



Auf Schritt und Tritt beobachtet

Kurzfassung der Studie von TA-SWISS «Automatisierte Erkennung von Stimme, Sprache und Gesicht: Technische, rechtliche und gesellschaftliche Herausforderungen»



Die Stiftung TA-SWISS, ein Kompetenzzentrum der Akademien der Wissenschaften Schweiz, setzt sich mit den Chancen und Risiken neuer Technologien auseinander.

Die hier vorliegende Kurzfassung basiert auf einer wissenschaftlichen Studie, die im Auftrag von TA-SWISS von einem interdisziplinären Projektteam vom Fraunhofer-Institut für System- und Innovationsforschung ISI in Karlsruhe und vom Institut für Europarecht der Universität Freiburg (Schweiz) unter der Leitung von Dr. Murat Karaboga durchgeführt wurde. Die Kurzfassung stellt die wichtigsten Resultate und Schlussfolgerungen in verdichteter Form dar und richtet sich an ein breites Publikum.

Automatisierte Erkennung von Stimme, Sprache und Gesicht: Technische, rechtliche und gesellschaftliche Herausforderungen

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

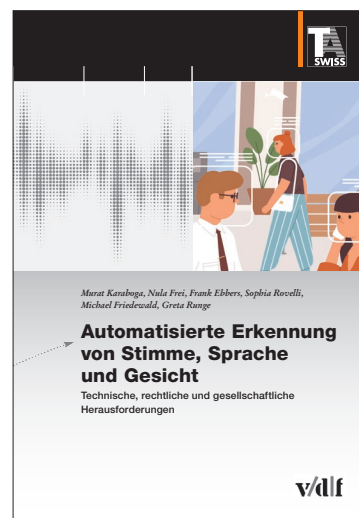
TA-SWISS, Stiftung für Technologiefolgen-Abschätzung (Hrsg.).

vdf Hochschulverlag an der ETH Zürich, 2022.

ISBN: 978-3-7281-4140-8

Die Studie steht als eBook zum freien Download bereit: www.vdf.ch

Die vorliegende Kurzfassung ist ebenfalls online verfügbar: www.ta-swiss.ch



Stimm-, Sprach- und Gesichtserkennung in aller Kürze	4
Einige Chancen ...	4
... und Risiken	4
Vordringliche Empfehlungen	4
Schwindende Anonymität	5
Hilfreiche Bürodrachen und andere Dienste	5
Das Gesicht aus Datenpunkten	6
Identifizieren und Verifizieren, Sprache und Sprechen	6
Biometrische Daten verraten viel mehr als ein blosses Abbild	7
Datenschutz mit weiterem Blickwinkel	7
Konformitätsdruck gefährdet Grundrechte und Demokratie	8
Wenn die Technik die Ohren spitzt	8
Smarte Lautsprecher: Es hören viele mit	8
Authentifizierung über die Sprache in der automatisierten Schalterhalle	9
Kriminelle abwehren	10
Für den Einlass Gesicht zeigen	11
Verhaltener Einsatz in der Schweiz	11
Gesichtserkennung in Echtzeit als Startschuss zur Massenüberwachung?	11
Optik vereint mit Akustik im Kampf gegen Rassismus	12
Wenn die Technik das Innerste nach aussen kehrt	13
Im Einsatz für die Medizin	13
Die Gefühlswelt ausloten	14
Keine Ablenkung in der Schule zugelassen	15
Nicht alle wollen mit allen bekannt sein	15
Big Brother soll draussen bleiben: Empfehlungen	17
Hoch riskante Anwendungen verbieten	17
Aufschub für die Emotions- und Krankheitserkennung	17
Gesetzliche Lücken füllen, Weiterbildung fördern, Betroffene unterstützen	18

Stimm-, Sprach- und Gesichtserkennung in aller Kürze

Die vielerorts im öffentlichen Raum aufgestellten Kameras liefern reichhaltiges Material, um die Programme der Stimm-, Sprach- und Gesichtserkennung zu alimentieren. Diese Technologie kann dazu beitragen, die individuelle wie auch die öffentliche Sicherheit zu erhöhen, indem sie den Behörden hilft, vermisste Personen zu finden oder Verdächtige zu beobachten. Allerdings birgt diese automatische Identifikation auch Missbrauchspotenzial und stellt die Menschen unter Druck, ihr Verhalten der sozialen Norm anzupassen. Dies hat nicht nur Folgen für die Individuen, sondern auch für die Demokratie.

Smarte Lautsprecher sind mittlerweile in etlichen Schweizer Haushalten anzutreffen, und wer sein Smartphone entsperren will, tut dies oft mit der Face-ID. Die Stimm-, Sprach- und Gesichtserkennung vereinfacht viele alltägliche Handlungen. Noch ist sie zwar nicht immer verlässlich: So identifiziert die Gesichtserkennung Frauen und Menschen mit dunklerer Hautfarbe weniger präzise als weisse Männer. Doch aufgrund der gegenwärtigen Entwicklungen ist absehbar, dass die technische Zuverlässigkeit in den nächsten Jahren einen hohen Stand erreichen wird. Grösser ist die Gefahr von falschen Ergebnissen, wenn Stimm- und Gesichtsdaten verwendet werden, um Rückschlüsse auf Emotionen oder den körperlichen und geistigen Zustand einer Person zu ziehen.

Einige Chancen ...

Dank smarter Lautsprecher lassen sich zahlreiche Geräte über Sprachbefehle bedienen, sodass die Vielzahl an Fernbedienungen überflüssig wird. Die smarten Assistenten vereinfachen das tägliche Leben, indem sie sprachgesteuert Einträge im Kalender speichern oder Stereoanlage und Beleuchtung regeln – während die Nutzerin, der Nutzer beide Hände freihalten kann.

In der Medizin könnten Stimm-, Sprach- und Gesichtserkennung dazu beitragen, frühzeitig schwere Leiden wie Parkinson, Alzheimer, Depressionen und Burn-outs aufzudecken. Auch wird an Programmen gearbeitet, um mittels Gesichtserkennung seltene Krankheiten zu identifizieren, die nur wenige ärztliche Fachkräfte je zu sehen bekommen.

Kriminelle Vorhaben lassen sich unter Umständen rechtzeitig aufdecken, wenn Verdächtige aufgespürt und in ihrem Handeln verfolgt werden können. Die Stimm-, Sprach- und Gesichtserkennung kann somit die öffentliche Sicherheit erhöhen.

... und Risiken

Stimm-, Sprach- und Gesichtsdaten sind biometrische Daten und verraten viel über eine Person. Sie verändern sich während des Erwachsenenlebens kaum mehr. Sind sie erst einmal gehackt, bleiben sie dauerhaft kompromittiert.

Der Einsatz der Stimm-, Sprach- und Gesichtserkennung gefährdet die Privatsphäre, weil sie die Menschen aus ihrer Anonymität holt. Smarte Lautsprecher stehen im Wohnzimmer und damit im Mittelpunkt des Haushaltes. Sie sind ständig auf Empfang und zeichnen unter Umständen vertrauliche Informationen auf.

Stimm-, Sprach- und Gesichtserkennung akzentuiert tendenziell das Machtgefälle zwischen staatlichen Stellen bzw. privatwirtschaftlichen Anwendern und den Bürgerinnen und Bürgern – insbesondere, wenn letztere nicht wissen, über welche Daten die Behörden oder privatwirtschaftliche Akteure über sie verfügen.

Vordringliche Empfehlungen

Besonders heikle Anwendungen der Stimm-, Sprach- und Gesichtserkennung sind zu verbieten. Dazu gehören die automatische Echtzeit-Überwachung und die Nutzung von Datenbrillen und anderen unauffälligen Geräten, die es gestatten, Menschen heimlich zu beschatten. Auch die Aufmerksamkeitsanalyse an Schulen sollte nicht zugelassen werden.

Die Verwendung der Stimm-, Sprach- und Gesichtserkennung durch die Polizei oder andere Behörden bedarf ausdrücklicher Gesetze, die sowohl rechtsstaatliche Sicherungsmechanismen definieren als auch festlegen, dass die Notwendigkeit des Technologieeinsatzes überprüft wird.

Es sollten Aus- und Weiterbildungsangebote für all jene geschaffen werden, welche die Technologie nutzen. Auch braucht es Anlaufstellen für Personen, die sich vor den Nachteilen aus dem Einsatz von Stimm-, Sprach- und Gesichtserkennung schützen und ihre Rechte kennen und durchsetzen möchten.

Die Vor- und Nachteile der Stimm-, Sprach- und Gesichtserkennung und ihr zulässiger Einsatz müssen in der Gesellschaft intensiv erörtert werden.

Die Studie zur Stimm-, Sprach- und Gesichtserkennung wurde von einer Projektgruppe aus dem Fraunhofer-Institut für System- und Innovationsforschung ISI und der Universität Freiburg i.Ue. unter der Gesamtleitung von Murat Karaboga durchgeführt. Methodisch stützt sie sich auf ausführliche Literaturstudien und auf die Auswertung von Medienberichten, die durch mehrere Diskussionen in Fokusgruppen mit Bürgerinnen und Bürgern und eine repräsentative online-Umfrage bei 1000 Personen ergänzt wurden.

