michel; Gratch, Jonathan; Cowie, Roddy; Scherer, Stefan et al. (2017): AVEC 2017. In: Fabien Ringeval (Hg.): Proceedings of the 7th Annual Workshop on AudioVisual Emotion Challenge. MM '17: ACM Multimedia Conference. Mountain View California USA, 23 10 2017. New York, NY: ACM, S. 3–9.

Riniker, Maja (2021): 21.7896 Frage, Gesichtserkennung zur Identifizierung im öffentlichen Raum. Online verfügbar unter https://www.parlament.ch/de/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20217896, zuletzt geprüft am 27.05.2022.

Rivas, Bruno (2021): Google Assistant bekommt Konkurrenz: Amazon Alexa startet als Pilotprojekt in der Schweiz. In: *vybe*. Online verfügbar unter https://vybe.ch/google-assistant-bekommt-konkurrenz-amazon-alexa-startet-als-pilotprojekt-in-der-schweiz/, zuletzt geprüft am 05.03.2021.

Roberts, Andrew (2015): A republican account of the value of privacy. In: *European Journal of Political Theory* 14 (3), S. 320–344. DOI: 10.1177/1474885114533262.

Rodrigues, Manuel; Durães, Dalila; Santos, Ricardo; Analide, Cesar (2021): Emotion Detection Throughout the Speech. In: Kohei Arai, Supriya Kapoor und Rahul Bhatia (Hg.): Intelligent Systems and Applications. Proceedings of the 2020 Intelligent Systems Conference (IntelliSys) Volume 2. Cham: Springer (Advances in Intelligent Systems and Computing), S. 304–314.

Rodríguez-Blanque, Raquel; Sanchez-Garcia, Juan Carlos; Sanchez-Lopez, Antonio Manuel; Expósito-Ruiz, Manuela; Aguilar-Cordero, Maria Jose (2019): Randomized Clinical Trial of an Aquatic Physical Exercise Program During Pregnancy. In: *Journal of obstetric, gynecologic, and neonatal nursing : JOGNN* 48 (3), S. 321–331. DOI: 10.1016/j.jogn.2019.02.003.

Rosenstein, M.; Foltz, P. W.; DeLisi, L. E.; Elvevåg, B. (2015): Language as a biomarker in those at high-risk for psychosis. In: *Schizophrenia research* 165 (2-3), S. 249–250. DOI: 10.1016/j.schres.2015.04.023.

Rosenthal, David (2020): Das neue Datenschutzgesetz. In: *Jusletter* (1045).

Rössler, Beate (2001): Der Wert des Privaten. Frankfurt am Main: Suhrkamp.

Roßnagel, Alexander; Bile, Tamer; Nebel, Maxi; Geminn, Christian; Karaboga, Murat; Ebbers, Frank et al. (2020): White Paper Einwilligung: Möglichkeiten und Fallstricke aus der Konsumentenperspektive. Hg. v. Forum Privatheit. Fraunhofer Institut für System- und Innovationsforschung. Karlsruhe. Online verfügbar unter https://www.forum-privatheit.de/download/einwilligung/.

Roßnagel, Alexander; Richter, Philipp (2017): Aufwachsen in virtuellen und technologisierten Welten. Herausforderungen der Datensammlung, Vernetzung, Kommerzialisierung und neuen Überwachungstechnologien für Jugendliche. In: Sachverständigenkommission 15 (Hg.): Zwischen Freiräumen, Familie, Ganztagsschule und virtuellen Welten – Persönlichkeitsentwicklung und Bildungsanspruch im Jugendalter. München, S. 205–260.

Rothrock, Kevin (2016): Facial Recognition Service Becomes a Weapon Against Russian Porn Actresses. In: *Global Voices Advox*. Online verfügbar unter https://advox.globalvoices.org/2016/04/22/facial-recognition-service-becomes-a-weapon-against-russian-porn-actresses/, zuletzt geprüft am 12.05.2021.

Rouast, Philipp V.; Adam, Marc; Chiong, Raymond (2019): Deep Learning for Human Affect Recognition: Insights and New Developments. In: *IEEE Trans. Affective Comput.*, S. 1. DOI: 10.1109/TAFFC.2018.2890471.

Rupareliya, Pratik (2021): AR Smart Glasses: Applications, Challenges & Future Potential [2020]. In: *Intuz*. Online verfügbar unter https://www.intuz.com/blog/augmented-reality-glass-application-usecases-challenges-future-potential, zuletzt geprüft am 04.05.2021.

Russell, Brandon (2020): Google no longer stores audio recordings by default for all Assistant users. Hg. v. xda.developers. Online verfügbar unter https://www.xda-developers.com/google-assistant-disables-saving-audio-recordings-all-users/, zuletzt aktualisiert am 05.08.2020, zuletzt geprüft am 10.03.2021.

Russell, Jon (2017): China's CCTV surveillance network took just 7 minutes to capture BBC reporter. In: *TechCrunch*. Online verfügbar unter https://techcrunch.com/2017/12/13/china-cctv-bbc-reporter/?guccounter=1, zuletzt geprüft am 15.06.2021.

Ryser, Daniel (2019): «Die Langstrasse ist komplett überwacht». Hg. v. Republik. Online verfügbar unter https://www.republik.ch/2019/02/21/die-langstrasse-ist-komplett-ueberwacht, zuletzt aktualisiert am 15.06.2021, zuletzt geprüft am 15.06.2021.

Saltman, Kenneth J. (2016): Scripted Bodies. Corporate Power, Smart Technologies, and the Undoing of Public Education. Milton: Taylor and Francis (Critical Interventions).

Samatas, Minas (2014): The Super-Panopticon Scandal of the Athens 2004 Olympics and its Legacy.

Sarikaya, Ruhi (2019): How Alexa Learns. Online verfügbar unter https://blogs.scientificamerican.com/observations/how-alexa-learns/, zuletzt aktualisiert am 08.02.2022, zuletzt geprüft am 08.02.2022.

SATW (Hg.) (2019): Recommendations for an AI Strategy in Switzerland. Online verfügbar unter https://www.satw.ch/fileadmin/user_upload/documents/02_Themen/08_Kuenstliche-Intelligenz/SATW-Swiss_AI_Strategy.pdf, zuletzt geprüft am 10.09.2021.

Saxena, Anvita; Khanna, Ashish; Gupta, Deepak (2020): Emotion Recognition and Detection Methods: A Comprehensive Survey. In: *AIS* 2 (1), S. 53–79. DOI: 10.33969/AIS.2020.21005.

Saz, Oscar; Yin, Shou-Chun; Lleida, Eduardo; Rose, Richard; Vaquero, Carlos; Rodríguez, William R. (2009): Tools and Technologies for Computer-Aided Speech and Language Therapy. In: *Speech Communication* 51 (10), S. 948–967. DOI: 10.1016/j.specom.2009.04.006.

Schachtler, Peter (2019): Stimmerkennung – Der neue digitale Fingerabdruck. Hg. v. Hochschule Luzern. Online verfügbar unter https://blog.hslu.ch/diginect/2018/05/03/stimmerkennung-der-neue-digitale-fingerabdruck/, zuletzt geprüft am 18.04.2021.

Schär Gmelch, Marcel (2019): Psychotherapie der Zukunft. In: Daniel Süss, Christoph Negri und Christoph Steinebach (Hg.): Angewandte Psychologie. Beiträge zu einer menschenwürdigen Gesellschaft. Berlin/Heidelberg: Springer, S. 79–92.

Scherer, Klaus R. (1996): Emotion. In: Wolfgang Stroebe (Hg.): Sozialpsychologie. Eine Einführung. 3., erw. und überarb. Aufl. Berlin: Springer, S. 293–330.

Scherer, Klaus R. (2005): What are emotions? And how can they be measured? In: *Social Science Information* 44 (4), S. 695–729. DOI: 10.1177/0539018405058216.

Scherer, Klaus R.; Moors, Agnes (2019): The Emotion Process: Event Appraisal and Component Differentiation. In: *Annual review of psychology* 70, S. 719–745. DOI: 10.1146/annurev-psych-122216-011854.

Scherer, Klaus R.; Mortillaro, Marcello; Rotondi, Irene; Sergi, Ilaria; Trznadel, Stéphanie (2018): Appraisal-driven facial actions as building blocks for emotion inference. In: *Journal of personality and social psychology* 114 (3), S. 358–379. DOI: 10.1037/pspa0000107.

Scherer, Klaus R.; Sundberg, Johan; Fantini, Bernardino; Trznadel, Stéphanie; Eyben, Florian (2017): The expression of emotion in the singing voice: Acoustic patterns in vocal performance. In: *The Journal of the Acoustical Society of America* 142 (4). DOI: 10.1121/1.5002886.

Scherer, Stefan; Stratou, Giota; Gratch, Jonathan; Morency, Louis-Philippe (2013): Investigating voice quality as a speaker-independent indicator of depression and PTSD. In: *Interspeech*.

