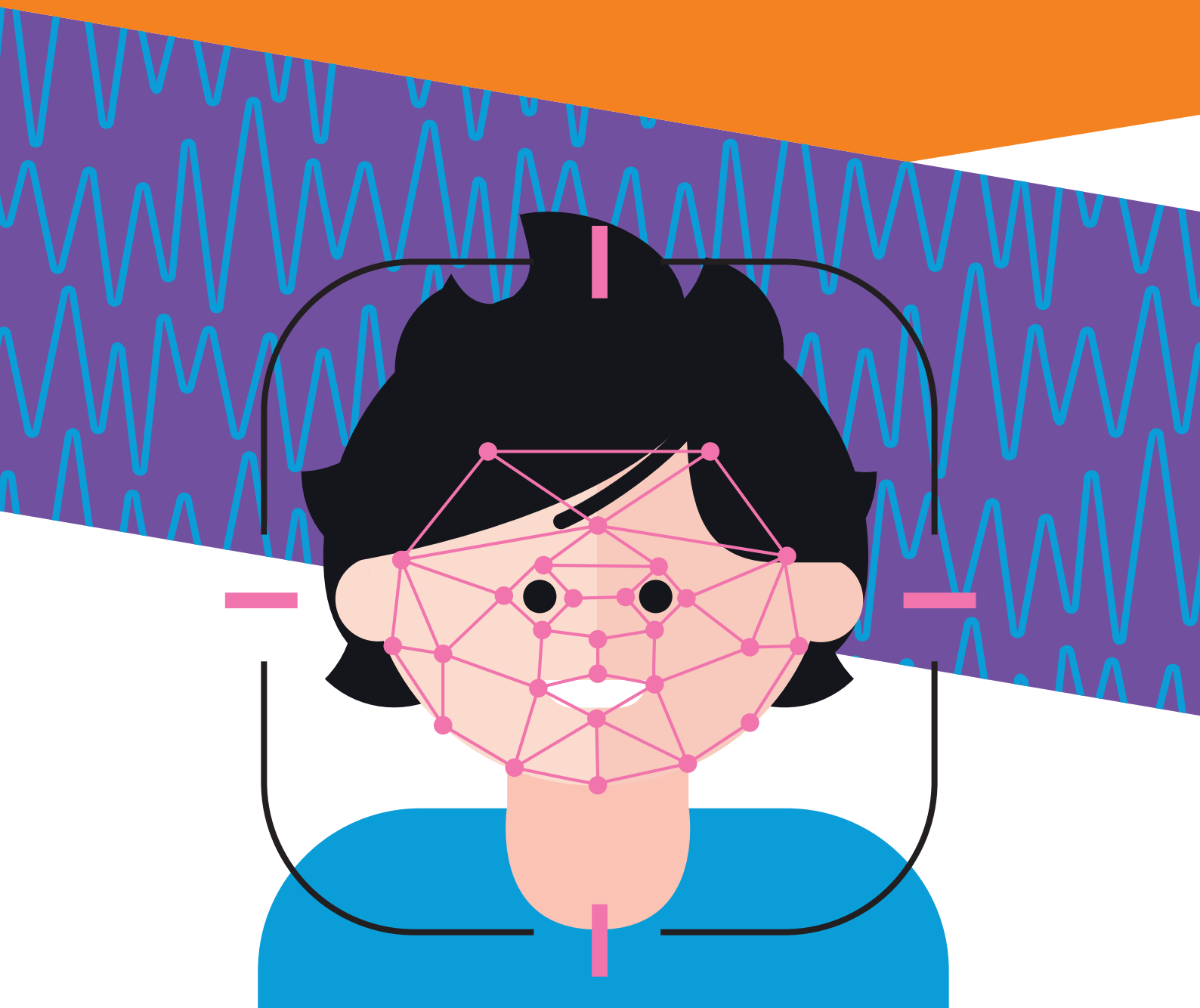
Observed at every turn

Summary of the TA-SWISS study: "Automated speech, speaker and facial recognition systems: technical, legal and societal challenges"



TA-SWISS, the Foundation for Technology Assessment and a centre for excellence of the Swiss Academies of Arts and Sciences, deals with the opportunities and risks of new technologies.

This abridged version is based on a scientific study carried out on behalf of TA-SWISS by an interdisciplinary project team with members from the Fraunhofer Institute for Systems and Innovation Research ISI in Karlsruhe, Germany, and the University of Fribourg i.Ue, Switzerland; Murat Karaboga was responsible for overall project leadership. The abridged version presents the most important results and conclusions of the study in condensed form and is aimed at a broad audience.

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

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

TA-SWISS, Foundation for Technology Assessment (Ed.)
vdf Hochschulverlag an der ETH Zürich, 2022.
ISBN: 978-3-7281-4140-8
Also available in open access: www.vdf.ch

This abridged version can be downloaded at no cost: www.ta-swiss.ch

Speech, speaker and facial recognition technologies: a brief introduction

Cameras installed in public spaces gather vast amounts of data to feed speech, speaker and facial recognition systems. Although these detection technologies have the potential to promote both individual and community safety by helping the authorities to find missing persons or monitor suspects in a crime, automated identification systems can easily be misused, and they put citizens under pressure to adjust their behaviours to conform with societal norms. This has consequences for individuals – but also for democracy itself.