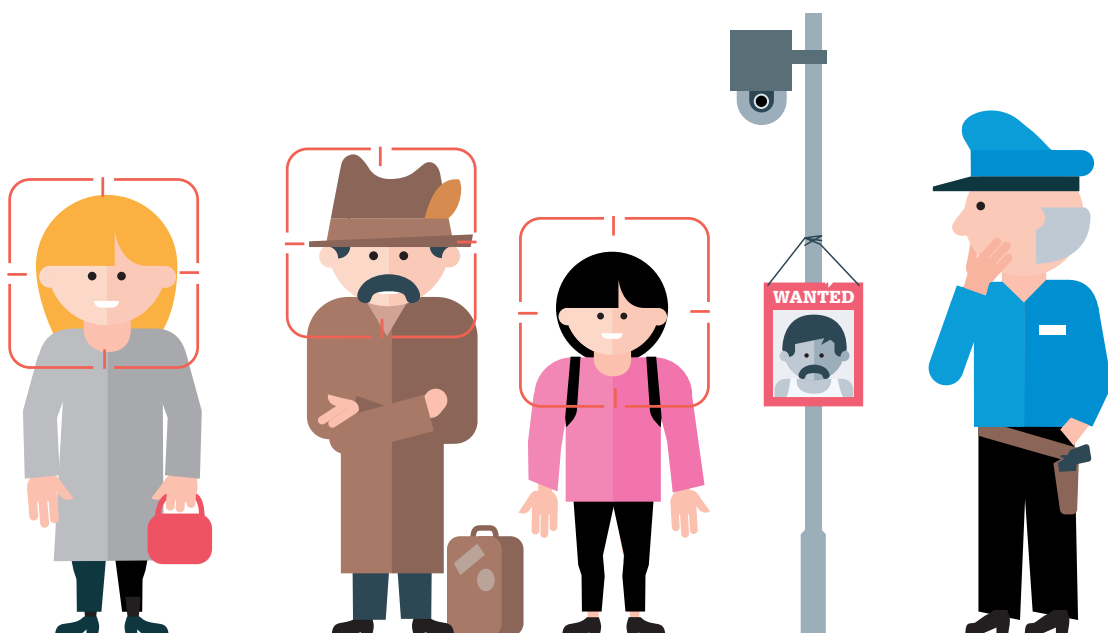
Schwindende Anonymität

Das schwarze Wunderauto K.I.T.T. aus der amerikanischen Krimi-Serie Knight Rider befolgte bereits in den frühen 1980er-Jahre gesprochene Befehle. Und schon in frühen Science-Fiction-Filmen brauchte die Protagonistin bloss ihr Gesichtsbild scannen zu lassen, um Zugang zum Cockpit des Raumschiffs zu erhalten. Mittlerweile ist die Stimm-, Sprach- und Gesichtserkennung auch in unserem Alltag angekommen.

Oft beeinflusst die Technik gesellschaftliche Entwicklungen auf unerwartete Weise. So gestattete es im Juni 2022 das Verwaltungsgericht Göttingen einem

Elternpaar, den Namen ihrer vierjährigen Tochter Alexa zu ändern. Denn die Kleine war gehänselt worden und galt schliesslich als seelisch belastet. Alexa heisst nämlich auch die Software von Amazon, die es ermöglicht, Applikationen über die Sprache zu steuern. Das Mädchen aus Niedersachsen musste sich entsprechend dämliche Befehle und ärgerliche Wortwitze anhören.

Ob der Klang, die Familientradition oder die Mode den Ausschlag gibt: Werdende Eltern machen sich viele Gedanken, um für ihr Baby den passenden Namen zu finden. Immerhin kann man diesen



wechseln, wenn es die Lebensumstände erfordern. Anders verhält es sich mit der Stimme, der Form des Gesichts oder auch der Art zu sprechen. Diese Eigenschaften verändern sich bei Erwachsenen kaum mehr. Sind diese mit einem Menschen untrennbar verbundenen Charakteristika vermessen, digitalisiert und entsprechend als biometrische Daten abgespeichert, werden wesentliche Persönlichkeitsmerkmale für eine weitere elektronische Aufbereitung verfügbar.

Stimm-, Sprach- und Gesichtserkennung blicken auf eine gut fünfzigjährige Geschichte zurück. Mögen die technischen Details auch unterschiedlich sein, so ist ihnen doch gemeinsam, dass sie auf der Verarbeitung biometrischer Daten beruhen, die viel über eine Person verraten.

Hilfreiche Bürodrachen und andere Dienste

Bereits in den frühen 1960er-Jahren beschäftigte sich die Forschung mit der maschinellen Verarbeitung von Sprache, wenngleich mit wenig Erfolg. Zu gering war damals die Leistung der Computer. Erst dank modernen Grossrechnern verbesserte sich die Performance, sodass IBM, Philips und Dragon Systems mit der Zeit kommerzielle Diktierprogramme entwickeln konnten, die sich ab den 1990er-Jahren in den Büros etablierten.

Als Diktierprogramm, das unabhängig von einem bestimmten Betriebssystem funktioniert, setzte sich Dragon durch. Nach einem kurzen Training, welches die Software mit der Sprecherin oder dem Sprecher vertraut macht, erreicht sie eine Erkennungsrate von bis zu 98 Prozent. Bei Diktaten in einer Fachsprache mit beschränktem Wortschatz arbeitet Dragon präziser als bei solchen mit literarischen Wendungen und reichem Vokabular.

An bestimmte Plattformen gebunden sind Siri, Google Assistant und Alexa, die in die Betriebssysteme von Apple und Android bzw. in die Lautsprecher von Amazon integriert wurden. Insbesondere die preiswerten Amazon-Lautsprecher verhalfen der Nutzung von Stimm- und Spracherkennung zum Durchbruch: 2017 stammten etwa 80 Prozent aller weltweit verkauften smarten Lautsprecher von Amazon. Freilich wurde seither das Angebot vielfältiger.

Auf Befehle wie «Alexa, stell den Wecker auf sieben Uhr» oder «Hey Siri, zeige mir die Wetterprognose für Zürich» reagiert das Smartphone mit der

gewünschten Aktion. Ist das Handy mit anderen Geräten im smarten Zuhause verbunden, können auch diese mittels Sprachbefehlen angesteuert werden: «Hey Siri, aktiviere den Drucker» oder «Alexa, schalte die Kaffeemaschine ein», funktioniert dann ebenfalls.

Das Gesicht aus Datenpunkten

Auch die Bemühungen um die Gesichtserkennung reichen bis in die 1960er-Jahre zurück. Greifbare Erfolge stellten sich allerdings später ein als bei der Spracherkennung. Eine erste schematische Kartographierung von Haaransatz, Augen und Nase, wie sie 1964 versucht wurde, ergab bereits bei geneigtem Kopf und ungünstigen Lichtverhältnissen falsche Ergebnisse. Diese steigerten sich erst ab 1970, als es die erhöhte Leistungsfähigkeit der Computer erlaubte, weitere Merkmale wie die Form der Lippen und die Haarfarbe einzubeziehen. Der Durchbruch gelang in den frühen 1990er-Jahren mit einem Algorithmus, der nicht mehr auf der Vermessung anatomischer Merkmale einzelner Gesichter beruht, sondern die Hauptkomponenten von Daten aus einer grossen Menge an Gesichtsbildern statistisch auswertet. Das Problem bei diesem Ansatz: Die Erkennungsraten in verschiedenen Forschungsgruppen liessen sich zunächst kaum miteinander vergleichen, da jede mit eigenen Bilddatenbanken arbeitete. Das verbesserte sich erst, nachdem das US-amerikanische Verteidigungsministerium im Rahmen eines 1993 lancierten Programms zur Gesichtserkennung eine grosse Datenbank aufzubauen begann, um die verschiedenen Algorithmen vergleichen zu können. Seither wetteifern die auf dem Markt erhältlichen Gesichtserkennungsprogramme jedes Jahr am «Facial Recognition Vendor Test» um die besten Resultate; in der Ausgabe von 2020 schickten 99 Hersteller insgesamt 189 Algorithmen in den Vergleichskampf.

Der Terroranschlag vom 11. September 2001 steht für einen Wendepunkt im Einsatz der Gesichtserkennung. Unter dem Eindruck dieses Traumas setzten die Vereinigten Staaten durch, dass auch für Kurzaufenthalte nur noch Personen mit einem biometrischen Pass in die USA einreisen durften. Die Europäische Union und die Schweiz passten sich dieser Forderung an und legten mit der Einführung standardisierter biometrischer Gesichtsbilder ab dem Jahr 2006 die Grundlagen für die staatliche Gesichtserkennung.

Identifizieren und Verifizieren, Sprache und Sprechen

In der Regel kommen Stimm-, Sprach- und Gesichtserkennung zum Einsatz, um eine Person automatisch zu identifizieren und ihr dabei allenfalls Zugang zu einem bestimmten Dienst zu gewähren. Bei der Spracherkennung geht es in der Regel darum, die Befehle am Handy oder am PC nicht mehr über die Tastatur, sondern über gesprochene Befehle auszuführen.

Technisch gilt es, zwischen der Identifikation und der Verifikation eines Individuums zu unterscheiden. Die Verifikation einer Person ist insofern einfacher, als man die charakteristischen Merkmale einer Ton- oder Bilddatei einzig mit einer Musteraufnahme der fraglichen Person vergleichen muss. Stimmen Tonaufnahme bzw. Bild mit dem Muster überein, ist die Identität der Person bestätigt. Bei der Identifizierung hingegen wird die Ton- oder Bilddatei mit einer Vielzahl von Aufnahmen aus einer Datenbank abgeglichen, um herauszufinden, wer die betreffende Person ist.

Stimm- und Spracherkennung sind ebenfalls nicht deckungsgleich. Ein System, das Worte erfasst – also deren Bedeutung «versteht» – identifiziert nicht zwangsläufig die Stimme bzw. die sprechende Person. Bei der Spracherkennung geht es nämlich darum, den Inhalt einer Aussage zu erkennen, um gegebenenfalls anschliessend den entsprechenden Befehl auszuführen. Die Stimmerkennung hingegen zielt darauf ab, Menschen aufgrund biometrischer Kenngrössen wie Tiefe und Klangfarbe der Stimme zu identifizieren. Damit kann beispielsweise die Sachbearbeiterin der Bank oder der Krankenkasse feststellen, ob tatsächlich diejenige Person am Draht ist, die zu sein sie vorgibt.