Schiller, Amy; McMahon, John (2019): Alexa, Alert Me When the Revolution Comes: Gender, Affect, and Labor in the Age of Home-Based Artificial Intelligence. In: *New Political Science* 41 (2), S. 173–191. DOI: 10.1080/07393148.2019.1595288.

Schimmel, Kimberly S. (2011): From 'Violence-complacent' to 'Terrorist-ready'. In: *Urban Studies* 48 (15), S. 3277–3291. DOI: 10.1177/0042098011422396.

Schindler, Stephan (2019). Super-Recognizer: Die menschliche Alternative zur bio-metrischen Gesichtserkennung? In: *ZD-Aktuell* (06730).

Schindler, Stephan; Hornung, Gerrit (2021): Datenschutz bei der biometrischen Gesichtserkennung. Künstliche Intelligenz und Mustererkennung als Herausforderung für das Recht. In: *Datenschutz und Datensicherheit*, S. 515–521.

Schmidlin, Remo R. (2022): Voice Recognition im Arbeitsverhältnis – eine datenschutzrechtliche Analyse. In: *sg*. DOI: 10.21257/sg.201.

Schmidt, Tobias (2019): Der Doktor in der Hosentasche: Was taugen Gesundheits-Apps? Online verfügbar unter https://www.noz.de/deutschland-welt/politik/artikel/1928416/der-doktor-in-der-hosentasche-was-taugen-gesundheits-apps, zuletzt aktualisiert am 07.11.2019, zuletzt geprüft am 15.09.2021.

Schneider, Jan (2020): Stimmanalyse per App – Corona-Diagnose am eigenen Smartphone. Hg. v. ZDF. Online verfügbar unter https://www.zdf.de/nachrichten/panorama/corona-erkennung-husten-100.html, zuletzt geprüft am 19.04.2021.

Schreiber, Markus; Joss, Mara (2020): Der «Chilling Effect» auf die Grundrechtsausübung. Hg. v. ZB1 (10). Online verfügbar unter https://www.zbl-online.ch/de/artikel/2504-0731-2020-0055/der-chilling-effect-auf-die-grundrechtsausubung.

Schuller, B.; Steidl, S.; Batliner, A.; Hantke, S.; Hönig, F.; Orozco-Arroyave, J. R. et al. (2015): The INTERSPEECH 2015 computational paralinguistics challenge: Nativeness, Parkinson's & eating condition. In: Proceedings of the Annual Conference of the Interna-

tional Speech Communication Association, INTERSPEECH 2015-January. Online verfügbar unter https://www.scopus.com/inward/record.uri?eid=2-s2.0-84959078348&partnerID=40&md5=fc7374727f364f131cfbac2a23382c48.

Schuller, B.; Steidl, S.; Batliner, A.; Schiel, F.; Krajewski, J. (2011): The INTERSPEECH 2011 speaker state challenge. In: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH. Online verfügbar unter https://www.scopus.com/inward/record.uri?eid=2-s2.0-84865716918&partnerID=40&md5=79de1b18f3ea246dc28c81f47669ef56.

Schuller, Björn; Steidl, Stefan; Batliner, Anton; Bergelson, Elika; Krajewski, Jarek; Janott, Christoph et al. (2017): The INTERSPEECH 2017 Computational Paralinguistics Challenge: Addressee, Cold & Snoring. In: Interspeech 2017. 20–24 August 2017. ISCA: ISCA, S. 3442–3446.

Schuller, Björn W.; Schuller, Dagmar M.; Qian, Kun; Liu, Juan; Zheng, Huaiyuan; Li, Xiao (2020): COVID-19 and Computer Audition: An Overview on What Speech & Sound Analysis Could Contribute in the SARS-CoV-2 Corona Crisis. Online verfügbar unter http://arxiv.org/pdf/2003.11117v1.

Schuller, Dagmar (2019): healthAI – Wie Audio Intelligence die Diagnostik und Therapie im Gesundheitswesen revolutionieren kann. Hg. v. Bitkom. Online verfügbar unter https://www.bitkom.org/Bitkom/Publikationen/healthAI-Wie-Audio-Intelligence-die-Diagnostik-und-Therapie-im-Gesundheitswesen-revolutionieren-kann.

Schulz, Sven (2012): Siri, Vlingo & Co.: So funktioniert die Spracherkennung mit dem Handy. Hg. v. Welt. Online verfügbar unter https://www.welt.de/wirtschaft/webwelt/article13811437/So-funktioniert-die-Spracherkennung-mit-dem-Handy.html, zuletzt geprüft am 14.12.2020.

Schulzki-Haddouti, Christiane (2007): Biometrische Gesichtserkennung für Fahndung ungeeignet. In: *ingenieur.de*. Online verfügbar unter https://www.ingenieur.de/karriere/arbeitsleben/arbeitssicherheit/biometrische-gesichtserkennung-fuer-fahndung-ungeeignet/, zuletzt geprüft am 11.03.2021.

Schweiz Tipps (Hg.) (2021): Die besten Alexa Tipps & Skills: Diese Tricks sollte jeder Amazon Echo Besitzer kennen. Online verfügbar unter https://www.schweiztipps.ch/die-besten-alexa-tipps-skills/3853/, zuletzt geprüft am 19.03.2021.

Schweizerische Eidgenossenschaft (2006): Einführung biometrischer Ausweise. Online verfügbar unter https://www.newsd.admin.ch/newsd/message/attachments/3981.pdf, zuletzt geprüft am 14.01.2021.

Scott, Jeramie (2020): Coalition for Ban on Corporate Use of Facial Recognition. Letter Facial Recognition Technology Suspension directed towards PCLOB. Online verfügbar unter https://epic.org/wp-content/uploads/privacy/facerecognition/PCLOB-Letter-FRT-Suspension.pdf, zuletzt geprüft am 15.02.2022.

Security Industry Association (2020): SIA Principles for the Responsible and Effective Use of Facial Recognition Technology. Online verfügbar unter https://www.securityindustry.org/report/sia-principles-for-the-responsible-and-effective-use-of-facial-recognition-technology/, zuletzt aktualisiert am 11.08.2020, zuletzt geprüft am 15.02.2022.

Security Industry Association (2021): Face Facts: Dispelling Common MYTHS Associated With Facial Recognition Technology. Online verfügbar unter https://www.securityindustry.org/wp-content/uploads/2019/06/facial-recognition-20193.pdf, zuletzt geprüft am 15.06.2021.

Selinger, Evan; Leong, Brenda (2021): The Ethics of Facial Recognition Technology. In: *SSRN Journal*. DOI: 10.2139/ssrn.3762185.

Selouani, Sid-Ahmed; Sidi Yakoub, Mohammed; O'Shaughnessy, Douglas (2009): Alternative Speech Communication System for Persons with Severe Speech Disorders. In: *EURASIP J. Adv. Signal Process.* 2009 (1), S. 1–12. DOI: 10.1155/2009/540409.

Setiowati, Sulis; Zulfanahri; Franita, Eka Legya; Ardiyanto, Igi (2017): A review of optimization method in face recognition: Comparison deep learning and non-deep learning methods. In: 9th International Conference on Information Technology and Electrical Engineering (ICITEE). 12–13 Oct. 2017. Phuket. ICITEE; Institute of Electrical and Electronics Engineers; IEEE Control Systems Society; IEEE Computational Intelligence Society; Electrical Engineering, Electronics Computer, Telecommunications and Information Technology Association. Piscataway, NJ: IEEE, S. 1–6.

Seubert, Sandra (2017): Das Vermessen kommunikativer Räume. In: *Forschungsjournal Soziale Bewegungen* 30 (2), S. 124–133. DOI: 10.1515/fjsb-2017-0033.

Sheldon, Robert (2020): Biometrische Authentifizierung kann mobile Geräte gefährden. Online verfügbar unter https://www.computerweekly.com/de/tipp/Biometrische-Authentifizierung-kann-mobile-Geraete-gefaehrden, zuletzt geprüft am 08.06.2021.

Sheu, Jia-Shing; Hsieh, Tsu-Shien; Shou, Ho-Nien (2014): Automatic Generation of Facial Expression Using Triangular Geometric Deformation. In: *Journal of Applied Research and Technology* 12 (6), S. 1115–1130. DOI: 10.1016/S1665-6423(14)71671-2.

Shneiderman, Ben (2000): The limits of speech recognition. In: *Commun. ACM* 43 (9), S. 63–65. DOI: 10.1145/348941.348990.

Siddiqui, Mohammad Faridul Haque; Javaid, Ahmad Y. (2020): A Multimodal Facial Emotion Recognition Framework through the Fusion of Speech with Visible and Infrared Images. In: *MTI* 4 (3). DOI: 10.3390/mti4030046.