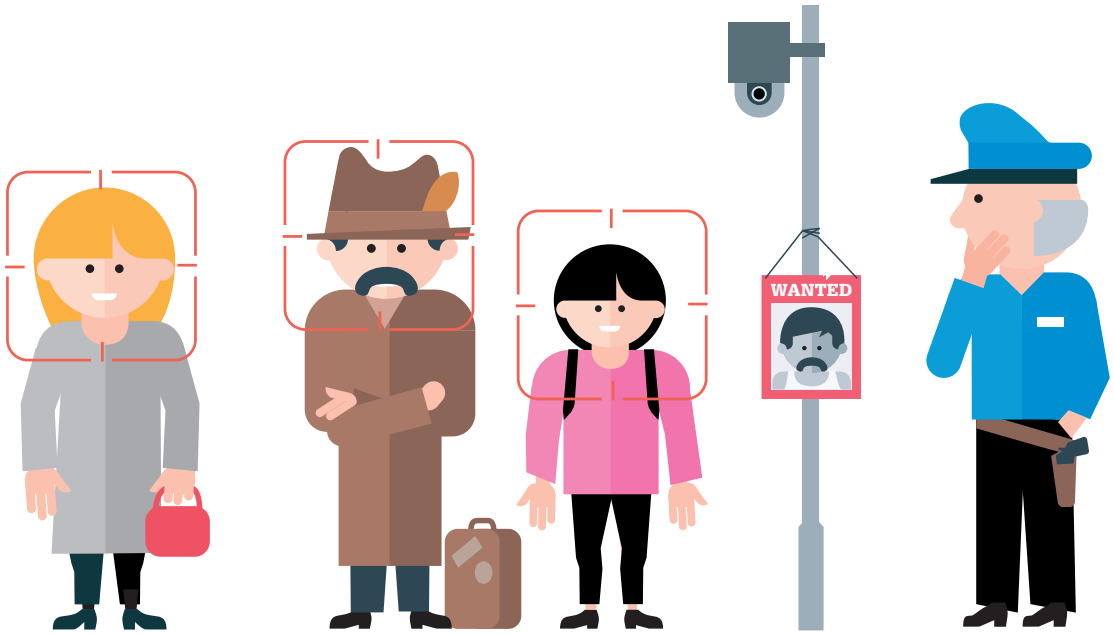
In recent years, smart speakers have become more common in Swiss homes, and facial recognition technology is often used to unlock smartphones. While speech, speaker and facial recognition simplifies many everyday activities, the underlying technology is not yet completely reliable: for instance, facial recognition systems fail to identify women and people of colour with the same precision as they identify white men. If the advances continue at the current pace, however, it is to be expected that a high level of technical reliability will be attained in the coming years. The danger of false matches is greater if voice and facial data are used to draw conclusions about an individual's emotional state or their physical or mental health.

Opportunities …

With smart speakers, multiple devices can be operated using voice commands, making it unnecessary to have different remote controls for different devices. And "smart assistants" simplify daily life by saving dates in our calendars or adjusting our stereos and lighting – and leaving our hands free for other tasks.

In medical care, it is hoped that speech, speaker and facial recognition systems will aid the early detection of serious conditions such as Parkinson's, Alzheimer's, depression and burnout. In addition, developers are working on programmes capable of using facial recognition to identify rare diseases that are seldom seen by most doctors.

The possibility of shadowing and observing criminal suspects means that their machinations might be discovered before a crime has been committed. Speech, speaker and facial recognition therefore can help to promote community safety.



… and risks

The data generated by speech, speaker and facial recognition systems are so-called biometric data. These data reveal a great deal about an individual, and they change very little after a person reaches adulthood. Once biometric data have been breached, they remain permanently compromised.

The use of speech, speaker and facial recognition technology endangers our privacy, as these systems erode our anonymity in various ways. For instance, the smart speakers that are often placed in living rooms, and thus at the very heart of our homes, are constantly listening in on us and can record potentially confidential information.

As such, speech, speaker and facial recognition technologies tend to accentuate the power imbalance between citizens and the government or private companies, especially if the former are unaware of the kinds of data public authorities or businesses have access to.

Key recommendations

It is recommended that particularly problematic uses of speech, speaker and facial recognition technologies be banned. This includes automated real-time surveillance systems as well as smartglasses and other hidden devices that can secretly observe people. Attention monitoring at schools should also be prohibited.

The use of speech, speaker and facial recognition systems by the police or other public authorities must be regulated through unambiguous legislation that defines rule-of-law safeguards and stipulates a mandatory review of whether a technology is necessary.

Training and continuing education programmes should be created for all persons who use detection technologies. Moreover, a support service should be set up to aid people who wish to protect themselves from the dangers of speech, speaker and facial recognition systems, and who wish to understand and assert their rights.

Finally, the advantages and disadvantages of speech, speaker and facial recognition systems and their lawful use must be made a topic of frank debate in society.

> The TA-SWISS study on speech, speaker and facial recognition technologies was conducted by a project group with members from the Fraunhofer Institute for Systems and Innovation Research ISI in Karlsruhe, Germany, and the University of Fribourg i.Ue, Switzerland; Murat Karaboga was responsible for overall project leadership. The study's methaodology draws on extensive literature reviews and analyses of media reports as well as several focus group discussions held with citizens and a representative online survey with 1000 respondents.

Vanishing anonymity

KITT, the souped-up Trans Am in the American action crime series Knight Rider from the early 1980s, carried out his driver's voice commands. And in early science fiction films, protagonists just needed to use a face scanner to gain access to the cockpit of a spaceship. In the intervening years, speech, speaker and facial recognition technologies have become a part of everyday, real-world life.

Technological developments often have unexpected impacts on society. For instance, in June 2022, the administrative court in Göttingen, Germany, granted the parents of a four-year-old girl named Alexa permission to change their daughter's name. The little girl was subject to so much teasing that she was diagnosed with emotional distress. 'Alexa' is also the name of Amazon's virtual assistant that controls various devices via voice command, and the German girl fell victim to predictably stupid commands and irritating jokes.

Whether sound, family tradition or current trends matters most, parents-to-be put a lot of thought into choosing the right name for their baby. But while a name may be changed later in life if circumstances

so require, it is an entirely different matter when it comes to our voices, the shape of our face or our manner of speaking. These traits hardly change at all after we reach adulthood. And once the characteristics that are inextricably linked to an individual have been measured, digitised and stored as biometric data, these key physical traits can be retrieved for later electronic processing.

It has now been over fifty years since speech, speaker and facial recognition technologies began emerging. Although the technological details are very different, all systems have one thing in common: they process biometric data that reveal a great deal about an individual.