Biometrische Daten verraten viel mehr als ein blosses Abbild

Einfache Bild- oder Tonaufnahmen lassen sich relativ leicht manipulieren und sind bei weitem noch keine biometrischen Daten. Vielmehr werden diese in einem aufwendigen Verfahren modelliert und berechnet. Bei der dreidimensionalen Methode etwa, auf welcher die Face-ID grosser Handy-Hersteller beruht, projiziert das Smartphone beim Scan mittels eines Punktprojektors 30 000 Punkte aus unsichtbarem Infrarotlicht auf das Gesicht der Besitzerin oder des Besitzers. Auf dieser Grundlage erstellt das Handy ein 3D-Modell, auf dem anschlies-

send charakteristische Strukturen und Merkmale des Gesichts identifiziert werden. Infrarotbild und 3D-Abbild werden zu einer mathematischen Formel umgewandelt und auf dem Chip des Geräts gespeichert. Bei jedem Entsperrvorgang ruft das Handy diese Formel ab, die sodann den Algorithmus ausführt.

Auch bei der Stimmanalyse erfasst und verarbeitet ein Algorithmus mehrere Tausend Merkmale, von denen die meisten für das menschliche Ohr nicht einmal hörbar sind. Sie umfassen weit mehr als Tonlage, Sprachmelodie, Atempausen und Rhythmus. Basierend darauf erstellt die Software den sogenannten Voiceprint, eine Art akustischen Fingerabdruck.

Weil biometrische Daten eng an eine Person gebunden sind und viel über diese verraten, gelten sie als besonders schützenswert. In einem Leitfaden definieren die schweizerischen Datenschützer biometrische Merkmale denn auch als «einmalige, spezielle körpereigene Kennzeichen einer Person, die – zumindest theoretisch – immer und überall eindeutig mit nahezu 100%iger Sicherheit nur diese einen Person zugeordnet werden können.»

Die Nutzung von Stimm-, Sprach- und Gesichtserkennungstechnologien durch Private widerspricht in der Regel dem Datenschutzrecht. Wer solche Daten bearbeiten will, kann schwerlich ein überwiegendes Interesse geltend machen, da es sich bei biometrischen Angaben um besonders schützenswerte Daten handelt. Hinzu kommt, dass die Einwilligung der Betroffenen in vielen Fällen kaum einzuholen ist, weil die Geräte ohne Rückgriff auf diese Daten gar nicht funktionieren. In manchen Fällen wie der verdeckten Identifikation wird die Einwilligung von vornherein gar nicht gesucht, was datenschutzrechtlich nicht zulässig ist.

Datenschutz mit weiterem Blickwinkel

Die automatisierte Erkennung von Personen aufgrund biometrischer Merkmale rief früh Kritik hervor. So wurde ins Feld geführt, die Ergebnisse automatisierter Personenerkennung seien allzu oft fehlerbehaftet und wiesen Verzerrungen auf: Bei der Gesichtserkennung werden beispielsweise Frauen und Menschen dunkler Hautfarbe systematisch schlechter erkannt als weisse Männer – ein Umstand, der nicht zuletzt darauf zurückzuführen ist, dass in den Datenbanken, die für das Training der Algorithmen verwendet werden, Bilder weisser

Männer lange in der Überzahl waren. Auch die Auswertung von Tondateien ist nicht vor Verzerrungen gefeit, indem beispielsweise allein schon eine schlechte Tonqualität Fehler bei der Interpretation begünstigt.

Als in den 1990er-Jahren der Einsatz biometrischer Identifikationssysteme noch in seinen Anfängen steckte, stand die Sorge um den Schutz der Daten im Vordergrund. Bemängelt wurde der drohende Verlust an Privatsphäre und persönlicher Freiheit, wenn Individuen bei jeder Bewegung und Handlung automatisch beschattet werden könnten. In den Folgejahren wurde dieser auf die Einzelnen fokussierte Blick jedoch als zu eng kritisiert. Fachleute machen sich daher zunehmend für eine umfassende Betrachtung der Folgen von Technologien stark, die eine Überwachung ermöglichen können. Denn ihr Einsatz gefährdet längst nicht nur das Recht auf Privatsphäre, sondern auch andere Grundrechte wie beispielsweise die Versammlungsfreiheit oder das Recht auf freie Meinungsäusserung, die essenziell für eine funktionierende Demokratie und das gesamtgesellschaftliche Zusammenleben sind.

Konformitätsdruck gefährdet Grundrechte und Demokratie

Wer befürchten muss, ständig beobachtet zu werden, neigt dazu, sich anzupassen und die eigenen Äusserungen zu zensieren. Dieser Druck zu linientreuem Verhalten widerspricht einer Rechtsetzung, die – wie in der schweizerischen Bundesverfassung – der Diskriminierung von Menschen aufgrund ihrer Identität, Lebensform und Überzeugungen eine Absage erteilt und die Meinungsäusserungsfreiheit hochhält.

Aus juristischer Sicht erweist sich der Datenschutz nach wie vor als Knackpunkt. Denn biometrische Daten geben Aufschluss über einzigartige Wesenszüge eines Menschen und sind daher besonders schützenswerte Personendaten. Werden solche Daten bearbeitet, droht ein schwerer Eingriff in die individuellen Grundrechte – darunter insbesondere das in der Bundesverfassung festgehaltene Recht auf Privatsphäre und der Schutz vor Missbrauch persönlicher Daten. Der Umgang mit biometrischen Daten wird durch den Umstand erschwert, dass sich die Anonymisierung mit den meisten Anwendungen von Stimm-, Sprach- und Gesichtserkennung praktisch nicht vereinbaren lässt. Der bereits

erwähnte Konformitätsdruck gefährdet ausserdem die Meinungsäusserungs-, Versamlungs- und Vereinigungsfreiheit, wie sie durch die Verfassung ebenfalls gewährleistet sind.

Zu den wichtigen Grundsätzen im Umgang mit biometrischen Daten gilt die Verhältnismässigkeit: Es sollen nur so viele Daten erhoben werden, wie es eine bestimmte Aufgabe erfordert. Wichtig ist auch die Zweckbindung, d.h., dass die Daten nur für dasjenige Vorhaben verwendet werden, für welches sie gesammelt wurden. Die Eignung einer Technologie ist ebenfalls wichtig. Nur wenn diese geeignet ist, um das mit ihr verfolgte Ziel zu erreichen, ist ihre Verwendung rechtmässig. Ausserdem ist Transparenz geboten: Menschen müssen wissen, dass ihre biometrischen Angaben erfasst werden – und sie müssen die Möglichkeit haben, der Erhebung zuzustimmen oder sie zu verweigern, ohne dass ihnen dadurch Nachteile erwachsen. Schliesslich muss auch die sichere Aufbewahrung der Daten gewährleistet sein, um zu verhindern, dass diese gehackt und missbraucht werden. Die Nutzung solcher Technologien erfordert daher ein präzise ausgestaltetes Gesetz. Dieses muss nicht nur den Zweck des Technikeinsatzes benennen, sondern auch seine Einschränkungen und insbesondere die Beschränkung der Datenbearbeitung auf das strikt Notwendige.

Unterschiedliche Prioritäten in der Forschung

Die Gesichtserkennung ist ein Forschungsfeld, das in China bereits seit längerem besonders intensiv bearbeitet wird. Dort wurden im Jahr 2020 gut doppelt so viele wissenschaftliche Artikel zu diesem Thema veröffentlicht wie in den USA oder Indien. Ganz Europa – die Schweiz eingeschlossen – kommt auf rund ein Drittel der chinesischen Publikationsmenge. Etwas anders sieht es aus bei der Spracherkennung, bei der die USA deutlich obenaus schwingen, gefolgt wiederum von China, das seit 2018 seine Bemühungen auf diesem Gebiet aber verstärkt. Auffällig ist das grosse Gewicht, das im Reich der Mitte der Stimm-, Sprach- und Gesichtserkennung im Vergleich zu anderen Forschungsthemen zugewiesen wird.

Wenn die Technik die Ohren spitzt

Smarte Lautsprecher haben in zahlreichen Büros und Haushalten Einzug gehalten. Statt Mails und Briefe zu tippen, spricht man sie dem PC laut vor, Wecker und Kalendereinträge werden dem Smartphone diktiert, und auch die Zimmertemperatur oder andere Komponenten der smarten Wohnung lassen sich über einen Sprachbefehl regeln.

Wenn es genügt, einen Befehl bloss auszusprechen, statt ihn über die Tastatur oder ein anderes Instrument einzugeben, liegen die Vorteile auf der Hand: Der smarte Lautsprecher ersetzt eine Vielzahl von Fernbedienungen. Zudem lassen sich die unterschiedlichen Geräte einfacher handhaben, und die Technik erleichtert somit auch all jenen den Alltag, die mit einer körperlichen Behinderung zurechtkommen müssen.

Smarte Lautsprecher: Es hören viele mit

Technisch bestehen smarte Lautsprecher mindestens aus einem Mikrophon, einem Lautsprecher und einer über WLAN und Internet laufende Verbindung zum jeweiligen Anbieter. Über diese laufen

die cloudbasierten Dienste und Funktionen, die den Nutzenden zur Verfügung stehen. Die Anbieter wiederum verwenden die Aufnahmen, um ihre Leistungen zu verbessern. Die erhobenen Daten werden somit an verschiedenen Orten gespeichert: Im smarten Lautsprecher und auf dem Server des Anbieters; sofern mit dem Lautsprecher ein Handy verbunden ist, werden die Aufnahmen auch darauf abgelegt.

Damit die smarten Lautsprecher das Aktivierungswort für einen gesprochenen Befehl hören können, müssen sie ständig auf Empfang sein. Sobald das Aktivierungswort fällt, beginnen sie mit der Aufnahme, die sodann dem Server des Herstellers übermittelt wird. Dieser wertet mittels Spracherkennung den Befehl aus und sendet das Ergebnis über eine automatisierte Sprachausgabe an die Nutzenden zurück.