Sigrist, Martin (2014): Staatsschutz oder Datenschutz? Die Vereinbarkeit präventiver Datenbearbeitung zur Wahrung der inneren Sicherheit mit dem Grundrecht auf informationelle Selbstbestimmung. Zugl.: Zürich, Univ., Diss., 2014. Zürich: Schulthess (Zürcher Studien zum öffentlichen Recht, 219).

Simmler, Monika; Canova, Giulia (2021): Gesichtserkennunstechnologie: Die «smarte» Polizeiarbeit auf dem rechtlichen Prüfstand. In: *Sicherheit & Recht*, 3, S. 105–117.

Simmons, Dan (2017): BBC fools HSBC voice recognition security system. Online verfügbar unter https://www.bbc.com/news/technology-39965545, zuletzt geprüft am 30.04.2021.

Simonazzi, Nicolas; Salotti, Jean-Marc; Dubois, Caroline; Seminel, Dominique (2021): Emotion Detection Based on Smartphone Using User Memory Tasks and Videos. In: Ahram und Di Cecco (Hg.): Human Interaction, Emerging Technologies and Future Applications III, Bd. 1253. [S.l.]: Springer (Advances in Intelligent Systems and Computing), S. 244–249.

Simonite, Tom (2018): When It Comes to Gorillas, Google Photos Remains Blind. Online verfügbar unter https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/, zuletzt geprüft am 14.01.2021.

Singer, Natasha; Metz, Cade (2019): Many Facial-Recognition Systems Are Biased, Says U.S. Study. In: *The New York Times*, 19.12.2019. Online verfügbar unter https://www.nytimes.com/2019/12/19/technology/facial-recognition-bias.html, zuletzt geprüft am 29.04.2022.

Skeem, Jennifer; Scurich, Nicholas; Monahan, John (2020): Impact of risk assessment on judges' fairness in sentencing relatively poor defendants. In: *Law and human behavior* 44 (1), S. 51–59. DOI: 10.1037/lhb0000360.

Söderlund, Göran B. W.; Jobs, Elisabeth Nilsson (2016): Differences in Speech Recognition Between Children with Attention Deficits and Typically Developed Children Disappear When Exposed to 65 dB of Auditory Noise. In: *Frontiers in psychology* 7, S. 34. DOI: 10.3389/fpsyg.2016.00034.

Softjourn (Hg.) (2021): Is Voice Authentication Secure Enough to be Your New Password? Online verfügbar unter https://softjourn.com/blog/article/security-considerations-in-voice-authentication, zuletzt geprüft am 30.04.2021.

Soleymani, M.; Lichtenauer, J.; Pun, T.; Pantic, M. (2012): A Multimodal Database for Affect Recognition and Implicit Tagging. In: *IEEE Trans. Affective Comput.* 3 (1), S. 42–55. DOI: 10.1109/T-AFFC.2011.25.

Sonde One (2021): Sonde Health. Boston, MA – USA. Online verfügbar unter https://www.sondehealth.com/sondeone-page, zuletzt geprüft am 19.04.2021.

Song, Lingxue; Liu, Changsong (2018): Face Liveness Detection Based on Joint Analysis of RGB and Near-Infrared Image of Faces. In: *Electronic Imaging* 2018 (10), 373-1-373-6. DOI: 10.2352/ISSN.2470-1173.2018.10.IMAWM-373.

Sonos Inc (Hg.) (2021): Alexa auf Sonos. Online verfügbar unter https://www.sonos.com/de-de/alexa-on-sonos, zuletzt geprüft am 19.03.2021.

Spitch (2019): Voice biometric verification for faster and more secure customer calls processing. Online verfügbar unter https://www.spitch.ch/upload/iblock/906/90615629e3609246629c45b1c93d61e9.pdf, zuletzt geprüft am 30.04.2021.

Spitch (2020): Migros Bank wählt Spitch-Lösung zur Identifizierung ihrer Kunden. Online verfügbar unter https://www.spitch.ch/de/news/2020.07.13.page, zuletzt geprüft am 30.04.2021.

Springer, Aaron; Hollis, Victoria; Whittaker, Steve (2018): Dice in the Black Box: User Experiences with an Inscrutable Algorithm. Online verfügbar unter http://arxiv.org/pdf/1812.03219v1.

SRF (2019): Heikle persönliche Daten – Widerstand gegen die elektronische Stimmerkennung der Postfinance. Online verfügbar unter https://www.srf.ch/news/schweiz/heikle-persoenliche-daten-widerstand-gegen-die-elektronische-stimmerkennung-der-postfinance, zuletzt aktualisiert am 17.08.2019, zuletzt geprüft am 18.02.2022.

SRF (2020): Automatische Gesichtserkennung – So einfach ist es, eine Überwachungsmaschine zu bauen. Online verfügbar unter https://www.srf.ch/news/schweiz/automa-

tische-gesichtserkennung-so-einfach-ist-es-eine-ueberwachungsmaschine-zu-bauen, zuletzt aktualisiert am 07.02.2020, zuletzt geprüft am 18.02.2022.

St. John, Allen (2020): Yes, Your Smart Speaker Is Listening When It Shouldn't. Hg. v. Consumer Reports. Online verfügbar unter https://www.consumerreports.org/smart-speakers/yes-your-smart-speaker-is-listening-when-it-should-not/, zuletzt geprüft am 05.03.2021.

Staben, Julian (2016): Der Abschreckungseffekt auf die Grundrechtsausübung.

Städeli, Markus (2019): Die Smartspeaker von Amazon und Google verstehen auch Dialekt. In: *NZZ*, 14.12.2019. Online verfügbar unter https://magazin.nzz.ch/wirtschaft/smart-speaker-von-amazon-und-google-verstehen-auch-schweizerdeutsch-ld.1528583?reduced=true, zuletzt geprüft am 29.04.2022.

Stadtpolizei Zürich (2021): Zeitweise Videoüberwachung am Utoquai und Stadelhofen – Stadt Zürich. Stadtpolizei Zürich. Online verfügbar unter https://www.stadt-zuerich.ch/pd/de/index/stadtpolizei_zuerich/medien/medienmitteilungen/2021/maerz/zeitweise_videoueberwachungamutoquaiundstadelhofen.html, zuletzt aktualisiert am 15.06.2021, zuletzt geprüft am 15.06.2021.

Stahl, Bernd Carsten (2021): Artificial Intelligence for a Better Future. Cham: Springer.

Stalder, Felix (2002): Privacy is not the Antidote to Surveillance. In: *SS* 1 (1), S. 120–124. DOI: 10.24908/ss.v1i1.3397.

Stanley, Jay; Steinhardt, Barry (2002): Drawing a Blank: The failure of facial recognition technology in Tampa, Florida. An ACLU Special Report. American Civil Liberties Union, ACLU.

Stapf, I.; Meinert, J.; Krämer, N.; Quinn, R. Ammicht; Bieker, F. (2021a): Das Recht von Kindern und Jugendlichen auf Privatheit in digitalen Umgebungen: Handlungsempfehlungen des Forum Privatheit. In: Ingrid Stapf, Regina Ammicht Quinn, Michael Friedewald, Jessica Heesen und Nicole Krämer (Hg.): Aufwachsen in überwachten Umgebungen. Nomos Verlag, S. 351–376.

Stapf, Ingrid; Ammicht Quinn, Regina; Friedewald, Michael; Heesen, Jessica; Krämer, Nicole (2021b): Aufwachsen in überwachten Umgebungen. Nomos Verlag.

Stark, Luke (2019): Facial recognition is the plutonium of AI. In: *XRDS* 25 (3), S. 50–55. DOI: 10.1145/3313129.

Stark, Luke; Hoey, Jesse (2020): The Ethics of Emotion in AI Systems: OSF Preprints.

Stasak, Brian; Huang, Zhaocheng; Razavi, Sabah; Joachim, Dale; Epps, Julien (2021): Automatic Detection of COVID-19 Based on Short-Duration Acoustic Smartphone Speech Analysis. In: *J Healthc Inform Res*, S. 1–17. DOI: 10.1007/s41666-020-00090-4.

Stelkens (2021): Zur Digitalisierung häuslicher Gewalt im Internet of Things. STREIT – Feministische Rechtszeitschrift. Online verfügbar unter https://www.streit-fem.de/ausgaben/ausgaben,id-2019,ausgabe-1-470.html, zuletzt aktualisiert am 15.06.2021, zuletzt geprüft am 15.06.2021.

Stern (2000): Sie sind erkannt. In: *Stern*.

Stettner, Elisa: Sicherheit am Bahnhof. Dissertation. Duncker & Humblot. Online verfügbar unter http://elibrary.duncker-humblot.de/9783428551576/U1.

Steve Wright (1998): An appraisal of technologies of political control. Scientific and Technological Options Assessment STOA. Working document. PE 166.499, 6 January 1998. Online verfügbar unter http://aei.pitt.edu/5538/.

Stevenson, Megan (2018): Assessing Risk Assessment in Action. In: *Minnesota Law Review*. Online verfügbar unter https://scholarship.law.umn.edu/mlr/58.