A helpful 'dragon' and other services

Already in the early 1960s, research was conducted on machine processing of language, albeit with little success. Computers were too inefficient. But with the advent of mainframe computers, performance levels increased, paving the way for IBM, Philips and Dragon Systems to develop commercial dictation programmes that became standard office tools as of the 1990s.

Dragon proved to be the best dictation programme that functions independently of a specific operating system. After a short training session to introduce the software to a speaker, the programme attains a detection rate of up to ninety-eight percent; Dragon is more efficient when used to dictate technical language with fixed terminology than for a language with literary expressions and a varied vocabulary.

Siri, Google Assistant and Alexa are all linked to specific platforms that are integrated in Apple's and Android's operating systems and Amazon's smart speakers. In particular, the affordable Amazon speakers were instrumental in promoting the use of speech and speaker recognition: in 2017, Amazon smart speakers made up about eighty percent of worldwide sales. To be sure, the market has grown more diverse since then.

Smartphones also react to commands such as "Alexa, set the alarm for seven" or "Hey Siri, what's the forecast for Zurich". If a smartphone is connected to other devices in a smart home, they, too, can be operated using voice commands, and orders like "Hey Siri, activate the printer" or "Alexa, turn on the coffee machine" will also be carried out.

Face made of data points

Research into facial recognition technology also dates back to the 1960s, but real advances came later than with speech recognition. The first schematic attempts to map hairlines, eyes and noses in 1964 yielded incorrect results if a person's head was tilted or the lighting conditions were poor. The technology improved first in 1970, when higher-performance computers made it possible to assess additional features such as lip shape or hair colour. But the real breakthrough came in the early 1990s with the development of an algorithm that – rather than measuring anatomical features of individual faces – conducts a statistical analysis of the main components of a large dataset of facial images. The problem with this approach is that the recognition rates attained by various research groups were at first difficult to compare, as each group worked with their own image databases. Matters improved only after the United States Department of Defense began to build up a large database to compare the different algorithms as part of a facial recognition programme launched in 1993. Since then, all facial recognition systems on the market have competed in the annual "Face Recognition Vendor Test"; in 2020, ninety-nine manufacturers sent a total of one hundred eighty-nine algorithms to the competition.

The terrorist attacks on 11 September 2001 mark a major turning point in the use of facial recognition technology. While still reeling from the shock of the events, the US government enforced a rule allowing only persons holding a biometric passport to enter the country, even for short stays. The European Union and Switzerland complied with the requirement and, with the introduction of standardised biometric facial images in 2006, laid the cornerstone for state-run facial recognition practices.

Identification and verification of speech and speakers

As a rule, speech, speaker and facial recognition systems are used to automatically confirm the identity of an individual and, under circumstances, grant him or her access to a specific service. Speech recognition technologies, by contrast, are generally used for voice commands that replace typing a command on a smartphone or computer.

It is important to differentiate between identification and verification. Verification of an individual's identity is a simpler process in which characteristic features from a sound or image file are compared to a sample recording of a specific person. If the new sound or image file matches the sample, the person's identity is confirmed. Identification by contrast is a process designed to discover who a person is by comparing a sound or image file with a large number of recordings or images in a database.

Another important point is that speech recognition and speaker recognition are not the same. A system that detects words – that 'understands' their meaning – will not necessarily be capable of identifying a voice or the person speaking. Indeed, the purpose of speech recognition technologies is to recognise the content of a statement in order to carry out a command (what is being said), whereas speaker recognition systems are designed to identify individuals on the basis of biometric parameters such as pitch and timbre of their voice (who is speaking). The latter technology is used at banks and health insurance companies to determine whether the person on the line is who they say they are.

Biometric data reveal much more than the eye sees

Simple images or sound files are relatively easy to manipulate, and they are far from constituting biometric data. The latter are modelled and calculated in a complex process. In the case of the three-dimensional methods used in the facial recognition technologies of major smartphone manufacturers, 30 000 dots of invisible infrared light are projected onto the user's face during the scanning process. The smartphone processes these data to create a 3D model that is then used to identify characteristic facial features and traits. Both the infrared image and 3D model are converted into a mathematical formula and stored on the device's chip. Each time a phone is unlocked, the formula is retrieved and the algorithm runs.

In speaker recognition, too, an algorithm records and processes several thousand features – most of which are inaudible to the human ear – that include much more than pitch, intonation, respiratory rate and rhythm. A software programme processes these data to create a so-called voiceprint, a type of acoustic fingerprint.

Because biometric data are so closely associated with an individual and reveal so much about personal characteristics, they are considered to be particularly worthy of protection. In a guideline issued by Swiss data protection officials, biometric data are defined as "unique, specific physical features of an individual that – at least theoretically – can be assigned to this individual with almost one-hundred percent certainty, always and everywhere."

The use of speech, speaker and facial recognition technologies by private actors generally represents an infringement of data protection law. Indeed, it is difficult for anyone wishing to process these data to prove an overriding interest, as biometric data are particularly sensitive. Moreover, obtaining consent from data subjects is often virtually impossible, as devices must first access the desired data in order to function. In situations such as covert identification processes, consent is not even sought, which is in violation of data protection law.

A broader view of data protection

The automated recognition of individuals on the basis of biometric data has drawn criticism from the start; in particular, it was argued that the results of automated detection systems are too often erroneous and subject to bias. For example, facial recognition systems detect women and people of colour less accurately than white men – not least due to the fact that, for years, the images used to train algorithms in databases were disproportionately of white men. The analysis of sound files is also not immune to error, as a single factor such as poor sound quality increases the likelihood of a false match.

When biometric identification systems were first being developed in the 1990s, protecting data was the greatest concern; the main point of criticism was the risk of losing privacy and personal freedoms if an individual's every movement and action were automatically monitored. Over time, however, the focus on individual rights and freedoms came to be regarded as too narrow. Experts now increasingly advocate for a comprehensive view of the impacts of technologies that could be used for surveillance. Indeed, the use of such devices not only poses a threat to the privacy of individuals: other fundamental rights are also at risk, including freedom of assembly and freedom of expression, which are essential for a functioning democracy.