Eine Funktion, die vielen nicht bekannt ist oder zumindest nur selten verwendet wird: Aufgezeichnete Tonaufnahmen lassen sich bei zahlreichen Geräten über eine App oder eine Website löschen. Aus Sicht des Datenschutzes ist dies zu begrüßen. Denn gewisse Anfragen oder Befehle sind heikel – etwa, wenn über einen Sprachbefehl ein Termin



beim Arzt oder der Scheidungsanwältin vereinbart wird. Ausserdem zeichnen die smarten Lautsprecher nebst dem Sprachbefehl weitere Daten wie Kalendertag und Uhrzeit auf, die eventuell Rückschlüsse auf die Gewohnheiten der betreffenden Person zulassen. Sind noch weitere Menschen anwesend, wird ihre Stimme mit aufgezeichnet. Das wiederum kann Hinweise auf die Zusammensetzung des Haushalts oder des Freundeskreises geben. Ausserdem sind nicht nur die gesprochenen Befehle und Anfragen aufschlussreich, sondern auch der Klang und die Tonhöhe der Stimme. Diese verrät einiges über den körperlichen Zustand eines Menschen – etwa, ob er erkältet ist oder zu viel getrunken hat. Auch die Gefühlslage ist oft herauszuhören, wenn jemand fröhlich lacht oder niedergedrückt seufzt.

Authentifizierung über die Sprache in der automatisierten Schalterhalle

Einige Schweizer Banken und die Post nutzen Sprachaufnahmen, um ihre Kundschaft zu identifizieren. Sie versprechen sich dadurch erhöhte Sicherheit und Zeitersparnis. Bei der Migros Bank konnte die Gesprächsdauer dank der Stimmauthentifizierung um zwanzig Prozent verkürzt werden. Bisher wird das Verfahren vorwiegend beim Telefon-Banking eingesetzt. Denkbar wäre, es künftig auch in Beratungsgesprächen oder für virtuelle Assistenzsysteme zu verwenden.

Bevor die Aufzeichnung beginnt, wird die Kundschaft gefragt, ob sie der Speicherung ihres Anrufs zustimmt, sodass dieser beim nächsten Anruf zur Authentifizierung genutzt werden kann. Bei der

PostFinance ist es ausserdem möglich, auf dem Webportal die Stimmerkennung zu aktivieren oder abzulehnen.

Ob sich die Identität einer Person über die Stimme eindeutig bezeugen lässt, ist umstritten. Die Einen sind der Ansicht, diese sei in hohem Grad individuell, und sie halten daher die biometrische Stimmerkennung für sicher. Andere weisen darauf hin, es sei Journalisten und Hackern auch schon gelungen, akustische Zutrittssperren zu übertölpeln. Sie nutzten dabei Tonschnipsel aus Youtube-Videos und Computerprogramme, die auch für die künstliche Herstellung täuschend echt wirkender Tonaufnahmen – sogenannter Deepfakes – verwendet werden.

Fest steht, dass die biometrische Authentifizierung nur so lange sicher ist, wie die Daten geschützt bleiben. Einmal ausgespäht, lassen sich diese für die unzweifelhafte Erkennung einer Person nicht mehr verwenden. Denn Stimm- und Gesichtsabdruck sind eng und dauerhaft mit einem Menschen verbunden und können – anders als ein Passwort – nicht ohne weiteres verändert werden.

Kriminelle abwehren

Kriminelle setzen viel daran, sich über die Stimmerkennung Zugriff auf sensible Daten zu verschaffen. Daher empfehlen Fachleute den Banken und anderen Unternehmen, die mit heiklen Angaben arbeiten, zusätzlich zur Stimmerkennung auch ein Passwort zu verwenden. Die Kombination beider Verfahren erhöht die Sicherheit.



Die zunehmende Verbreitung smarter Lautsprecher macht auch diese zum Ziel von Hackern. Sie können versuchen, über die Sprachaufnahmen einer Person den Authentifizierungsmechanismus ihres smarten Lautsprechers zu täuschen, um sich sensibler Daten der Betroffenen zu bemächtigen.

Die Forschung nimmt sich derzeit der Sicherheitsprobleme smarter Lautsprecher an. Ein Ansatz besteht darin, Sprachmerkmale von der Aufnahme zu entfernen, die es für die Interaktion mit dem smarten Lautsprecher gar nicht braucht. Um Unbefugten den Zugriff auf die Daten zu verwehren, arbeiten andere wissenschaftliche Teams an Programmen zur Netzwerkanalyse, welche erkennen, wenn im Internet Tondateien übertragen werden. So können Nutzende gewarnt werden, wenn ihr Aktivierungswort ausgelöst wurde.

Zurückhaltung in der Schweiz

Im Jahr 2020 hielt Amazon mit 22 Prozent aller weltweit verkauften smarten Lautsprecher den Spitzenplatz inne. Dichtauf folgte Google mit 17 Prozent. 2018 nutzten in den USA 18 Prozent der Bevölkerung smarte Lautsprecher, in Deutschland waren es zehn Prozent. Im Vergleich dazu ist die Schweiz zurückhaltend. 2018 nutzte gerade mal ein Prozent der Schweizer Bevölkerung smarte Lautsprecher; ein Jahr später waren es drei Prozent. Die im Rahmen der Studie von TA-SWISS im Oktober 2021 durchgeführte repräsentative Umfrage ergab, dass sich in 63 Prozent der Haushalte kein smarter Lautsprecher befindet. Knapp die Hälfte der Nutzerinnen und Nutzer eines solchen Gerätes haben es im Laufe des letzten Jahres angeschafft. Nur neun Prozent besitzen eines seit mehr als drei Jahren. Von den Befragten, die keinen smarten Lautsprecher nutzen, geben 41 Prozent an, auch künftig darauf verzichten zu wollen – in erster Linie, weil sie keinen sinnvollen Einsatzzweck für diese Geräte sehen, und aus Datenschutzbedenken.

Für den Einlass Gesicht zeigen

Mit der Corona-Pandemie haben berührungslose Bezahl- und Zugangssysteme beträchtlich an Terrain gewonnen. Ausserdem haben sich viele Menschen daran gewöhnt, das Handy mittels Face-ID zu entsperren. Mag der selbstbestimmte Einsatz der Gesichtserkennung am eigenen Smartphone auch praktisch sein, der Einsatz dieser Technik beim Besuch eines Fussballspiels oder eines Rockkonzerts ist selten erwünscht und kann für den Einzelnen gar bedrohliche Folgen haben.

Behörden, insbesondere Zoll und Polizei, bedienen sich bereits seit Längerem der Gesichtserkennung. Allerdings ist es schwierig, Aufschluss über die Art des Einsatzes dieser Technik und dessen Ergebnisse zu bekommen – findet doch gerade die polizeiliche Nutzung der Gesichtserkennung oft im Verborgenen statt.

Die in den 1990er-Jahren durch US-amerikanische Polizeistellen verwendeten Systeme machten vorwiegend durch ihr Scheitern von sich reden. So wur-

de 2003 bei einem mehrjährigen, gross angelegten Test am Flughafen von Boston kein einziger Treffer gelandet. Auch in Florida führte der Einsatz einer entsprechenden Software nicht zum Erfolg; an einigen Tagen gab sie gar ausschliesslich falsch-positive Meldungen aus. Etwas bessere Resultate lieferte ein 2006 am Mainzer Bahnhof durchgeführter Test des deutschen Bundeskriminalamtes. Dieser erreichte immerhin eine Trefferrate von 30 Prozent, lag damit allerdings deutlich unter den anvisierten 80 Prozent.

Verhaltener Einsatz in der Schweiz

Hierzulande kommt die Gesichtskontrolle in erster Linie am Flughafen Zürich zum Einsatz, und zwar auf freiwilliger Basis: Nach einer halbjährigen Pilotphase ab dem Herbst 2017 haben Flugreisende seit 2020 die Möglichkeit, freiwillig durch die automatisierte Gesichtskontrolle zu gehen. Hinweisschilder machen auf die Gesichtserkennung aufmerksam, und die

Leute können ihre Personalien weiterhin auf konventionelle Art am Schalter überprüfen lassen.

Vom Flughafen Zürich abgesehen, nehmen in der Schweiz soweit bekannt einzig die Polizei Aargau und die Polizei St. Gallen Rückgriff auf die Gesichtserkennung. Andere Polizeistellen testen entsprechende Systeme oder beschäftigen menschliche Profis für Gesichtserkennung. Die Kantonspolizei Basel-Stadt besitzt sieben Tesla-Fahrzeuge, die je mit acht Kameras ausgestattet sind; diese werden zwar nicht zur Gesichtserkennung genutzt, es wäre jedoch ein Leichtes, sie mit der erforderlichen Software zu verknüpfen.

Im Austausch mit den zwei Kantonspolizeien, die in der Schweiz von der Gesichtserkennung Gebrauch machen, konnte sich die Projektgruppe von TA-SWISS vergewissern, dass sich beide bemühen, einen sicheren und auch rechtlich wie ethisch vertretbaren Einsatz der Technik zu gewährleisten. So haben beide eine Datenschutzfolgenabschätzung (DSFA) durchgeführt. Und um die Daten sicher aufzubewahren, speichern sie diese auf Servern, die nicht mit dem Internet oder anderen Netzwerken verbunden sind.

Die Studie von TA-SWISS deckt aber auch Schwächen bei der polizeilichen Nutzung der Gesichtserkennung auf. Weder die Polizei des Kantons Aargau noch die aus St. Gallen veröffentlichen die Ergebnisse ihrer DSFA. Auch haben beide darauf verzichtet, Leistungsfähigkeit und Sicherheit ihrer Systeme von unabhängiger Stelle prüfen zu lassen. Welche Stellen auf welche Daten zurückgreifen können, bleibt ebenfalls im Dunkeln. Zudem vertreten beide Kantonspolizeien die Haltung, es sei nicht nötig, Personen zu informieren, wenn sie ins Visier der Gesichtserkennung geraten seien. Denn es mache keinen grundsätzlichen Unterschied, ob eine Software oder das menschliche Auge die Bilder durchsuche.

Gesichtserkennung in Echtzeit als Startschuss zur Massenüberwachung?