Stolcke, Andreas; Droppo, Jasha (2017): Comparing Human and Machine Errors in Conversational Speech Transcription. In: Interspeech 2017. 20–24 August 2017. ISCA: ISCA, S. 137–141.

Stolyarov, Gleb; Tétrault-Farber, Gabrielle (2021): «Face control»: Russian police go digital against protesters. In: *Reuters Media*. Online verfügbar unter https://www.reuters.com/article/us-russia-politics-navalny-tech-idUSKBN2AB1U2, zuletzt geprüft am 15.06.2021.

Strategy Analytics (2020): Global Smart Speaker Sales Rose 6% to 30 Million Units in Q2 2020. Online verfügbar unter https://news.strategyanalytics.com/press-releases/press-release-details/2020/Strategy-Analytics-Global-Smart-Speaker-Sales-Rose-6-to-30-Million-Units-in-Q2-2020/default.aspx, zuletzt geprüft am 05.03.2021.

Suarez, Beatriz (2021): The Ethics of Digital Voice Assistants. Hg. v. Viterbi Conversations in Ethics. Online verfügbar unter https://vce.usc.edu/semester/spring-2021/the-ethics-of-digital-voice-assistants/, zuletzt aktualisiert am 26.10.2021, zuletzt geprüft am 04.05.2022.

Suchy, B. H.; Wolf, S. R.; Gebhard, A.; Paulus, D. (2001): Bildanalysesystem zur Erkennung einer Fazialisparese. In: *HNO* 49 (10), S. 814–817. DOI: 10.1007/s001060170029.

Surveillance under Surveillance (2022): Statistics – country: CH – all areas – all types all years – all months. Online verfügbar unter https://sunders.uber.space/en/stats/?pie=area&cols=country&vals=CH&year=all&month=all&time=single, zuletzt geprüft am 15.06.2021.

Swiss Digital Initiative (2021): Labels and Certifications for the Digital World. Online verfügbar unter https://www.swiss-digital-initiative.org/news/labels-and-certifications-for-the-digital-world/, zuletzt geprüft am 10.09.2021.

Swissmedic (2021): Regulierung Medizinprodukte. Online verfügbar unter https://www.swissmedic.ch/swissmedic/de/home/medizinprodukte/regulierung-medizinprodukte.html, zuletzt geprüft am 15.09.2021.

The Telegraph (Hg.) (2015): A history of banking: from coins to pings. Online verfügbar unter https://www.telegraph.co.uk/sponsored/finance/your-bank/10912973/history-banking-early-coins-contactless.html, zuletzt geprüft am 30.04.2021.

TheCGBros (2012): A Sci-Fi Short Film : «Sight» – by Sight Systems. YouTube. Online verfügbar unter https://www.youtube.com/watch?v=lK_cdkpazjI, zuletzt aktualisiert am 01.08.2012, zuletzt geprüft am 31.03.2022.

Theunissen, Georg (2019): «Voice» ist «the next big thing». In: *Wirtsch Inform Manag* 11 (3), S. 158–159. DOI: 10.1365/s35764-019-00182-w.

Thomas, Oliver; Ickerott, Ingmar; Berkemeier, Lisa; Werning, Sebastian; Zobel, Benedikt; Vogel, Jannis et al. (2020): GLASSHOUSE – Smart Glasses zur Unterstützung von Logistikdienstleistungen. In: Oliver Thomas und Ingmar Ickerott (Hg.): Smart Glasses:

Augmented Reality zur Unterstützung von Logistikdienstleistungen. Berlin/Heidelberg: Springer, S. 2–18.

Thornsby, Jessica (2020): 4 Ways to Delete Amazon's Voice Recordings. Hg. v. Make Tech Easier. Online verfügbar unter https://www.maketecheasier.com/smart-home/delete-amazons-voice-recordings/, zuletzt aktualisiert am 06.07.2020, zuletzt geprüft am 08.02.2022.

Thouvenin, Florent (2019): Privatversicherungen: Datenschutzrecht als Grenze der Individualisierung? In: Astrid Epiney und Déborah Sangsue (Hg.): Datenschutz und Gesundheitsrecht. Schulthess Verlag (Forum Europarecht, 40), S. 15–42.

Timms, Michael J. (2016): Letting Artificial Intelligence in Education Out of the Box: Educational Cobots and Smart Classrooms. In: *Int J Artif Intell Educ* 26 (2), S. 701–712. DOI: 10.1007/s40593-016-0095-y.

Tobler, Lukas (2019): Überwachungskameras: Was soll das eigentlich?. In: *Das Lamm*. Online verfügbar unter https://daslamm.ch/ueberwachungskameras-was-soll-das-eigentlich/, zuletzt geprüft am 15.06.2021.

t-online (Hg.) (2012): Telefonbanking – Vorteile und Risiken. Online verfügbar unter https://www.t-online.de/finanzen/geld-vorsorge/sparen-finanzieren/id_45995034/telefonbanking-vorteile-und-risiken.html, zuletzt aktualisiert am 29.04.2021, zuletzt geprüft am 30.04.2021.

Tremmel, Moritz (2014): Gesichtserkennung mit Google Glass: Nach ersten Apps jetzt auch die Polizei in Dubai. In: *netzpolitik.org*. Online verfügbar unter https://netzpolitik.org/2014/gesichtserkennung-mit-google-glass-nach-ersten-apps-jetzt-auch-die-polizei-in-dubai/, zuletzt geprüft am 04.05.2021.

Tsanas, Athanasios; Little, Max A.; McSharry, Patrick E.; Spielman, Jennifer; Ramig, Lorraine O. (2012): Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. In: *IEEE transactions on bio-medical engineering* 59 (5), S. 1264–1271. DOI: 10.1109/TBME.2012.2183367.

Turk, M.; Pentland, A. (1991): Eigenfaces for recognition. In: *Journal of cognitive neuroscience* 3 (1), S. 71–86. DOI: 10.1162/jocn.1991.3.1.71.

Universität St. Gallen (2019): Erklärbare KI: Einblick in die Black Boxes des Machine Learning. Online verfügbar unter https://www.unisg.ch/de/wissen/newsroom/aktuell/rssnews/meinung/2019/mai/explainable-ai-27mai2019, zuletzt geprüft am 17.09.2021.

Urech, Rahel (2019): Die Stimme entlarvt den Verbrecher. In: *Tages-Anzeiger*, 10.06.2019. Online verfügbar unter https://www.tagesanzeiger.ch/wissen/technik/die-stimme-entlarvt-den-verbrecher/story/22923932, zuletzt geprüft am 29.04.2022.

Uttinger, Ursula (2015): § 10 Datenschutz im Gesundheitswesen. In: Nicolas Passadelis (Hg.): Datenschutzrecht. Beraten in Privatwirtschaft und öffentlicher Verwaltung. Basel: Helbing Lichtenhahn (Handbücher für die Anwaltspraxis).

Vaidya, Tavish; Sherr, Micah (2019): You Talk Too Much: Limiting Privacy Exposure Via Voice Input. In: SPW. IEEE Symposium on Security and Privacy Workshops : proceedings : 23 May 2019, San Francisco, California, USA. 5/19/2019 – 5/23/2019. Los Alamitos, California: IEEE Computer Society, S. 84–91.

Vallor, Shannon (2015): Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character. In: *Philos. Technol.* 28 (1), S. 107–124. DOI: 10.1007/s13347-014-0156-9.

Valstar, Michel; Gratch, Jonathan; Schuller, Björn; Ringeval, Fabien; Cowie, Roddy; Pantic, Maja (Hg.) (2016): AVEC'16. Proceedings of the 6th International Workshop on Audio/Visual Emotion Recognition Challenge : October 16, 2016, Amsterdam, The Netherlands : co-located with: ACM Multimedia 2016. Unter Mitarbeit von Michel Valstar. MM ,16: ACM Multimedia Conference. Amsterdam The Netherlands, 16 10 2016 16 10 2016. AVEC; Association for Computing Machinery; International Workshop on Audio/Visual Emotion Recognition Challenge; Audio-Visual Emotion Recognition Challenge – Depression, Mood, and Emotion. New York, NY, USA: ACM.

van den Broek, Tijs; Merel Ooms; Michael Friedewald; van Lieshout, Marc; Sven Rung (2017): Privacy and security: Citizens\textquoteright desires for an equal footing. In: Michael Friedewald, J. Peter Burgess, Johann Cas, Walter Peissl und Rocco Bellanova (Hg.): Surveillance, Privacy and Security: Citizens' Perspectives. United Kingdom: Taylor & Francis, S. 15–35.

van Noorden, Richard (2020): The ethical questions that haunt facial-recognition research. In: *Nature* 587 (7834), S. 354–358. DOI: 10.1038/d41586-020-03187-3.

van Steenkiste, Sjoerd; Kurach, Karol; Schmidhuber, Jürgen; Gelly, Sylvain (2020): Investigating object compositionality in Generative Adversarial Networks. In: *Neural networks : the official journal of the International Neural Network Society* 130, S. 309–325. DOI: 10.1016/j.neunet.2020.07.007.