Pressure to conform endangers basic rights and democracy itself

People who fear they are always being watched tend to adapt their behaviours and to self-censor what they say. This pressure to conform is at variance with legislation (including Switzerland's Federal Constitution) that both upholds freedom of expression and rejects discrimination against people on the basis of their identity, their lifestyle or their personal convictions.

From a legal point of view, data protection has been, and remains, a sticking point. Because biometric data reveal information about an individual's unique characteristics and are thus considered to be particularly worthy of protection, the processing of these data potentially endanger basic individual rights. At particular risk are fundamental rights such as the right to privacy and the protection against misuse of personal data that are guaranteed by the Federal Constitution. Handling biometric data is also made more difficult because the processes to anonymise data are practically incompatible with most speech, speaker and facial recognition systems. Moreover, the aforementioned pressure to conform endangers the fundamental freedoms of expression, assembly and association guaranteed by the Federal Constitution.

With regard to biometric data, one of the key principles is proportionality, meaning that as little data as possible should be collected for a specific task. So-called purpose limitation is also important, i.e. data may only be processed for the purpose they were collected for. It is also important to review the appropriateness of a technology: use of a technology should be deemed lawful only if it is appropriate for achieving the stated objective. Transparency is also a key requirement: it is crucial that people know when their biometric information is being collected – and they must be given the opportunity to consent or refuse to take part, with no negative repercussions. Lastly, secure data storage must be guaranteed in order to prevent hackers or other persons from gaining unlawful access. The use of such technologies therefore requires legislation that has been framed clearly and precisely. In addition to naming the purpose of the technology, such a law must also define restrictions and, in particular, limit data processing to what is strictly necessary.

Different priorities in the world of research
In China, facial recognition technologies have been an intense research focus for quite some time. In 2020, Chinese researchers published at least twice as many scientific articles on the topic as their counterparts in the US or India. Publications from all of Europe, Switzerland included, equal about a third of Chinese output. The situation is a little different for speech recognition technologies, where the United States is a clear leader, followed by China, which began intensifying its efforts in this area in 2018. The importance of speech, speaker and facial recognition systems in China compared to other research areas is very evident.

When technology eavesdrops

Smart speakers have become a common feature in many offices and homes. Instead of typing e-mails and letters, we dictate them to our computers. We make appointments and set alarms by talking to our smartphones, and we give voice commands to regulate room temperature and other features in a smart home.

The advantages of using a voice command to replace typing it or otherwise entering it into a device are obvious: smart speakers replace a multitude of remote controls. In addition, the various devices are easier to handle, which also makes everyday life more manageable for people with a physical disability.

Smart speakers, many listeners

From a technical point of view, smart speakers consist of at least one microphone and loudspeaker as well as a connection to a provider via WLAN and internet; the provider offers a range of cloud-based services and functions to users, but also analyses the recordings to improve services. As such, the data collected are stored in different locations: in the smart speaker itself and on the provider's server; if the loudspeaker is connected to a smartphone, the recordings will also be stored there.

To ensure that a smart speaker will 'hear' the activation word for a command, it is constantly in reception mode. As soon as the device registers a command, a recording begins, which is then transmitted to the manufacturer's server, where it is analysed. Speech recognition technology processes the command and sends the result back to the user via an automated synthetic voice output.

Many people are either unaware of – or at least rarely take advantage of – the following possibility: audio recordings can be deleted from numerous devices via an app or a website. From a data protection perspective, this is welcome news. Indeed, many requests or commands are sensitive – making a doctor's appointment via a voice command, for instance, or arranging a meeting with a divorce lawyer. In addition to the voice command itself, smart speakers also record other data such as the time and date, making it possible to draw conclusions about a specific individual's habits. If others are present, their voices are also recorded, which in

turn offers clues as to the identity of the people who belong to a household or a circle of friends. Moreover, spoken commands and requests are not the only revealing data collected – the sound and pitch of a voice are also significant. In particular, information can be gleaned about an individual's physical condition – if someone has a cold, for instance, or has had too much to drink. An individual's emotional state can also be detected on the basis of merry laughter or dejected sighs.

Speaker authentication at the automated welcome desk

The Swiss post office and several Swiss banks use voice recordings to confirm the identity of their clients. It is believed these processes will lead to greater security and efficiency. At Migros Bank, for instance, call duration was reduced by twenty percent thanks to speaker authentication. Thus far, the technology has been applied primarily for processes in telephone banking. Future use in advisory services or virtual assistance systems is also plausible.

Before a recording begins, clients are asked whether they agree to having their call saved and used for authentication purposes in future calls. At PostFinance, bank customers also have the option of activating or rejecting speaker recognition on the web portal.

Whether a person's identity can be confirmed with complete accuracy based on his or her voice is open to debate. Some people believe a voice is extremely individual and thus view biometric speaker recognition as secure. However, others point out that journalists and hackers have already succeeded in outsmarting acoustic access barriers by combining sound snippets from YouTube videos and computer programmes – a practice also used for the production of deceptively authentic-sounding fakes, so-called deep fakes.

What is certain is that authentication based on biometric data is secure only if the data remain protected. Once they have been breached, they can no longer be used to unequivocally verify the identity of an individual. This is because voice and facial features are closely and permanently linked to a specific person and – unlike a password – cannot be changed easily.

Blocking criminals

Criminals go to great lengths to gain access to sensitive data via speaker recognition technologies. For this reason, experts recommend that banks and other businesses dealing with sensitive information require a password in addition to speaker recognition. The two-factor process increases security.

The growing use of smart speakers has increasingly made them a target for hackers, who may try to obtain sensitive data by using a voice recording of an individual to trick the authentication mechanism in their smart speaker.

Researchers are currently addressing the security risks in smart speakers. One approach is to remove features from the recording that are not needed to interact with the smart speakers. To prevent unauthorised access to the data, other researchers are developing network analysis programmes that recognise when sound files are transmitted on the internet. This warns users when their activation word has been triggered.