In der Schweiz findet keine Massenüberwachung in Echtzeit statt. Sie wäre rechtlich auch nicht statthaft, da sie fundamentale Grundrechte wie die Versammlungsfreiheit gefährdet. Auch der Grundsatz der Verhältnismässigkeit würde verletzt, wenn beispielsweise alle Menschen überwacht würden, die an einer Demonstration teilnehmen, bloss um Vandalenakte Einzelner zu ahnden. Dennoch ist es die Massenüberwachung, die auch hierzulande intensiv diskutiert wird. Die Corona-Pandemie war nicht dazu angetan, die Befürchtungen zu dämpfen. Denn verschiedene Länder setzten die Gesichtserkennung ein, um die Einhaltung der Corona-Regeln zu überwachen.

Die Fokusgruppen, die im Rahmen der Studie von TA-SWISS zur Diskussion eingeladen wurden, forderten, die Massenüberwachung durch den Staat sei zu verhindern. Denn dieser erhalte gegenüber den Bürgerinnen und Bürgern zu viel Macht und büsse dadurch an Vertrauen ein. Befürchtet wird ein Dammbruch, wonach die Gesichtserkennung zunächst zwar nur zur Aufdeckung schwerer Verbrechen gegen Leib und Leben eingesetzt werde; habe sich die Technik dabei bewährt, werde ihr Anwendungsbereich auf geringfügigere Vergehen wie Taschendiebstahl ausgeweitet, bis sich die Bevölkerung an die flächendeckende und ständige Überwachung gewöhnt habe. Schliesslich drohten Zustände wie in China, das mit seinem Sozialkreditsystem gerne als abschreckendes Beispiel angeführt wird: Bürgerinnen und Bürger des Landes, die identifiziert werden, wenn sie bspw. bei Rot über den Zebrastreifen gehen oder gegen andere Vorgaben des Staates verstossen, haben mit Sanktionen zu rechnen, indem sie etwa bei Stellenbewerbungen oder auf Wohnungssuche benachteiligt werden.

Optik vereint mit Akustik im Kampf gegen Rassismus

Während in europäischen Ländern der Staat zurückhaltend auf die Gesichtserkennung zurückgreift, geben sich bestimmte private Akteure Forscher: Nachdem verschiedentlich Affenlaute, Naziparolen und andere rassistische Beleidigungen dunkelhäutiger Spieler mehrmals zum Abbruch von Fussballspielen geführt haben, erwägen italienische Clubs, mit der Gesichts- und Spracherkennung fremdenfeindliche Pöbeleien zu bekämpfen. Rassistische Sprechchöre könnten mittels Spracherkennung ermittelt und die Fehlbaren anschliessend dank Gesichtserkennung identifiziert werden. Noch sind in den italienischen Stadien keine entsprechenden Systeme im Einsatz. Doch Luigi De Siervo, Generaldirektor der höchsten italienischen Fußballliga, erklärte im Herbst 2019 seine Bereitschaft, die Gesichtserkennung einzusetzen, um «die Leute, die diesen wunderbaren Sport ruinieren, einen nach dem anderen zu erwischen».

In der Schweiz fehlt eine gesetzliche Grundlage, um die Sprach- und Gesichtserkennung in Sportstadien oder im Rahmen anderer privatwirtschaftlicher Anlässe anzuwenden. Allerdings sprach sich eine knappe Mehrheit der im Rahmen der Studie Befragten zugunsten der Gesichtserkennung in Sportstadien aus; diese Anwendung der Technik stiess im Vergleich zur Nutzung für andere Zwecke auf die grösste Akzeptanz, die ansonsten für die Gesichtserkennung gering ist.

Klare Direktiven für den polizeilichen Einsatz der Gesichtserkennung gefordert

In der im Rahmen der TA-SWISS-Studie durchgeführten repräsentativen Umfrage wurde der polizeiliche Einsatz der Gesichtserkennung nicht rundweg abgelehnt; ein Drittel sprach sich für deren Einsatz aus, ein Drittel erklärte sich zu wenig vertraut mit der Materie, ein knappes Drittel lehnte die Verwendung dieser Technik ab. Aus Sicht der Befürworterinnen und Befürworter bietet sie sich in erster Linie für das Auffinden vermisster Personen an, gefolgt von der Terrorismusbekämpfung. Wer die Gesichtserkennung ablehnt, tut dies in erster Linie aus Angst vor einem Missbrauch der Daten, weil diese Technik das öffentliche Leben zu stark beeinflussen und weil sie zur anlasslosen Massenüberwachung führen könnte. Alle Befragten äusserten die Meinung, nur zugelassenes Personal solle die Gesichtserkennung einsetzen dürfen, müsse jeden Einsatz protokollieren und auch transparent darüber kommunizieren. Zudem brauche es entsprechende gesetzliche Grundlagen, und die Verwendung der Technik müsse regelmässig von unabhängigen Fachleuten begleitet und evaluiert werden.



Wenn die Technik das Innerste nach aussen kehrt

Eine erfahrene Ärztin oder ein feinfühligler Arzt erfährt anhand Veränderungen des Hautbildes, des Gesichtsausdrucks und der Stimme seines Gegenübers viel über dessen Gesundheitszustand. Es zeichnet sich ab, dass dereinst die Stimm-, Sprach- und Gesichtserkennung Krankheiten gar früher erkennen könnten als das medizinische Fachpersonal. Wenn es aber darum geht, Emotionen zu erfassen, stehen der Technik noch grössere Hindernisse entgegen.

«Für Transgender ist die App echt hilfreich, um die Stimme zu trainieren bzw. Veränderungen festzustellen», hält ein Kommentar zur App «Voice Pitch Analyzer» fest. «Hi, bin selbst FtM (female to male) und benutze diese App, weil das echt interessant ist, selbst zu sehen, wie sich die Tonlage mit der Hormontherapie verändert», schreibt jemand anderes. Voice Pitch Analyzer von Purr Programming ist eine App, die kostenlos heruntergeladen und verwendet werden kann, um eine Stimme zu analysieren – ein Merkmal, das in der Regel verrät, ob eine Frau oder ein Mann spricht. Applikationen für die Analyse und das Training der Stimme sollen Transpersonen dabei unterstützen, ihre Sprechstimme so zu verändern, dass sie sich dem geschlechtsspezifischen Ziel annähert, dass sie also je nach Wunsch weiblicher oder männlicher klingt.

Ob sportliche Leistung, Schlafrhythmus, Herzschlag oder Eisprung – im Internet gibt es mittlerweile zahlreiche Applikationen, um den eigenen Körper zu überwachen. Künftig könnten die Programme verstärkt auch auf die Stimm-, Sprach- und Gesichtserkennung zurückgreifen, um den physischen Zustand einer Person einzuschätzen. Der berühmte «Dr. Google», der bei unklaren Symptomen oft als erstes konsultiert wird, erhielt damit weitere leistungsstarke Analyseinstrumente zur Hand.

Was in Gestalt der eingangs erwähnten mobilen Stimmanalyse über PC oder Handy harmlos wirkt, könnte bei ausgeklügelteren Versionen problematisch werden. Denn mit den Fortschritten der Diagnosetools wächst die Gefahr, dass Laien überfordert werden, wenn sie sich über die Kamera und das Mikrofon des Computers eigenständig medizinisch

überwachen. Derzeit sind allerdings die auf Stimm- oder Gesichtserkennung beruhenden leistungsstarken Diagnoseinstrumente den medizinischen Fachpersonen vorbehalten.

Im Einsatz für die Medizin

Wenn wir sprechen, regelt das Gehirn das Zusammenspiel von bis zu hundert Muskeln. Die Stimme kann daher auf eine Vielzahl von Krankheiten hinweisen. Dabei können auch Laute aufschlussreich sein, die das menschliche Ohr leicht überhört. So untersuchten Forschende den «Aahh»-Laut von gesunden und an Parkinson erkrankten Menschen und fanden zehn akustische Merkmale, anhand derer sich die Krankheit mit einer Genauigkeit von fast 99 Prozent bestimmen lässt. Die Wortwahl bzw. Schwierigkeiten beim Auffinden von Wörtern wiederum kann auf Alzheimer hindeuten. Diese Krankheit spüren die Stimm- und Spracherkennungssysteme in Testumgebungen mit einer Erkennungsrate von 80 bis 90 Prozent auf.

Seit der Corona-Pandemie wird am charakteristischen Hustenton geforscht, um die Krankheit möglichst frühzeitig zu diagnostizieren. Auch Herzleiden könnten anhand der Stimme aufgedeckt werden, weil bestimmte Muster der Stimmfrequenz mit schweren Leiden der Herzkranzgefässe einhergehen.

Schliesslich widerspiegelt die gesprochene Stimme die psychische Verfassung eines Menschen. Tempo, Rhythmus, Tonhöhe und Lautstärke verraten, ob er aufgeregt, ängstlich, niedergeschlagen oder manisch ist. Bei der Behandlung von Depressionen ist die Spracherkennung mittlerweile zum Standard geworden: Eine monotone und gegenüber der üblichen Stimme etwas erhöhte Tonlage kann darauf hindeuten, dass jemand suizidgefährdet ist. Auch für die Diagnose von Autismus sowie von Aufmerksamkeitsdefizit- oder Hyperaktivitätsstörungen wurden sprachliche Merkmale herausgearbeitet.

Die Gesichtserkennung ist in der Medizin weniger verbreitet als die Sprech- und Spracherkennung,

wird aber vorangetrieben. Im Rahmen des EU-Projekt SEMEOTICONS wurde eine Art Spiegel mit verschiedenen Sensoren und Kameras entwickelt. Dieser war in der Lage, nebst psychischen Faktoren auch die Fitness und den Ernährungszustand von Personen zu erkennen; dazu analysierte er die Farbe von Haut und Schleimhäuten sowie die Verteilung des Fettgewebes unter der Haut und das Schwitzmuster. Das an der Universität Bonn entwickelte System namens DeepGestalt soll anhand eines Gesichtsfotos einer Person seltene Krankheiten oder Gendefekte aufdecken können. Die Software vergleicht hierfür den Gesichtsausdruck auf dem Bild mit zahlreichen Fotos von Personen mit bestimmten diagnostizierten Krankheiten; die Datenbank umfasst 17'000 Fotos von mehr als 200 komplexen Krankheitsbildern. Verschiedene Forschungsteams arbeiten an weiteren Systemen, um mit Gesichtsbildern seltene Krankheiten aufzudecken.