Vasella, David (2015): Zur Freiwilligkeit und zur Ausdrücklichkeit der Einwilligung im Datenschutzrecht. In: *Jusletter* (21).

Veale, Michael; Zuiderveen Borgesius, Frederik (2021): Demystifying the Draft EU Artificial Intelligence Act.

Velazco, Chris (2012): Google's 'Project Glass' Augmented Reality Glasses Are Real And In Testing. In: *TechCrunch*. Online verfügbar unter https://social.techcrunch.com/2012/04/04/google-project-glas/, zuletzt geprüft am 04.05.2021.

Venutti, Dario (2008a): Fans auf Schritt und Tritt überwachen. In: *Tages-Anzeiger*, S. 1.

Venutti, Dario (2008b): Fanüberwachung: Datenschützer kritisiert Pilotprojekt. In: *Tages-Anzeiger*, S. 3.

Verbraucherzentrale.de (2020): Gesundheits-Apps: medizinische Anwendungen auf Rezept. Online verfügbar unter https://www.verbraucherzentrale.de/wissen/gesundheit-pflege/aerzte-und-kliniken/gesundheitsapps-medizinische-anwendungen-auf-rezept-41241, zuletzt aktualisiert am 18.02.2022, zuletzt geprüft am 18.02.2022.

Villongco, Christopher; Khan, Fazal (2020): «Sorry I Didn't Hear You.» The Ethics of Voice Computing and AI in High Risk Mental Health Populations. In: *AJOB neuroscience* 11 (2), S. 105–112. DOI: 10.1080/21507740.2020.1740355.

Vincent, James (2017): Moscow says its new facial recognition CCTV has already led to six arrests. In: *The Verge*. Online verfügbar unter https://www.theverge.com/2017/9/28/16378164/moscow-facial-recognition-cctv-arrests-crime-surveillance, zuletzt geprüft am 15.06.2021.

Vincent, James (2019): Facial recognition smart glasses could make public surveillance discreet and ubiquitous. In: *The Verge*, 06.10.2019. Online verfügbar unter https://www.theverge.com/2019/6/10/18659660/facial-recognition-smart-glasses-sunglasses-surveillance-vuzix-nntc-uae, zuletzt geprüft am 30.03.2022.

Vincent, James (2020): Moscow rolls out live facial recognition system with an app to alert police. In: *The Verge*. Online verfügbar unter https://www.theverge.com/2020/1/30/21115119/moscow-live-facial-recognition-roll-out-ntechlab-deployment, zuletzt geprüft am 15.06.2021.

Viola, P.; Jones, M. (2001): Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001. 8 – 14 December 2001, Kauai, Hawaii, USA. Los Alamitos, Calif.: IEEE Computer Society, I-511-I-518.

Vogel, Jannis; Koßmann, Cosima; Schuir, Julian; Kleine, Nadine; Sievering, Jost (2020): Virtual- und Augmented-Reality-Definitionen im interdisziplinären Vergleich. In: Oliver Thomas und Ingmar Ickerott (Hg.): Smart Glasses: Augmented Reality zur Unterstützung von Logistikdienstleistungen. Berlin/Heidelberg: Springer, S. 19–50.

Vokinger, Kerstin Noëlle (2012): Das Berufsrecht in der Arzt-Patienten-Beziehung – veranschaulicht an einem Fallbeispiel.

Vokinger, Kerstin Noëlle (2020): Gesundheitsdaten im digitalen Zeitalter. In: *Jusletter*.

von Kaenel, Adrian (2006): Medizinische Untersuchungen und Tests im Arbeitsverhältnis. In: *ArbR*, S. 93–117.

Vuichard, Florence (2020): Reden statt tippen. In: *Bilanz* (4), S. 21. Online verfügbar unter https://www.unilu.ch/fileadmin/fakultaeten/wf/Dekanat/Dok/Professuren/Hofstetter/Bilanz_Sprachassistenten.pdf, zuletzt geprüft am 06.12.2021.

Wakefield, Jane (2021): AI emotion-detection software tested on Uyghurs. In: *BBC News*. Online verfügbar unter https://www.bbc.com/news/technology-57101248, zuletzt geprüft am 15.06.2021.

Wakefield, Jane (2020): PimEyes facial recognition website 'could be used by stalkers'. In: *BBC News*. Online verfügbar unter https://www.bbc.com/news/technology-53007510, zuletzt geprüft am 12.05.2021.

Waldmann, Bernhard (2015): Kommentar zu Art. 8 BV. In: Bernhard Waldmann, Eva Maria Belser und Astrid Epiney (Hg.): Schweizerische Bundesverfassung (BV), Basler Kommentar. Basel.

Walker, Shaun (2016): Face recognition app taking Russia by storm may bring end to public anonymity. In: *The Guardian*. Online verfügbar unter http://www.theguardian.com/technology/2016/may/17/findface-face-recognition-app-end-public-anonymity-vkontakte, zuletzt geprüft am 12.05.2021.

Wang, DeLiang; Chen, Jitong (2018): Supervised Speech Separation Based on Deep Learning: An Overview. In: *IEEE/ACM transactions on audio, speech, and language processing* 26 (10), S. 1702–1726. DOI: 10.1109/TASLP.2018.2842159.

Warman, Matt (2013): Google Glass: we'll all need etiquette lessons; What happens when we can all record everything, asks Matt Warman. In: *The Telegraph*, S. 2. Online verfüg-

bar unter https://www.telegraph.co.uk/technology/google/10015697/Google-Glass-well-all-need-etiquette-lessons.html, zuletzt geprüft am 04.05.2021.

Warren, Samuel D.; Brandeis, Louis (1890): The Right to Privacy. In: *Harvard Law Review* Vol. IV (5).

WBF, SBFI (Hg.) (2019): Herausforderungen der künstlichen Intelligenz: Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat. Online verfügbar unter https://www.sbfi.admin.ch/sbfi/de/home/bfi-politik/bfi-2021-2024/transversale-themen/digitalisierung-bfi/kuenstliche-intelligenz.html, zuletzt aktualisiert am 10.09.2021, zuletzt geprüft am 10.09.2021.

WBF, UVEK und Interdepartementale Arbeitsgruppe künstliche Intelligenz (Hg.) (2020): Leitlinien «Künstliche Intelligenz» für die Bundesverwaltung verabschiedet. Online verfügbar unter https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-81319.html, zuletzt geprüft am 10.09.2021.

Weber, Christian (2010): «Wie viel Verlogenheit hätten‘s denn gern?». In: *Süddeutsche Zeitung*, 12.05.2010. Online verfügbar unter https://www.sueddeutsche.de/wissen/neurophilosoph-metzinger-im-interview-totale-transparenz-ist-unertraeglich-1.1031809, zuletzt geprüft am 18.02.2022.

Webster, William R. (2004): The Diffusion, Regulation and Governance of Closed-Circuit Television in the UK. In: *SS* 2 (2/3). DOI: 10.24908/ss.v2i2/3.3376.

Wei, Wenqi; Wang, Jianzong; Cheng, Ning; Chen, Yuanxu; Zhou, Bao; Xiao, Jing (2020): Epidemic Guard: A COVID-19 Detection System for Elderly People. In: Xin Wang, Rui Zhang, Young-Koo Lee, Le Sun und Yang-Sae Moon (Hg.): Web and Big Data. 4th International Joint Conference, APWeb-WAIM 2020, Tianjin, China, September 18-20, 2020, Proceedings, Part II. Cham: Springer (Information Systems and Applications, incl. Internet/Web, and HCI), S. 545–550.

Weinmann, Benjamin (2019): Brisante Pläne: Scannt Migrolino bald die Gesichter der Kunden? In: *Aargauer Zeitung AG*, 24.07.2019. Online verfügbar unter https://www.aargauerzeitung.ch/wirtschaft/brisante-plaene-scannt-migrolino-bald-die-gesichter-der-kunden-ld.1137935, zuletzt geprüft am 29.04.2022.

Wernli, Reto (2020): Studie_MobilePayment_20210103_final. Online verfügbar unter https://blog.hslu.ch/retailbanking/files/2021/01/Mobile-Payment-Studie-2020.pdf, zuletzt geprüft am 16.06.2021.

White, Daniel (2017): Affect: An Introduction. In: *1* 32 (2), S. 175–180. DOI: 10.14506/ca32.2.01.

Wiggers, Kyle (2020): Researchers find evidence of bias in facial expression data sets. Hg. v. venturebeat. Online verfügbar unter https://venturebeat.com/2020/07/24/researchers-find-evidence-of-bias-in-facial-expression-data-sets/, zuletzt aktualisiert am 24.07.2020, zuletzt geprüft am 22.09.2021.