Moderation in Switzerland
In 2020, Amazon led the market with twenty-two percent of all smart speakers sold worldwide. Google, with seventeen percent, was a close second. In 2018, eighteen percent of the US population was using smart speakers, and in Germany ten percent. In comparison, the Swiss practise restraint: in 2018, just one percent of the Swiss population used smart speakers; one year later, it was three percent. A representative TA-SWISS study conducted in October 2021 revealed that sixty-three percent of all homes have no smart speakers. Almost half of all smart speaker users had bought the device within the past year. Only nine percent said they had owned their device for more than three years. Among the persons surveyed who do not have smart speakers, forty-one percent said they had no plans to buy them in future – mainly because they see no good use for these devices and are concerned about data protection.

Showing our true face to gain access

During the coronavirus pandemic, the use of contactless payment and access systems increased significantly, and using technology like Face ID to unlock our phones is now very common. While facial recognition systems on a personal device may be practical, they lose their appeal when used at football matches or rock concerts; for individual people, they can even have adverse consequences.

Public authorities, first and foremost the police and custom officials, have been using facial recognition technologies for quite some time. However, gaining precise information about how and to what end the technology is used is difficult, especially as police use of facial recognition tools is often covert.

In the 1990s, the systems used by US police forces were notable mainly for their poor execution. In 2003, not one single match was registered in a long-term, large-scale test at the Boston airport. In Florida, too, a facial-recognition software programme proved highly faulty: on some days, the system yielded nothing but false-positives. A test conducted by the German Federal Criminal Police Office in 2006 at the Mainz train station attained somewhat better results, with a hit rate of thirty percent – which, however, fell drastically short of the targeted eighty percent.

Limited use in Switzerland

In Switzerland, facial detection technology is used mainly at Zurich Airport, and this on a voluntary basis. Following a six-month pilot project starting in autumn 2017, air travellers have had the option of choosing the automated face detection system since 2020. Signs lead travellers to the facial recognition system, but it is still possible to have documents checked by a person at the control desk.

Apart from Zurich Airport, the police in Aargau and St. Gallen are the only other known instances of facial detection technology in operation in Switzerland. Other police forces are testing similar systems, or they employ humans for facial recognition work. In Basel-Stadt, the cantonal police have purchased seven Teslas, each of which is equipped with eight

cameras. Although not set up for facial recognition, connecting the vehicles to detection software would be easy.

Through discussions with the two cantonal police forces using facial recognition technology, the TA-SWISS project group was able to ascertain that they are endeavouring to guarantee secure, lawful and ethically responsible use of the technology. For instance, both forces have conducted a data protection impact assessment (DPIA). And to ensure security, data are stored on servers that are not connected to the internet or other networks.

Nevertheless, the TA-SWISS study also identified weaknesses in their use of facial recognition systems. Neither police force published the results of the DPIA, and both cantonal police authorities have chosen not to have the performance and security of their systems independently controlled. Who has access to which data is also unclear. In addition, both cantonal police forces take the stance that it is unnecessary to inform individuals when their data have been processed in police facial recognition systems. Their reasoning is that it makes no fundamental difference whether a human or a software programme is responsible for scanning images.

Will real-time facial recognition lead to mass surveillance?

In Switzerland, no real-time mass surveillance systems are used. Indeed, they would be unlawful, as they endanger basic rights such as freedom of assembly. The principle of proportionality would also be violated if, for instance, everyone taking part in a political demonstration was under surveillance simply so the police could catch a few individuals who might commit acts of vandalism. Despite not being used, mass surveillance forms the topic of much contentious debate in Switzerland. The coronavirus pandemic did little to assuage the fears, as various countries used facial recognition systems to monitor compliance with Covid-related restrictions.

Participants in the focus groups conducted for the TA-SWISS study called for state-operated mass surveillance to be prohibited, because the state would otherwise have too much power over citizens, leading to a loss of trust in government. The participants also believe that, once introduced, mass surveillance would be unstoppable: at first, they say, facial recognition systems would be used only to solve serious crimes, but after proving successful, the technology would soon find application for misdemeanours such as pickpocketing, until the population was accustomed to constant and blanket surveillance. In the end, Switzerland would be confronted with conditions like those in China, where the social credit system is often cited as a cautionary tale: people caught jaywalking or violating other state rules have to reckon with sanctions, such as being disadvantaged in job applications or when looking for housing.

Two-pronged technical approach to fighting racism

Although European countries tend to use facial recognition systems sparingly, certain private actors have less compunction. For instance, Italian football clubs began considering the use of facial and speech recognition systems to prevent racism after several matches were cancelled because fans were making ape sounds, yelling Nazi slogans and otherwise casting xenophobic insults at dark-skinned players. With speech recognition technology, it would be possible to understand the content of racist chants, and then use facial recognition software to identify the offenders. To date, no such systems have been introduced to Italian stadiums. However, Luigi De Siervo, director general of Italy's top football league, stated in the autumn of 2019 that he was prepared to start using facial recognition systems to "catch the people who are ruining this wonderful sport, one by one'.

In Switzerland, no legal basis exists for speech and facial recognition systems in sports stadiums or at other private events. Nevertheless, a slight majority of respondents surveyed in the study approved of facial recognition technology at sports stadiums, with the highest approval ratings for the technology found in this context; as a rule, however, the acceptance of facial recognition systems is low.

Clear directives needed for police use of facial recognition technology

In the representative TA-SWISS study, police use of facial recognition was not rejected outright: one third of all respondents were in favour of using it, one third said they were not familiar enough with the subject and just under one third rejected use of the technology. Those who voiced approval believe the technology primarily lends itself to seeking missing persons, followed by counter-terrorism efforts. Respondents who reject facial recognition technology do so mainly because they fear these data will be misused, because the technology would have a major impact on public life and because it could lead to unwarranted mass surveillance. All respondents said that only authorised staff should be allowed to use facial recognition technologies and that these individuals should also be required to log and communicate every use. In addition, the participants in the survey believe an appropriate legal basis is necessary and that independent experts should be consulted to regularly monitor and evaluate the technology.