Eine Schwierigkeit für den medizinischen Einsatz von Stimm-, Sprach- und Gesichtserkennung besteht im Aufbau einer verlässlichen Datenbasis. Aktuell gibt es nur wenige Datenbanken für die Stimmanalyse. Zudem zweifeln gewisse Fachleute an der Verlässlichkeit und Aussagekraft von Sprachmerkmalen: Zu rasch wird die gesprochene Sprache verfälscht, etwa durch Heiserkeit oder allergischen Schnupfen. Auch Herkunft und Kultur eines Menschen beeinflussen die Lautstärke und das Tempo seiner Rede. Der

Forschungsbedarf rund um den Einsatz von Stimm-, Sprach- und Gesichtserkennung in der Medizin ist also noch beträchtlich.

Beim Einsatz der Erkennungstechnologien in der Medizin stellen sich im Sinn des Arztgeheimnisses besondere Anforderungen an eine sichere Handhabung und Speicherung der Daten. Applikationen, die allenfalls bei der Selbstdiagnose zum Einsatz kommen könnten, gelten als Medizinprodukt. Da solche Applikationen zurzeit noch kaum zu finden sind, lässt eine Regulierung noch auf sich warten.

Die Gefühlswelt ausloten

Die Stimm-, Sprach- und Gesichtserkennung könnte künftig auch in Bewerbungsverfahren zum Einsatz kommen. Sie könnte beispielsweise das Auswahlverfahren beschleunigen, indem in einem ersten Schritt eine Software die Vorauswahl besonders geeigneter Kandidatinnen und Kandidaten unterstützen würde. Die Technologie liesse sich einsetzen, um die Wirkung des Auftretens einer Person zu testen und damit zu prüfen, ob sie zur ausgeschriebenen Tätigkeit passt.

Denkbar wäre auch, dass Privatversicherungen sich der Technik bedienen, um ihrer Kundschaft personalisierte Angebote zu unterbreiten. Denn ähnlich wie in der Medizin, könnte auch die Versicherungs-



branche die Hinweise auf Fitness und körperliche Eigenschaften ausnutzen, die sich aus Gesichts- und Stimmdateien gewinnen lassen. Überhaupt dürfte die Emotionserkennung in unterschiedlichen Kreisen auf beträchtliches Interesse stossen: Eine Bank könnte sich daraus Vorteile beim Marketing erhoffen, die Polizei dürfte versucht sein, sie zum Aufdecken von Lügen einzusetzen, wie es bereits beim von der EU geförderten Projekt iBorderCtrl erprobt wurde, und Stadionbetreiber könnten sich von der Technik versprechen, frühzeitig gewaltbereite Fans zu entdecken.

Gleich wie beim psychologischen Test oder bei der Handschriftenanalyse müssten die Betroffenen in die Analyse ihres Gesichtsbildes oder ihrer Stimme zustimmen. Sie müssten zudem ihre Einwilligung freiwillig geben – eine Voraussetzung, die angesichts der Machtasymmetrie zwischen Arbeitgeber und Arbeitnehmer oder Versicherung und Versicherten schwer zu schaffen ist. Würde die Technologie während eines Bewerbungsgesprächs oder beim Abschluss einer Versicherungspolice versteckt eingesetzt, käme dies einer Persönlichkeitsverletzung gleich; insbesondere würde gegen den Grundsatz von Treu und Glauben verstossen.

Bei vielen Fachleuten stösst die Emotionserkennung generell auf Skepsis. Sie stellen infrage, dass Gefühle, die sich individuell ganz unterschiedlich äussern, von einer Software verlässlich interpretiert werden können. Allein der Umstand, dass auch die kulturelle Herkunft eines Menschen einen erheblichen Einfluss darauf ausübt, wie er Gefühle zeigt, belegt die Schwierigkeit der Emotionsanalyse.

Bedenken hinsichtlich der Emotionsanalyse

In der Schweiz weckt sowohl ein medizinischer Einsatz der Gesichts-, Stimm- und Sprachanalyse als auch deren Verwendung zur Emotionserkennung oder zur Aufmerksamkeitsanalyse in der Schule beträchtliche Vorbehalte. Ein knappes Viertel der Befragten hätte keine Bedenken, wenn die Technologie zur Diagnose körperlicher Krankheiten verwendet würde, während 37 Prozent grosse Bedenken hätten; dabei wiegt die Sorge am schwersten, dass Krankenkassen Zugriff auf die Ergebnisse erhielten. Mit Blick auf die Emotionsanalyse äusserten 65 Prozent der Befragten die Sorge, die Software könnte nicht korrekt funktionieren. Auch denken 62 Prozent der Befragten, Emotionen seien zu vielschichtig, als dass sie eine Software korrekt erkennen könne.

Keine Ablenkung in der Schule zugelassen

Viele Gesellschaften verfolgen die Leistungen ihrer Schulkinder mit Argusaugen. Denn vom gut ausgebildeten Nachwuchs hängt der wirtschaftliche Erfolg einer Nation ab. Schulen in verschiedenen Staaten – neben China sind es insbesondere englischsprachige Länder – setzen Technologien zur Aufmerksamkeitsanalyse ein. US-Schulen investierten in jüngerer Vergangenheit jährlich 2,7 Milliarden US-Dollar in Überwachungs- und Sicherheitsprodukte, um sich gegen Amokläufe und andere Gewalttaten zu wappnen. Zwar gibt es bis jetzt keine Hinweise darauf, dass in



den USA die Aufmerksamkeit der Kinder überwacht wird. Sind aber Videosysteme erst einmal installiert, lassen sie sich mit relativ geringem Aufwand und zu bescheidenen Kosten von knapp 200'000 Schweizer Franken pro Schule für die Aufmerksamkeitsanalyse aufrüsten.

In China werden in bestimmten Schulen nicht nur die Kinder, sondern auch die Lehrkräfte observiert. Weder bei den Schülerinnen und Schülern noch bei ihren Eltern wird die Zustimmung eingeholt. Die Systeme zeichnen beispielsweise auf, wie lange die Kinder ihren Blick auf die Tafel richten und errechnen daraus eine Aufmerksamkeits-Punktzahl, die Lehrkräfte und Eltern einsehen können. Der Einsatz solcher Technik zielt auf die Verbesserung des Schulerfolgs ab: Nicht nur wird äusserer Ablenkung ein Riegel geschoben, sondern man verspricht sich einen Unterricht, der besser auf die individuellen Stärken und Schwächen der Kinder zugeschnitten ist. Aus der Psychologie ist allerdings bekannt, dass zu viel Kontrolle auch Stress hervorrufen kann und damit den Lernerfolg behindert. Auch zweifeln etliche Fachleute, dass die Analyse von Gesichtsausdrücken zuverlässig Aufschluss über die Aufmerksamkeit eines Menschen gibt.

Emotionen können grundsätzlich auf allen Systemen und Plattformen analysiert werden, auf denen Bilddaten von Schulkindern oder Studierenden entstehen. Die Corona-Pandemie, die den Einsatz von Lern- und Kommunikationsplattformen wie Moodle, Google Classroom oder auch Zoom Meetings vorangetrieben hat, hat damit zugleich eine Basis gelegt, um künftig allenfalls die Emotionserkennung oder Aufmerksamkeitsanalyse einzuführen – zumal an gewissen Schulen solche Software bereits genutzt wird, um bei Tests das ordnungsgemässe Verhalten der Prüflinge zu überwachen.

Nicht alle wollen mit allen bekannt sein

Die Stimm-, Sprach- und Gesichtserkennung rückt die sog. Jedermann-Identifikation in Griffweite: Ausgerüstet mit den erforderlichen Geräten wären alle in der Lage, jede Person zu identifizieren, die ihren Weg kreuzt. Einen ersten Schritt in diese Richtung ging die Firma Alphabet 2014 mit der Lancierung ihrer Brille «Google Glass». Die Idee hinter dem Mini-Computer vor dem Auge: Auf das Glas sollten Informationen aller Art über die Umwelt eingeblendet werden, also beispielsweise über Gebäude oder andere Elemente in der Landschaft. Die dazu benötigten Angaben bezog die Brille aus dem Internet. Zwar schloss der Konzern die Gesichtserkennung ausdrücklich aus. Fachleute des Datenschutzes wiesen indes darauf hin, die Brille lasse sich relativ einfach mit bestehenden Programmen der Gesichtserkennung verbinden – zumal verschiedene Firmen darauf setzten, Alphabet werde vom selbstaufgelegten Verbot der Gesichtserkennung abrücken.

Datenbanken für die Identifikation Privater stehen jedenfalls zur Verfügung; allein die Firma NameTag hatte mehrere Millionen Gesichtsbilder zusammengetragen und verkündete, sich künftig aus sozialen Netzwerken weitere Portraitfotos zu besorgen. Nicht zuletzt aufgrund anhaltender Kritik aus der Öffentlichkeit und wegen des zu geringen Kaufinteresses stellte Alphabet 2015 den Vertrieb ihrer Brille an Private ein. Doch längst sind leistungsstarke neue Datenbrillen erschienen, etwa die Ray-Ban Stories, die in Kooperation mit Facebook entwickelt wurde. Hacker könnten auch diese Brillen um eine Gesichtserkennungsfunktion ergänzen. Unter Rückgriff auf eine Gesichter-Datenbank könnten dann beliebige Menschen im Vorbeigehen identifiziert werden. Die dank allgegenwärtiger Smartphones bereits heute schwindende Anonymität in der Öffentlichkeit wäre mit der Jedermann-Identifikation passé.