Wildhaber, Isabelle (2010): Genetische und medizinische Informationen in der Arbeitswelt. In: *Jusletter* (6. Dezember 2010). Online verfügbar unter https://jusletter.weblaw.ch/juslissues/2010/596/_8858.html__ONCE&login=false.

Willing, Richard (2003): Airport anti-terror systems flub tests; Face-recognition technology fails to flag ’suspects’. In: *USA Today,* zuletzt geprüft am 09.03.2021.

Wohlsen, Marcus (2014): Failure Is the Best Thing That Could Happen to Google Glass. In: *Wired*. Online verfügbar unter https://www.wired.com/2014/04/failure-is-the-best-thing-that-could-happen-to-google-glass/, zuletzt geprüft am 04.05.2021.

Wolfangel, Eva (2018a): Automatische Stimmanalysen übertreffen menschliche Experten. Heise Medien. Online verfügbar unter https://www.heise.de/newsticker/meldung/Emotionserkennung-fuer-Therapie-und-Marketing-4058882.html, zuletzt aktualisiert am 31.03.2021, zuletzt geprüft am 31.03.2021.

Wolfangel, Eva (2018b): Computer und große Gefühle: Wie Emotionserkennung gelingen kann. t3n magazin. Online verfügbar unter https://t3n.de/news/computer-grosse-gefuehle-953501/, zuletzt aktualisiert am 25.02.2018, zuletzt geprüft am 15.03.2021.

Wright, D.; Friedewald, M. (2013): Integrating privacy and ethical impact assessments. In: *Science and Public Policy* 40 (6), S. 755–766. DOI: 10.1093/scipol/sct083.

Wright, D.; Friedewald, M.; Gellert, R. (2015): Developing and testing a surveillance impact assessment methodology. In: *International Data Privacy Law* 5 (1), S. 40–53. DOI: 10.1093/idpl/ipu027.

Wright, David; Raab, Charles D. (2012): Constructing a surveillance impact assessment. In: *Computer Law & Security Review* 28 (6), S. 613–626. DOI: 10.1016/j.clsr.2012.09.003.

Wright, Steve (1998): An appraisal of technologies of political control. Scientific and Technological Options Assessment STOA. Working document. PE 166.499, 6 January 1998. Online verfügbar unter http://aei.pitt.edu/5538/, zuletzt geprüft am 09.03.2021.

Wu, Chao; Cao, Shuyang; Zhou, Fuchun; Wang, Chuanyue; Wu, Xihong; Li, Liang (2012a): Masking of speech in people with first-episode schizophrenia and people with chronic schizophrenia. In: *Schizophrenia research* 134 (1), S. 33–41. DOI: 10.1016/j.schres.2011.09.019.

Wu, Ting; Fu, Siyao; Yang, Guosheng (2012b): Survey of the Facial Expression Recognition Research. In: Huaguang Zhang (Hg.): Advances in brain inspired cognitive systems. 5th international conference, BICS 2012, Shenyang, China, July 11 – 14, 2012 ; proceedings. Berlin/Heidelberg: Springer (Lecture notes in computer science Lecture notes in artificial intelligence, 7366), S. 392–402.

Xia, Stephen; Jiang, Xiaofan (2020): PAMS: Improving Privacy in Audio-Based Mobile Systems. In: Proceedings of the 2nd International Workshop on Challenges in Artificial Intelligence and Machine Learning for Internet of Things. SenSys '20: The 18th ACM Conference on Embedded Networked Sensor Systems. Virtual Event Japan, 16 11 2020 – 19 11 2020. New York, NY, United States: Association for Computing Machinery (ACM Digital Library), S. 41–47.

Xu, Pengfei; Xie, Fei; Su, Tongsheng; Wan, Zhaoxin; Zhou, Zhaoyong; Xin, Xiaoyu; Guan, Ziyu (2020a): Automatic evaluation of facial nerve paralysis by dual-path LSTM with deep differentiated network. In: *Neurocomputing* 388, S. 70–77. DOI: 10.1016/j.neucom.2020.01.014.

Xu, Shihao; Fang, Jing; Hu, Xiping; Ngai, Edith; Guo, Yi; Leung, Victor C. M. et al. (2020b): Emotion Recognition From Gait Analyses: CurrentResearch and Future Directions. Online verfügbar unter https://arxiv.org/pdf/2003.11461.pdf.

Yang, D.; Alsadoon, Abeer; Prasad, P. W. C.; Singh, A. K.; Elchouemi, A. (2018): An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment. In: *Procedia Computer Science* 125, S. 2–10. DOI: 10.1016/j.procs.2017.12.003.

Yoffie, David B.; Wu, Liang; Sweitzer, Jodie; Eden, Denzil; Ahuja, Karan; Baldwin, Eric (2018): Voice War: Hey Google vs. Alexa vs. Siri, S. 25.

Yoo, Sanghyun; Song, Inchul; Bengio, Yoshua (2019): A Highly Adaptive Acoustic Model for Accurate Multi-dialect Speech Recognition. In: IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings: May 12–17, 2019, Brighton Conference Centre, Brighton, United Kingdom. ICASSP 2019. Piscataway, NJ: IEEE, S. 5716–5720.

Yoon, Seunghyun; Byun, Seokhyun; Jung, Kyomin (2018): Multimodal Speech Emotion Recognition Using Audio and Text.

You, Yue; Gui, Xinning (2020): Self-Diagnosis through AI-enabled Chatbot-based Symptom Checkers: User Experiences and Design Considerations. In: The AMIA 2020 Annual Symposium.

Yujie, Xue (2019): Camera Above the Classroom. Hg. v. Sixth Tone. Peking. Online verfügbar unter https://www.sixthtone.com/news/1003759/camera-above-the-classroom, zuletzt geprüft am 08.04.2021.

Zhang, Nan; Mi, Xianghang; Feng, Xuan; Wang, XiaoFeng; Tian, Yuan; Qian, Feng (2019): Dangerous Skills: Understanding and Mitigating Security Risks of Voice-Controlled Third-Party Functions on Virtual Personal Assistant Systems. In: IEEE Symposium on Security and Privacy. SP 2019: San Francisco, California, USA, 19–23 May 2019: proceedings. Piscataway, NJ: IEEE, S. 1381–1396.

Zhang, Zhan; Song, Yufei; Cui, Liqing; Liu, Xiaoqian; Zhu, Tingshao (2016): Emotion recognition based on customized smart bracelet with built-in accelerometer. In: *PeerJ* 4, e2258. DOI: 10.7717/peerj.2258.

Zhao, W.; Chellappa, R.; Phillips, P. J.; Rosenfeld, A. (2003): Face recognition. In: *ACM Comput. Surv.* 35 (4), S. 399–458. DOI: 10.1145/954339.954342.

Zick, Andreas (2014): Bambule und Randale – Gewalt im Fussball: Im Abseits? In: *Bundeszentrale für politische Bildung*, zuletzt geprüft am 15.06.2021.

Zloteanu, Mircea; Krumhuber, Eva G.; Richardson, Daniel C. (2018): Detecting Genuine and Deliberate Displays of Surprise in Static and Dynamic Faces. In: *Frontiers in psychology* 9, S. 1184. DOI: 10.3389/fpsyg.2018.01184.

Zuboff, Shoshana (2018): Das Zeitalter des Überwachungskapitalismus. Frankfurt a. M.: Campus Verlag.

Zuchowski, Matthias; Pashayeva, Aydan; Wohlrab, Martin (2020): Medizinische Spracherkennung im stationären und ambulanten Einsatz – Eine systematische Übersicht. In: *Gesundh ökon Qual manag* 25 (02), S. 83–90. DOI: 10.1055/a-1115-6980.

Anhang

Tabelle 19: Übereinstimmungen zwischen zentralen Schweizer Policy-Dokumenten zu Leitlinien und Zielen hins. KI (eigene Zusammenstellung)

Übereinstimmende Prinzipien	**Herausforderungen der künstlichen Intelligenz**	**Leitlinien für den Umgang mit den Herausforderungen der künstlichen Intelligenz**	**Ziele aus der Strategie Digitale Schweiz**
Nachvollziehbarkeit und Transparenz	Nachvollziehbarkeit und Transparenz	Transparenz, Nachvollziehbarkeit und Erklärbarkeit	Sicherheit, Vertrauen und Transparenz gewährleisten
Fairness, Bias und Nicht-Diskriminierung	Bias und Diskriminierung	Den Menschen in den Mittelpunkt stellen	Chancengleiche Teilhabe aller ermöglichen und Solidarität stärken
Verantwortlichkeit, Rechenschaftspflicht	Autonomie, Verantwortlichkeit und Haftung	Verantwortlichkeit	
Datenschutz/ Privatheit	Datenzugang und Datenschutz		Sicherheit, Vertrauen und Transparenz gewährleisten
Sicherheit		Sicherheit	Sicherheit, Vertrauen und Transparenz gewährleisten
		Rahmenbedingungen für Entwicklung und Anwendung von KI	
		Aktive Mitgestaltung der Governance von KI	
		Einbezug aller relevanten nationalen und internationalen Akteure	
			Digital Befähigung und Selbstbestimmung der Menschen weiter stärken

Tabelle 20: Ethik-Kriterien-Fragekatalog

Privatheit, Datenschutz

- Ist der Zweck der Datenbearbeitung rechtmässig?
- Wird von den Betroffenen eine Einwilligung eingeholt?
- Bestehen Auskunftsrechte und können sie effektiv wahrgenommen werden?
- Können Betroffene die sie betreffenden personenbezogenen Daten löschen oder korrigieren lassen?
- Können Betroffene einer Bearbeitung widersprechen?