When technology reveals our innermost secrets

Experienced, empathetic doctors can learn a great deal about their patients' health by observing changes in skin tone, facial expression or voice quality. It is now becoming apparent that speech, speaker and facial recognition systems will one day be capable of detecting diseases even earlier than medical professionals. However, with regard to understanding emotions, the technology faces greater obstacles.

"This app can really help transgender people train their voices and track their progress," is one comment about the app Voice Pitch Analyzer. "Hi, I'm FtM (female to male), and I use the app because it's so interesting to watch how your voice changes when you start hormone therapy," is another person's comment. Voice Pitch Analyzer created by Purr Programming is a free app that can be downloaded and used to analyse a voice – one of the features that generally reveals whether a speaker is a man or a woman. Apps to analyse and train voices are designed to help trans people align their speaking voice as closely as possible to the pitch of the gender they identify with.

Whether the aim is to track physical fitness, sleep quality, pulse or ovulation – there are now numerous apps available to monitor our own bodies. In future, these programmes might increasingly rely on speech, speaker and facial recognition technologies to assess a person's physical state. And because people nowadays tend to consult "Dr Google" before going to their general practitioner, the Google search engine would have access to even more powerful analytical tools.

Technology that seems harmless in a voice analysis app on a computer or smartphone might become problematic with more sophisticated versions. Indeed, the advances in diagnostic tools go hand in hand with the danger that non-specialists will encounter problems when using them to monitor their health via their computer's camera and microphone. At present, however, powerful diagnostic tools based on speaker or facial recognition technologies are reserved for medical professionals.

In action for medicine

When we speak, our brains regulate the interaction processes of up to a hundred muscles. The quality of our voices can therefore be indicative of numerous illnesses. In such cases, sounds that the human ear easily misses are potentially revealing. For instance, researchers examined the 'aah' sound made by healthy people and those suffering from Parkinson's disease, identifying ten acoustic characteristics that can be used to diagnose the disease with an accuracy of almost ninety-nine percent. As another example, word choice or difficulty in finding words can be an indication of Alzheimer's. In a controlled environment, speech and speaker recognition tools detect the disease with a recognition rate of eighty to ninety percent.

Since the start of the coronavirus pandemic, researchers have been studying the characteristic Covid cough in order to promote early diagnosis. It is also possible that heart conditions could be detected on the basis of a person's voice, as certain patterns of voice frequency are associated with severe disease of the coronary arteries.

Lastly, our speaking voices mirror our mental health. Tempo, rhythm, pitch and volume can reveal whether we are excited, afraid, depressed or manic. Speech recognition is now a standard tool for diagnosing depression: a monotone, somewhat higher pitch compared to the regular speaking voice can indicate that a person is suicidal. Linguistic features have also been identified to help diagnose autism as well as attention-deficit/hyperactivity disorder.

Facial recognition in medical care is less common than speech and speaker recognition, but potential applications for the technology are nonetheless being explored. In the EU project SEMEOTICONS, a type of mirror with various sensors and cameras was developed that, in addition to recognising psychological traits, was able to detect a person's physical and nutritional state. It also analysed the colour of skin and mucous membranes as well as the distribution of subcutaneous fatty tissue and

perspiration patterns. A system called DeepGestalt developed at the University of Bonn is programmed to detect rare diseases or genetic defects based on a photograph of an individual's face. The software programme compares the facial expression on the picture to images with numerous other people who have been diagnosed with a specific disease; the database comprises 17,000 photographs of more than 200 complex diseases. Several other research teams are working on additional systems with the aim of using facial images to detect rare diseases.

One difficulty in using speech, speaker and facial recognition systems in medical care lies in creating a reliable database. Currently, there are very few databases for speaker analysis. Moreover, some experts doubt the reliability and informative value of speaker traits. Too often, spoken language is distorted because the speaker has a cold or is suffering from allergies. Culture and background will also influence how loudly or quickly a person speaks. It is readily apparent that there is great need for more research into the use of speech, speaker and facial recognition technologies in the world of medicine.

When using detection technologies in medical care, it is important that specific requirements regarding secure handling and storage of patient data also comply with physician-patient confidentiality. Apps that could be used for self-diagnosis should be deemed medical products. Because such applications are at present very rare, regulatory measures are not to be expected in the near future.

Sounding out our emotional world

Another potential use for speech, speaker and facial recognition technologies would be in job application procedures. For example, a software programme designed to preselect especially suitable candidates could speed up the selection process. Technology could also be used to assess an individual's overall demeanour and thus help to determine whether he or she is a good match for the advertised position.

It is also plausible that private insurance companies would be interested in using the technology to offer their clients personalised services. The insurance industry is similar to medical care in that here, too, information about a person's physical fitness or other physical traits gleaned through facial and voice data could be exploited. Generally speaking, emotion detection could meet with great interest

in various circles: banks might hope for marketing advantages, police forces could be tempted to use the technology as a lie detector – as has already been tested in the EU-funded iBorderCtrl project – and sport stadium operators might see an opportunity to identify violent fans before they have a chance to cause damage.

In such situations, data subjects must grant their consent to having their faces or voices analysed, as is presently the case with psychological tests or handwriting analyses. Moreover, this consent would have to be voluntary – a requirement that is difficult to fulfil considering the power asymmetry between employer and employee, or insurance company and persons insured. And if a technology were to be used secretly during a job interview or when taking out an insurance policy, this would amount to an invasion of personal privacy; in particular, it would violate the principle of good faith.

Many experts maintain a general scepticism towards emotion detection technology. They doubt whether a software programme could be capable of reliably interpreting feelings, which are expressed very differently from person to person. The mere fact that our cultural background has a major influence on how we show our feelings is proof of the difficulties surrounding emotion detection.

Reservations about emotion detection
In Switzerland, the medical use of speech, speaker and facial recognition technology as well as its use for emotion detection or attention monitoring in schools meets with considerable reservations. Just under a quarter of all respondents would have no concerns if the technology were used to diagnose physical illnesses, with thirty-seven percent voicing reservations; the main concern of the latter group is that health insurance companies would have access to the data generated. With regard to emotion detection, sixty-five percent worry about the accuracy of the software. And sixty-two percent of the respondents believe that emotions are too complex to be accurately detected by a software programme.