Dass datenschützerische Befürchtungen rund um die sogenannte Jedermann-Identifikation nicht aus der Luft gegriffen sind, hat der Skandal um die Firma Clearview AI gezeigt: Das 2017 gegründete US-amerikanische Unternehmen hatte im Internet – insbesondere aus sozialen Netzwerken – Milliarden von Portraitfotos abgegriffen, daraus eine umfangreiche Datenbank erstellt und in seine Gesichtersuchmaschine eingefügt. Diesen Dienst hatte die Firma sodann weltweit staatlichen und privaten Stellen

angeboten, ohne dass die Öffentlichkeit der betroffenen Länder oder gar die Eigentümer der verwendeten Bilder bzw. die auf den Bildern figurierenden Personen je darüber informiert worden wären. Gegen das Unternehmen wurden verschiedene Klagen eingereicht, die dazu führten, dass seit 2020 nur noch Sicherheitsbehörden auf die Software zugreifen können. Der Ukraine-Krieg hat die Firma wieder in die Schlagzeilen gebracht, als bekannt wurde, dass ukrainische Behörden die Software verwenden, um gefallene oder gefangene russische Soldaten zu identifizieren und deren Angehörigen zu informieren. Auch die afghanischen Taliban machen sich die Gesichtserkennung zunutze. Sie erbeuteten bei ihrer Machtergreifung nämlich nicht nur Fahrzeuge und Waffen der westlichen Streitkräfte, sondern auch Datensätze mit Millionen von Fingerabdrücken und Gesichtsbildern. Diese nutzen die Gotteskrieger nun auf ihrer Jagd nach Personen, die mit Organisationen aus dem Westen «kollaboriert» haben und nach anderen «Verrätern».

Absage an die Aufmerksamkeitsanalyse und die Jedermann-Identifikation

Der Einsatz der Stimm-, Sprach- und Gesichtserkennung in der Schule wird in der repräsentativen Umfrage von 56 Prozent der Befragten rundweg abgelehnt. Noch ausgeprägter fällt der Widerstand gegen die Jedermann-Identifikation aus – und zwar sowohl in der repräsentativen Befragung (51 Prozent äussern grosse Bedenken) als auch in den Fokusgruppen: Eine Welt mit Jedermann-Identifikation wäre eine Welt, «in der sie nicht leben möchte», so die Aussage einer teilnehmenden Person, die den Grundtenor aus den Diskussionsrunden wiedergibt. Unter anderem befürchteten die Teilnehmenden, die Jedermann-Identifikation werde zu vermehrtem und intensiverem Stalking führen. Diejenigen Personen, denen in der repräsentativen Umfrage die Jedermann-Identifikation keine Sorgen bereitet, begründen ihre Haltung u.a. damit, es gebe heute ohnehin keine Anonymität mehr.

Big Brother soll draussen bleiben: Empfehlungen

Biometrische Daten sind hoch sensibel und besonders schützenswert. Umso dringlicher ist die Regulierung der Stimm-, Sprach- und Gesichtserkennung bei besonders heiklen Einsätzen, etwa in der Medizin, in der Strafverfolgung, bei der Vergabe von Krediten oder Versicherungen und im Arbeitsumfeld.

Digitalisierte Sprechaufnahmen und Gesichtsbilder widerspiegeln körperliche Eigenschaften einer erwachsenen Person, die sich kaum mehr verändern. Sind solche biometrischen Angaben erst einmal kompromittiert, lässt sich ihre Unversehrtheit nicht wiederherrichten. Daher empfiehlt es sich generell, die Erhebung biometrischer Daten zu beschränken und bei Verfahren der Authentifizierung mehrere Faktoren zu verwenden, also beispielsweise das Gesichtsbild mit einem Passwort zu ergänzen.

Hoch riskante Anwendungen verbieten

Die automatisierte Echtzeit-Überwachung durch die Stimm-, Sprach- und Gesichtserkennung stösst in westlichen Demokratien auf weitestgehende Ablehnung und ist unvereinbar mit mehreren Grundrechten, die in der Bundesverfassung der schweizerischen Eidgenossenschaft verankert sind. Daher ist eine solche Echtzeit-Überwachung zu verbieten – ebenso wie die Einrichtung eines Systems von Sozialkrediten, welche auf Grundlage einer flächendeckenden Überwachung der Menschen das Wohl- oder Fehlverhalten der einzelnen Bürgerinnen und Bürger bewertet.

Der Einsatz von Datenbrillen und von anderen unauffälligen technischen Vorrichtungen, die eine heimliche Überwachung anderer Menschen mittels der Gesichtserkennung ermöglichen, ist im öffentlichen Raum ebenfalls zu verbieten.

Zu unterlassen ist des Weiteren die Aufmerksamkeitsanalyse an Schulen. In besonders heiklen Situationen, insbesondere in Spitälern, im Banken- und Versicherungswesen, in der Strafverfolgung oder am Arbeitsplatz, ist der Rückgriff auf vollautomatische Entscheidungssysteme, die auf der Stimm-, Sprach- und Gesichtserkennung beruhen, zu untersagen. Stattdessen müssen die Ergebnisse teilautomatisierter Entscheidungssysteme stets durch geschulte Fachleute kritisch überprüft und freigegeben werden.

Aufschub für die Emotions- und Krankheitserkennung

Solange die technische und organisatorische Zuverlässigkeit von Emotions- und Krankheitserkennung durch Gesichts- und Sprachdaten nicht gesichert ist, sollte die Anwendung dieser Technik in bestimmten Lebensbereichen einem Moratorium unterstellt werden. Besondere Vorsicht ist beispielsweise in den Bereichen Strafverfolgung und Versicherungen geboten. Insbesondere darf die Stimme einer beim Call-Center anrufenden Person oder während der Bewerberauswahl nicht auf Krankheiten oder Emotionen untersucht werden; dies wäre eine zweckwidrige Verwendung dieser biometrischen Daten.

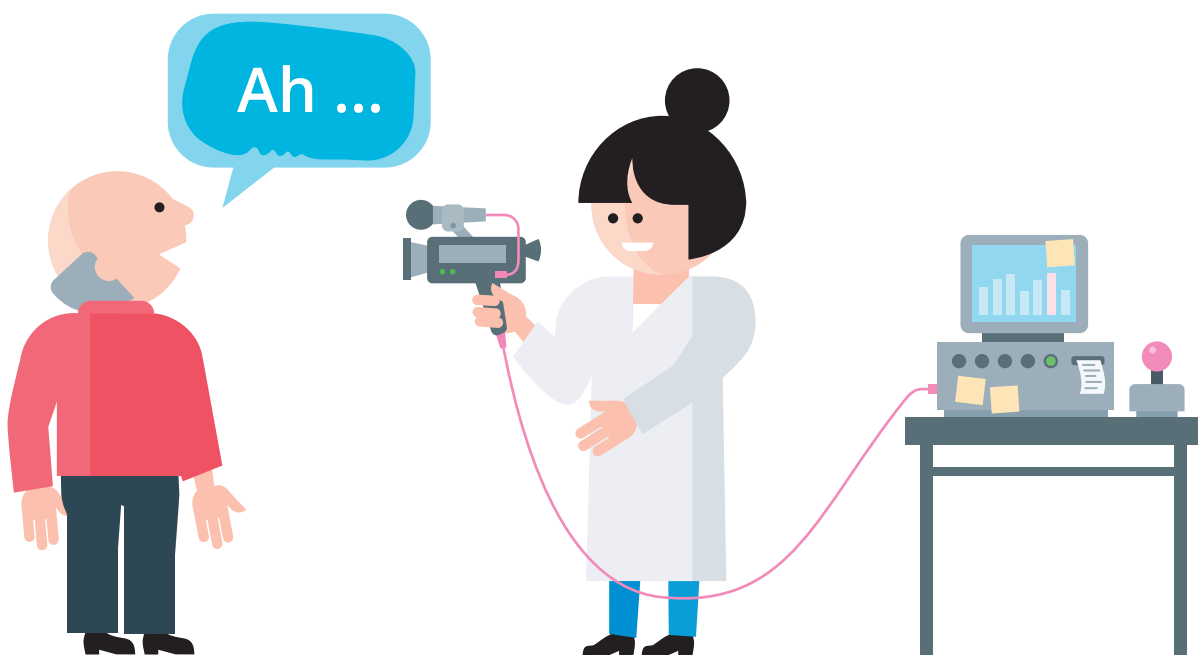
Bestimmte erwünschte Anwendungen könnten als Hoch-Risiko-Einsatz klassifiziert und damit unter Vorbehalt behördlicher Zustimmung gestattet werden.

Gesetzliche Lücken füllen, Weiterbildung fördern, Betroffene unterstützen

Damit öffentliche Stellen die Stimm-, Sprach- und Gesichtserkennung nutzen können, braucht es eine ausdrückliche gesetzliche Grundlage. Diese muss sowohl rechtsstaatliche Sicherungsmechanismen definieren als auch sicherstellen, dass die Notwendigkeit des Technologieeinsatzes überprüft wird.

Fachleute, die sich der Stimm-, Sprach- und Gesichtserkennung bedienen, auf die erhobenen Daten zugreifen und die Ergebnisse freigeben, müssen sich angemessen aus- und weiterbilden. Auch braucht es Leitlinien und Hilfestellungen, damit die Betreiber entsprechender Systeme den Grundsätzen des Datenschutzes folgen.