Verantwortlichkeit, Rechenschaftspflicht

- Wie werden die Daten gespeichert? (eine zentrale Datenbank auf einem Server; eine zentrale verteilte Datenbank auf mehreren Servern; mehrere einzelne Datenbanken auf verschiedenen Servern bei verschiedenen Behörden, Unternehmen usw.)
- Gibt es innerhalb der für die Anwendung verantwortlichen Institution eine klare Verantwortungsverteilung?
  - Ist eine natürliche oder juristische Person als Verantwortungsadressat bestimmt?
  - Sind ggf. die Verantwortlichkeiten zwischen verschiedenen beteiligten Institutionen klar geregelt?
  - Zugriffsrechtemanagement: Wer hat innerhalb der Institution Zugang zu den Daten?
- Ex-ante-Kontrolle
  - Wurde vor der Inbetriebnahme des Dienstes ein Impact Assessment (z.B. eine DSFA) durchgeführt?
- Ex-post-Kontrolle
  - Wird der Betrieb nach Inbetriebnahme laufend von unabhängiger Seite evaluiert?
  - Wie sind die Kosten des Systems (Hardware, Software als auch laufender Betrieb)?
- Betroffenenperspektive
  - Könne Betroffene gegen ein Analyseergebnis bzw. auf Entscheidungen oder Handlungen, die darauf basieren, Widerspruch einlegen?
  - Wie einfach ist dies möglich?
  - Wie transparent wird diese Möglichkeit kommuniziert?
  - Gibt es Beschwerdemöglichkeiten? Wer nimmt die Beschwerde entgegen?
  - Sind Haftungsfragen (insb. die Frage, welche Stelle für entstandenen Schaden aufkommen muss und wohin sich Betroffene wenden können) klar geregelt?

Sicherheit

- Wird die Sicherheit und Belastbarkeit des Systems von unabhängiger Stelle evaluiert?
- Werden (kritische/ Sicherheits-)Updates von Herstellern angeboten?
- Werden diese Updates auf den Systemen der Institutionen installiert?
- Ist die Vertraulichkeit des Systems gewährleistet? Hat nur autorisiertes Personal Zugang zu den Analyseergebnissen und/oder Rohdaten?
- Ist die Integrität des Systems gewährleistet?
- Ist die Verfügbarkeit des Systems gewährleistet?
- Sind die Systeme sicher gegen absichtliche Sabotage oder Workarounds? (Resilienz)
- Ist eine Überprüfbarkeit der Echtheit der Ergebnisse/Datenobjekte möglich? (Authentizität)
- Werden Zugriffe auf das System für jeden Nutzer einzeln geloggt? Sind diese Zugriffe zuordenbar? (Zurechenbarkeit)
- Können Nutzer den Zugriff auf das System abstreiten? Werden unlöschbare Logfiles erstellt? (Verbindlichkeit)

Transparenz, Erklärbarkeit

- Transparenz des Algorithmus und der genutzten (Trainings-)Daten
  - Ist der Ursprung der Trainingsdaten bekannt?
  - Sind die Inhalte der Trainingsdaten bekannt?
  - Ist der verwendete Algorithmus bekannt?
  - Ist der verwendete Algorithmus frei verfügbar / open source?
  - Gibt es Prozeduren zur unabhängigen Begutachtung des genutzten Algorithmus?
  - Welche Zuverlässigkeit wird dem Algorithmus in unabhängigen Tests (z.B. NIST) bescheinigt?
- Transparenz der Anwendung
  - Wird der Herstellername der eingesetzten Software angegeben?
  - Gibt es Prozeduren zur unabhängigen Begutachtung des verwendeten Algorithmus im konkreten Anwendungsfall?
  - Welche Zuverlässigkeit wird dem Algorithmus in unabhängigen Tests für den konkreten Anwendungsfall bescheinigt?
  - Woher stammen die gespeicherten Stimm-, Sprach- und Gesichtsdaten?
  - Werden Betroffene über den Einsatz der Stimm-, Sprach- bzw. Gesichtserkennung auf verständliche Weise informiert?
  - Werden Daten per Opt-In oder per Opt-Out erhoben?
  - Können Betroffene Informationen darüber erhalten, welche Personen(gruppen) Zugriff auf die gespeicherten Daten haben?
  - Wird über den Einsatz der Stimm-, Sprach- bzw. Gesichtserkennung regelmässig öffentlich Bericht erstattet? (z.B. auch darüber, wie viele Stellen auf Daten zugegriffen haben)

Gerechtigkeit, Fairness, Nicht-Diskriminierung

- Dient das System dem Zweck der Authentifikation, Identifikation oder Verfolgung (Tracking)?
- Ist das System echtzeit- oder ex-post-basiert?
- Was ist der erwartete Nutzen des Systems?
  - Welchen Interessen ist dieser Nutzen zuträglich (privat/öffentlich, Staat, Zivilgesellschaft, einzelne Bürger)?
  - Welche Auswirkungen hat das System auf die übrigen Interessengruppen?
- Welche Alternativen zur Stimm-, Sprach- bzw. Gesichtserkennung existieren? Was sind die Vor- und Nachteile der Alternativen?
- Bias-Vermeidung
  - Wurden die Trainingsdaten im Hinblick auf ein potenzielles Bias untersucht?
  - Wurde der Algorithmus im Hinblick auf ein potenzielles Bias untersucht?
  - Wurden die Inputdaten im Hinblick auf Qualität untersucht und bewertet? (z.B. Lichtempfindlichkeit der Kamera, die zu Verfälschungen führen kann)
  - Wird potenzieller Bias transparent kommuniziert?
- Sind Mechanismen zur Evaluation der Daten und Betriebsprozesse in Kraft?
- Wurde bei der Gestaltung der Anwendung auf partizipative Elemente gesetzt? (etwa die Beteiligung von Betroffenen oder zivilgesellschaftlichen Akteuren)

Menschliche Kontrolle der Technik

- Werden die Ergebnisse der Software von einem Menschen kontrolliert und interpretiert?
- Werden Menschen hins. der korrekten Kontrolle und Interpretation von Ergebnissen geschult?

Tabelle 21: Übersicht der am häufigsten genannten Vorteile der diskutierten Stimm-, Sprach- und Gesichtserkennungstechnologien aus Sicht der Fokusgruppen-Teilnehmenden

Fokusgruppen-Fall	Steigerung der (gefühlten) persönlichen Sicherheit	Steigerung des Komforts	Vereinfachte Aufklärung bei Ermittlungen	Verbesserte (Früh)Erkennung von Krankheiten
Stadionüberwachung (DE)			X	
Erkennung psychischer Krankheiten (DE)				X
Smarte Lautsprecher (FR)	X	X		
Polizeil. Überwachung (FR)	X	X	X	
Smarte Lautsprecher (DE)	X	X		
Jedermann-Identifikation (DE)				
Polizeil. Überwachung (DE)			X	
Authentifizierung via Stimme (DE)	X	X		
Emotionserkennung Aufmerksamkeitsanalyse (FR)	X			
Erkennung physischer Krankheiten (FR)				X
Häufigkeit der Nennung	5 Mal	4 Mal	3 Mal	2 Mal

Tabelle 22: Übersicht der am häufigsten genannten Nachteile der diskutierten Stimm-, Sprach- und Gesichtserkennungstechnologien aus Sicht der Fokusgruppen-Teilnehmenden

Fokusgruppen-Fall	Fehlende Transparenz (inkl. Weitergabe und Zweckentfremdung)	Wahrnehmung als ungerechtfertigter Privatheitseingriff	Unzuverlässigkeit von Softwareergebnissen	Furcht vor Diskriminierung und weiteren sozialen Folgen	Angst vor Manipulation	Bedenken hins. der Datensicherheit	Angst vor Technologieabhängigkeit (z.B. social deskilling)
Stadionüberwachung (DE)	X	X	X	X		X	
Erkennung psychischer Krankheiten (DE)	X		X	X			
Smarte Lautsprecher (FR)	X	X	X	X	X		X
Polizeil. Überwachung (FR)		X	X	X			
Smarte Lautsprecher (DE)	X	X		X	X	X	X
Jedermann-Identifikation (DE)	X	X					
Polizeil. Überwachung (DE)	X	X					
Authentifizierung via Stimme (DE)	X	X	X			X	
Emotionserkennung Aufmerksamkeitsanalyse (FR)	X	X	X	X	X		
Erkennung physischer Krankheiten (FR)	X		X	X			
Häufigkeit der Nennung	**9**	**8**	**7**	**7**	**3**	**3**	**2**