Banishing distractions at school

In many societies, performance at school is monitored very closely, as a nation's economic success depends on educating the next generation. Various countries – especially China and in the English-speaking world – use technologies to monitor attention levels of schoolchildren. In the recent past, US schools have invested 2.7 billion dollars annually in surveillance and security products with the aim of being prepared for mass shootings and other acts

of violence. Although there is currently no evidence that the attention levels of schoolchildren in the United States are being monitored, once video systems are installed, they can be upgraded to perform attention monitoring with relatively little effort and at the modest price of just under 200,000 Swiss francs per school.

At some schools in China, not only children but also teachers are monitored. Neither students nor parents are asked to grant consent. The monitoring systems make various recordings – for instance, how long children keep their eyes on the blackboard – and calculate attention scores that teachers and parents have access to. Using this kind of technology is designed to improve scholastic achievement: not only are all external distractions eliminated, it is believed that the individual strengths and weaknesses of learners can be better accommodated by using attention monitoring. However, psychological research has demonstrated that too much monitoring can also cause stress, which is an obstacle to learning. In addition, numerous experts cast doubt on whether analysing facial expressions can deliver reliable information about a person's attention levels.

In principle, emotions can be analysed on all systems and platforms where image data of school children or students are generated. The coronavirus pandemic, which has driven the use of online learning and communication platforms such as Moodle, Google Classroom or Zoom, has also helped to lay the foundation for later introducing emotion detection or attention monitoring – especially as these kinds of software are already being used for examination supervision at some schools.

Not everyone wants to have their name known

Speech, speaker and facial recognition technologies are getting closer to making universal identification a reality. Equipped with the right devices, everyone would be able to identify every single person they encounter. Alphabet took an initial step in this direction in 2014 when the company launched its Google Glass smartglasses. The idea behind the miniature computer in front of our eyes is to superimpose all kinds of information about the environment onto the glasses, including data on buildings or other objects in the surroundings. The smartglasses access the necessary data from the internet. While it is true that the company explicitly excluded facial recognition technology in the product, data protection experts pointed out that it would be relatively easy to connect the smartglasses to existing facial recognition programmes – and various companies were betting that Alphabet would move away from its self-imposed ban on facial recognition technology.

Databases for identifying private individuals are, however, already available; the company NameTag alone collected several million facial images and announced it would continue to mine for more portraits in social networks. Due to low demand – and not least to persistent public criticism – Alphabet stopped selling its smartglasses to private individuals in 2015. However, powerful new smartglasses have long since entered the market, including Ray-Ban Stories, which were developed in partnership with Facebook. It would be easy for hackers to add a facial recognition function to the glasses. If this were to happen, random passers-by could be identified by tapping into an image database. Anonymity in the public sphere, which is already dwindling due to the ubiquitous use of smartphones, would be a thing of the past with the advent of universal identification technologies.

The Clearview AI scandal offers proof that data-protection related fears linked to universal identification are justified. Established in 2017, the US company mined billions of portrait photos on the internet – especially from social networks – to create its own, extensive image database, which it then linked to its inhouse search engine. The company offered its search-engine services to governments and private

companies throughout the world, albeit without informing the citizens of these countries or even the owners of the images or the persons depicted on them. Various lawsuits were filed against the company, with the result that only security agencies have been granted access to the search engine since 2020. The company returned to the headlines with the war in Ukraine, when it was discovered that Ukrainian authorities were using the software to identify fallen or captured Russian soldiers and inform their relatives. The Taliban in Afghanistan are also making use of facial recognition technology. When they seized power, they not only confiscated vehicles and weapons left behind by the Western armies – they also took possession of datasets containing millions of fingerprints and facial images. The fundamentalist group is now hunting 'traitors' and other individuals who have 'collaborated' with Western organisations.

Attention monitoring and universal identification are unwanted

In the representative survey, fifty-six percent of all respondents absolutely rejected the use of speech, speaker and facial recognition technology at schools. The resistance to universal identification was even more pronounced, both in the representative survey (with fifty-one percent expressing major reservations) and in the focus groups. One participant said a world with universal identification would be "a world I wouldn't want to live in"; this statement captures the basic tenor of the discussions. Participants also voiced concerns that universal identification would lead to more, and more extreme, cases of stalking. Respondents in the representative survey who are not worried about universal identification partially justify their opinion with the claim that anonymity is already a thing of the past.

Keeping big brother out: recommendations

Biometric data are highly sensitive and thus particularly worthy of protection. This makes it all the more vital that regulatory measures are adopted for using speech, speaker and facial recognition technologies in contexts such as medical care, law enforcement, lending or insurance institutions, and in the work environment.

Digitised voice recordings and facial images capture physical traits that are unlikely to change much after a person reaches adulthood. Once these kinds of biometric data have been compromised, their integrity is lost. For this reason, it is in general recommended that the collection of biometric data is limited and that authentication processes rely on several factors, for example an image and a password.

Ban on high-risk applications

Automated real-time surveillance through speech, speaker and facial recognition technologies meets with almost universal disapproval in Western democracies; it is also incompatible with several basic rights enshrined in Switzerland's Federal Constitution. Real-time surveillance systems in general must therefore be banned – as should the introduction of a social credit system that evaluates good or bad behaviours of individual citizens by means of a blanket surveillance system.

The use of smartglasses and other hidden devices that use facial recognition programmes to secretly observe people must also be prohibited in public spaces.

A further recommendation is that attention monitoring at schools be prohibited. In addition, fully automated decision-making processes based on speech, speaker and facial recognition technologies should be banned in particularly sensitive settings, especially at hospitals, banks and insurance companies, in law enforcement or at work. Instead, the results of semi-automated decision-making systems should be closely monitored by trained staff, who then authorise the decisions.

Moratorium on emotion monitoring and disease detection tools

As long as the technical and organisational reliability of emotion monitoring and disease detection systems that rely on facial and speaker data is not

given, this technology should be placed under a moratorium in certain life domains. Particular caution is advised in areas such as law enforcement or at insurance companies. It is especially important that, during a call to a customer centre, a client's voice is not analysed for diseases or emotions; this would represent an improper use of biometric data. Certain desirable applications could be classified as high-risk tools and thus permitted, subject to approval by the authorities.