Kommen Technologien zur Stimm-, Sprach- oder Gesichtserkennung zum Einsatz, sollte transparent darauf hingewiesen werden. Wenn immer möglich, sind Alternativen anzubieten, auf die Betroffene



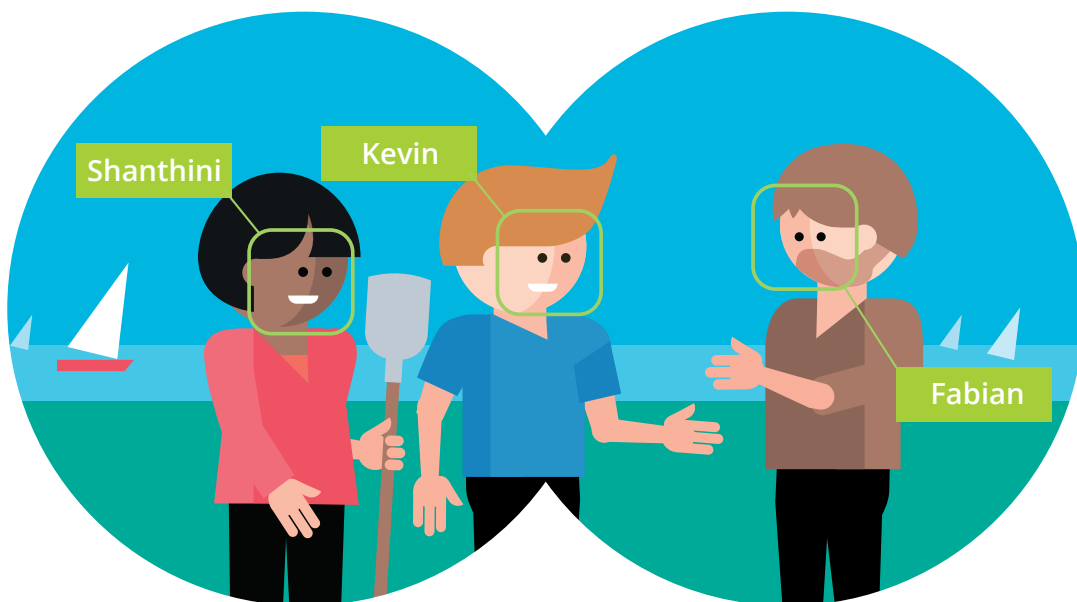
ausweichen können, ohne dadurch Nachteile – etwa eine längere Wartezeit oder Kosten – in Kauf nehmen zu müssen.

Zudem müssen Fachstellen eingerichtet werden, an die sich Personen wenden können, die sich vor den Nachteilen aus dem Einsatz von Stimm-, Sprach- und Gesichtserkennung schützen wollen sowie ihre Rechte kennen und diese auch durchsetzen möchten.

An die Adresse der Entwickler von Stimm-, Sprach- und Gesichtserkennungsgeräten und von entsprechenden Applikationen richtet sich die Empfehlung, die erhobenen Daten, wenn immer möglich, nur in den Geräten selber zu speichern und zu verarbeiten. Dadurch liessen sich sowohl der Datenschutz

als auch die Verfügungsgewalt der Nutzerinnen und Nutzer über die eigenen Daten stärken.

Schliesslich sollte auch die gesellschaftliche Debatte über die Vor- und Nachteile von Stimm-, Sprach- und Gesichtserkennung und über die Zulässigkeit und Grenzen ihres Einsatzes angeregt werden. Zudem sollte sich die Öffentlichkeit mit der Bedeutung der Stimm-, Sprach- und Gesichtserkennung für unser Zusammenleben und das Funktionieren unserer Demokratie befassen – nicht zuletzt vor dem Hintergrund, dass der Einsatz dieser Technik den sozialen Druck auf die Menschen erhöht und der belehrenden Bewertung der Bürgerinnen und Bürger resp. ihrer Handlungen Vorschub leistet. Die vorliegende Studie gibt einen Anstoss zu dieser öffentlichen Auseinandersetzung.



Mitglieder der Begleitgruppe

- Dr. Bruno Baeriswyl, Datenschutzexperte, Mitglied des Leitungsausschusses von TA-SWISS, Leiter der Begleitgruppe
- Dominik Brumm, Head of Development Cubera
- Prof. Dr. Volker Dellwo, Institut für Computerlinguistik, Universität Zürich
- Dr. Jean Hennebert, Université de Fribourg
- Dr. Anna Jobin, Soziologin, Alexander von Humboldt Institut für Internet und Gesellschaft
- 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

Projektleitung bei TA-SWISS

- Dr. Elisabeth Ehrensperger, Geschäftsführerin
- Dr. Christina Tobler, Projektverantwortliche (2020 – 2021)
- Dr. Laetitia Ramelet, Projektverantwortliche (2022)

Impressum

Auf Schritt und Tritt beobachtet

Kurzfassung der Studie « Automatisierte Erkennung von Stimme, Sprache und Gesicht: Technische, rechtliche und gesellschaftliche Herausforderungen»

TA-SWISS, Bern 2022

TA 79A/2022

Autorin: Dr. Lucienne Rey, TA-SWISS, Bern

Produktion: Dr. Laetitia Ramelet und Fabian Schluelp, TA-SWISS, Bern

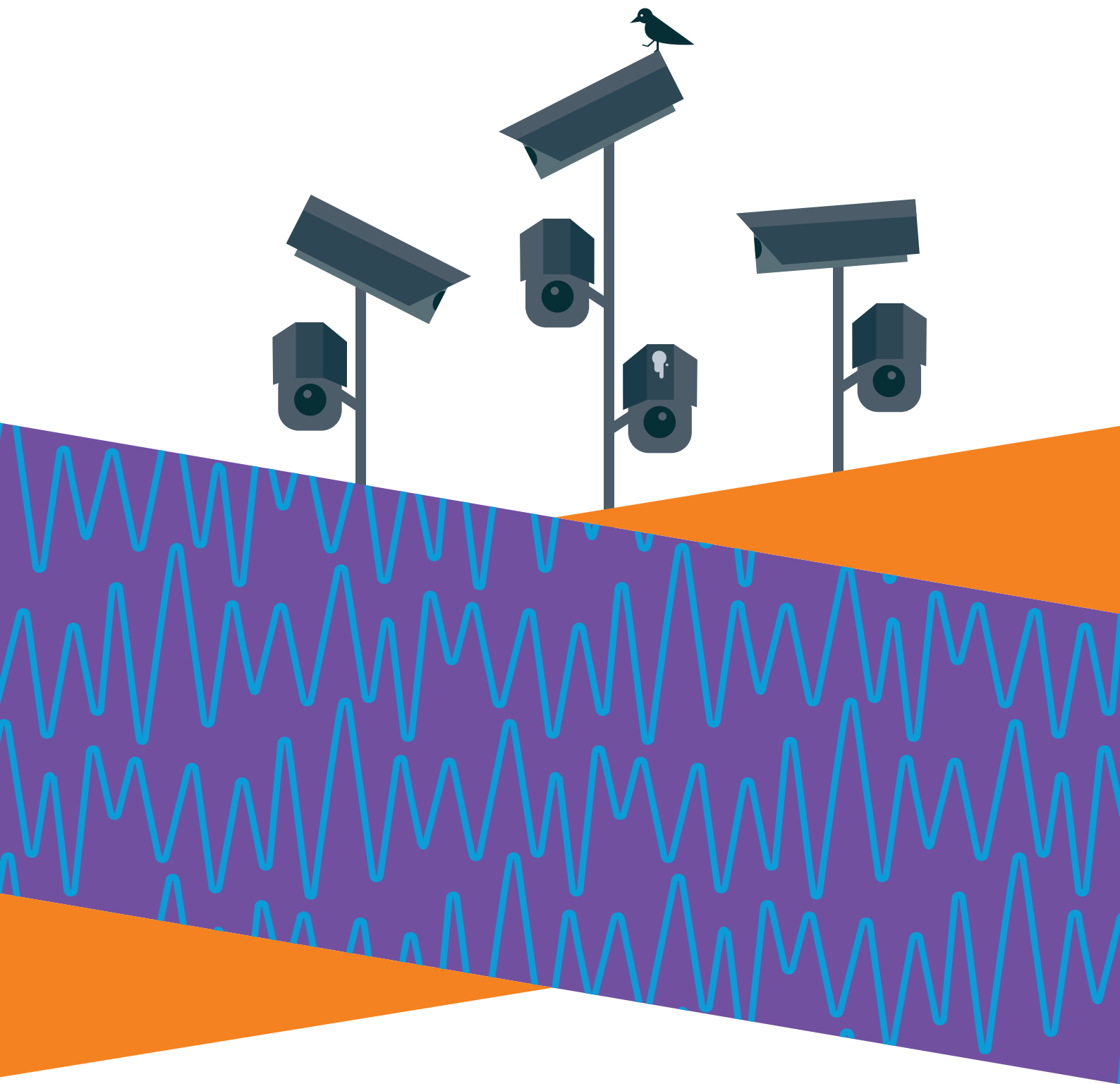
Gestaltung und Illustrationen: Hannes Saxer, Bern

Druck: Jordi AG – Das Medienhaus, Belp

TA-SWISS – Stiftung für Technologiefolgen-Abschätzung

Neue Technologien bieten oftmals entscheidende Verbesserungen für die Lebensqualität. Zugleich bergen sie mitunter aber auch neuartige Risiken, deren Folgen sich nicht immer von vornherein absehen lassen. Die Stiftung für Technologiefolgen-Abschätzung TA-SWISS untersucht die Chancen und Risiken neuer technologischer Entwicklungen in den Bereichen «Biotechnologie und Medizin», «Informationsgesellschaft» und «Mobilität / Energie / Klima». Ihre Studien richten sich sowohl an die Entscheidungstragenden in Politik und Wirtschaft als auch an die breite Öffentlichkeit. Ausserdem fördert TA-SWISS den Informations- und Meinungsaustausch zwischen Fachleuten aus Wissenschaft, Wirtschaft, Politik und der breiten Bevölkerung durch Mitwirkungsverfahren. Die Studien von TA-SWISS sollen möglichst sachliche, unabhängige und breit abgestützte Informationen zu den Chancen und Risiken neuer Technologien vermitteln. Deshalb werden sie in Absprache mit themenspezifisch zusammengesetzten Expertengruppen erarbeitet. Durch die Fachkompetenz ihrer Mitglieder decken diese Begleitgruppen eine breite Palette von Aspekten der untersuchten Thematik ab.

Die Stiftung TA-SWISS ist ein Kompetenzzentrum der Akademien der Wissenschaften Schweiz.



TA-SWISS
Stiftung für Technologiefolgen-Abschätzung
Brunngasse 36
CH-3011 Bern
info@ta-swiss.ch
www.ta-swiss.ch

mitglied der
 akademien der
wissenschaften schweiz