Tabelle 23: Detaillierte Übersicht aller Fokusgruppen-Empfehlungen

Empfehlungskategorie und konkrete Empfehlung	Häufigkeit der Nennung
Gewährleistung von Transparenz	
• Transparenz über den jeweiligen Einsatz	4
• Transparenz über die Art, Menge, Speicherung und Auswertung der erhobenen Daten	4
• Transparenz über die Weitergabe von Daten an Dritte und über eine mögliche Sekundärnutzung	2
• Transparenz im Hinblick auf die Nachvollziehbarkeit von Softwareergebnissen	2

Empfehlungskategorie und konkrete Empfehlung	Häufigkeit der Nennung
Erwartungen an die Politik:	
• Spezifische gesetzliche Vorschriften (inkl. klarer Zweckbestimmung usw.) bzw. spezifischere Bestimmung der Einsatzzwecke	5
• Gesetzliches Verbot spezifischer, besonders problematischer Zwecke (z.B. politische Werbung bei sm. LS FR, von (Gruppen-)Profiling mittels anonymisierter Daten bei sm. LS DE). Z.B. auch mittels Zahlung einer Gebühr für die Dienstenutzung (smarte Lautsprecher DE), polizeilicher Gesichtserkennung und des Tragens smarter Brillen in der Öffentlichkeit mit Erlaubnisvorbehalt	4
• Gewährleistung der techn. Zuverlässigkeit durch Regulierung, insb. Zertifizierung	3
• Vorantreiben von Aufklärungskampagnen über Technologierisiken	3
• Sicherstellung der Kontrolle durch Dritte (unabhängige Institutionen oder staatliche Stellen), dass die Technologie die datenschutzrechtlichen Anforderungen auch einhält (im Falle der pol. GE unabhängige Kommission des Parlaments)	2
• Sicherstellung eines geregelten Vertriebs von Diensten zur Erkennung von Krankheiten, um Missbrauch im persönlichen Bereich auszuschliessen	1
• Gewährleistung der Betroffenenrechte hins. Herausgabe, Korrektur, Löschung	1
• Vorantreiben von Aufklärungskampagnen über den Technologiegebrauch	1
• Vorantreiben grenzüberschreitender Regulierung zur Adressierung grenzüberschreitender Risiken	1
• Durchsetzung geltender Gesetze und hohe Strafen bei Verstössen	1
• Staatliche Unterstützung bei der Entwicklung datenschutzkonformer Konkurrenzprodukte	1
• Garantierung technischer Grenzen für den Datenzugriff (Verhinderung von Missbrauch)	1
• Klarstellung von Haftbarkeit im Schadensfall	1
Erwartungen an Technologiehersteller und -betreiber:	
• Schulung des Personals, das die Technologie einsetzt	3
• Förderung alternativer Methoden anstelle oder parallel zum Einsatz von Stimm-, Sprach- und Gesichtserkennungstechnologien	2
• Gewährleistung der Datensicherheit	2
• Datenschutzfreundliches Technikdesign (so wenige Daten wie möglich erheben)	2
• Bearbeitung der Daten auf den Geräten bzw. in der Schweiz oder Europa / Keine Übertragung in die Cloud	2

Empfehlungskategorie und konkrete Empfehlung	Häufigkeit der Nennung
• Bereinigung der Sprachbefehle um prosodische Merkmale (Stimmhöhe, -geschwindigkeit etc.), um die Daten zu entemotionalisieren und zu entpersonalisieren, bevor die Daten zur Ausführung des Befehls gesendet werden	1
• Schaffung offener Schnittstellen bzw. Trennung von Hardware und Software	1
• Opt-In bevorzugt	1
• Kein alleiniger Verlass auf die Technologie (Begleitung des Einsatzes von poliz. Gesichtserkennung von Personen aus der Sozialhilfe)	1
• Gesetzeskonforme Befüllung der Datenbank	1
• Entwurf einer berufsethischen Konvention als Handlungsanleitung (technologische Tools als Unterstützung der ärztlichen Diagnose, nicht als Ersatz)	1
Gesellschaftliche Handlungsmöglichkeiten	
• Aussenden von Störsignalen zur Beeinträchtigung der Gesichtserkennungsfunktion	1
• Bekleben des eigenen Gesichts mit Störmustern zur Beeinträchtigung der Gesichtserkennungsfunktion	1
• Kritische Reflexion des Technologieeinsatzes	1

Autorinnen und Autoren

Murat Karaboga, Dr. des. (Studienleiter): Politikwissenschaftler und Projektleiter im Geschäftsfeld Informations- und Kommunikationstechnologien (IKT) am Competence Center (CC) Neue Technologien des Fraunhofer ISI.

Nula Frei, Dr. iur. (Leitung des Projektteams der Universität Freiburg): Oberassistentin am Institut für Europarecht der Universität Freiburg i.Ue. und Lehrbeauftragte am Global Studies Institute der Universität Genf.

Frank Ebbers, M.Sc.: Wirtschaftsinformatiker und Doktorand im Geschäftsfeld IKT am CC Neue Technologien am Fraunhofer ISI.

Sophia Rovelli, Mlaw: Diplomassistentin und Doktorandin am Lehrstuhl für Europa-, Völker- und öffentliches Recht an der Universität Freiburg i.Ue.

Michael Friedewald, Dr. Ing. (Co-Studienleiter): Ingenieur der Elektrotechnik und Informationstechnik sowie Wirtschaftswissenschaftler. Koordinator des Geschäftsfelds IKT und Senior-Researcher im CC Neue Technologie am Fraunhofer ISI.

Greta Runge, M.A.: Politikwissenschaftlerin und Doktorandin im Geschäftsfeld IKT am CC Neue Technologien am Fraunhofer ISI.

Mitglieder der Begleitgruppe

Dr. Bruno Baeriswyl (Leiter der Begleitgruppe), Datenschutzexperte, Mitglied des Leitungsausschusses von TA-SWISS

Dominik Brumm, Head of Development Cubera

Prof. Dr. Volker Dellwo, Institut für Computerlinguistik, Universität Zürich

Dr. Jean Hennebert, Université de Fribourg, Mitglied des Leitungsausschusses von TA-SWISS

Dr. Anna Jobin, Soziologin

Prof. Dr. Annett Laube, Technik und Informatik, Berner Fachhochschule

Prof. Dr. Klaus Scherer, Swiss Center for Affective Sciences, Universität Genf

Remo Schmidlin, Jurist, Lenz & Staehelin

Prof. Dr. Thomas Vetter, Departement Mathematik und Informatik, Universität Basel

Patrick Walder, Amnesty International Schweiz

Projektmanagement TA-SWISS

Dr. rer. soc. Elisabeth Ehrensperger, Geschäftsführerin

Dr. Christina Tobler, Projektleiterin (2020–2021)

Dr. Laetitia Ramelet, Projektleiterin (2022)

www.ta-swiss.ch

Ein Kompetenzzentrum der
akademien der wissenschaften schweiz

Technologien zur Stimm-, Sprach- und Gesichtserkennung werden im Alltag bereits angewendet. Etwa in smarten Lautsprechern, die auf Kommando die gewünschte Musik abspielen oder Nutzerfragen beantworten, durch Strafverfolger zur Suche nach dem Gesicht von Verdächtigen in Videomaterial oder bei Bankkunden, die am Telefon anhand ihrer Stimme identifiziert werden.

Diese Technologien versprechen, den Alltag ihrer Nutzerinnen und Nutzer zu vereinfachen, ihnen frühzeitige Hinweise auf Krankheiten zu geben und der Polizei neue Möglichkeiten bei der Verbrechensbekämpfung zu eröffnen. Dabei ist gar nicht sicher, wie zuverlässig sie funktionieren und ob sie mit geltendem Recht in Einklang stehen. Hinzu kommen umstrittene Nutzungen, etwa die Möglichkeit, auf den Gesundheitszustand, Emotionen und Gewohnheiten einer Person zu schliessen.

In dieser Studie werden zahlreiche Anwendungen aus technischer, rechtlicher und ethischer Sicht untersucht und daraus Handlungsempfehlungen abgeleitet. Wie kann die Technologie verantwortungsbewusst eingesetzt werden, und wo wäre ein Verbot sinnvoll? Die Diskussion über Stimm-, Sprach- und Gesichtserkennung ist in vollem Gang, die Studie bietet dazu fundierte Orientierung.

v/d|f

TA-SWISS 79/2022
ISBN 978-3-7281-4140-8 (Printversion)
ISBN 978-3-7281-4141-5 (E-Book)
DOI-Nr. 10.3281/4141-5