Filling legal gaps, promoting training programmes, supporting data subjects

An explicit legal basis must be introduced for the use of speech, speaker and facial recognition technology in the public sphere. Such legislation must establish rule-of-law safeguards and ensure that the necessity of a technology is reviewed.

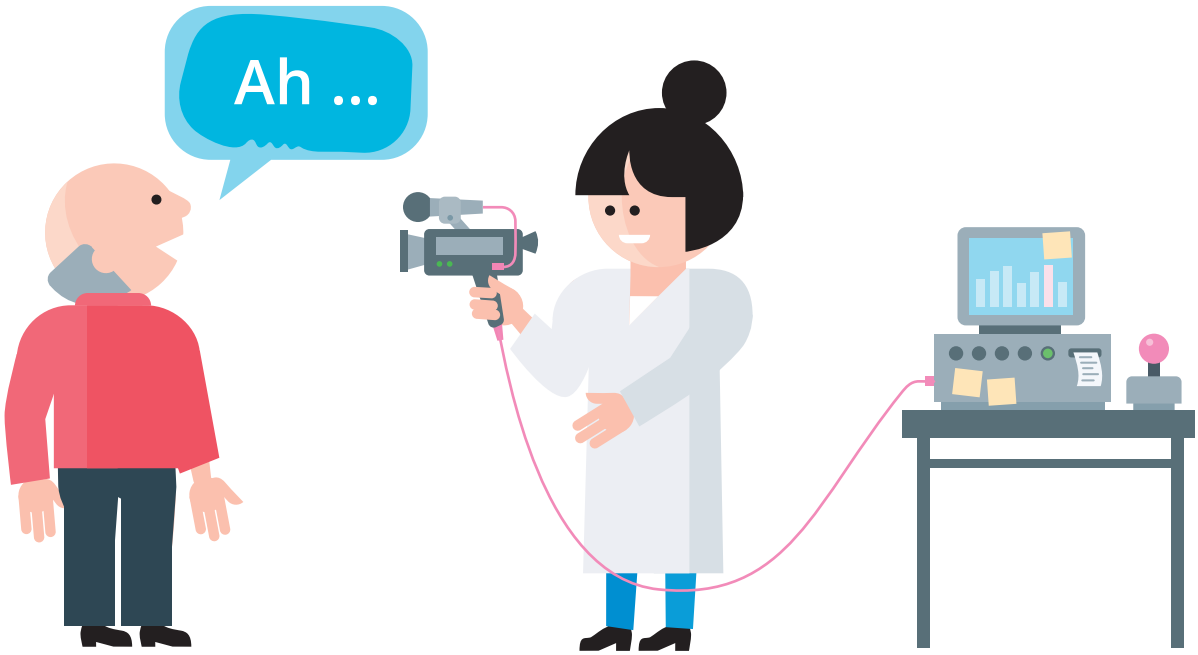
Professionals who use speech, speaker and facial recognition systems, have access to sensitive data and share results must receive proper training and further education. In addition, guidelines and support services are needed to ensure that operators of detection systems adhere to data protection principles.

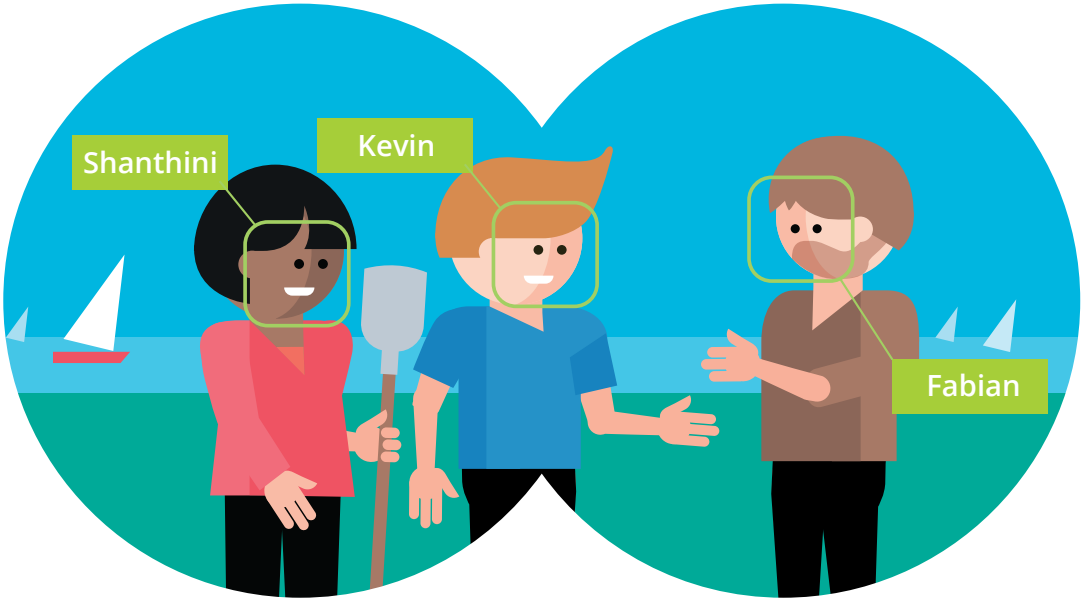
Whenever speech, speaker or facial recognition systems are used, it is important that this is communicated clearly. Wherever possible, data subjects should be able to choose an alternative without incurring negative consequences – longer waits or extra costs, for instance.

In addition, there is a need for a support service to aid people who wish to protect themselves from the disadvantages associated with speech, speaker and facial recognition systems, and who wish to understand and assert their rights.

Developers of devices that use speech, speaker and facial recognition technologies or related apps are advised, whenever possible, to store and process all data collected only in the device itself. This would both protect data and also grant users greater autonomy over their own data.

Lastly, it is important to foster societal discourse on the advantages and disadvantages of speech, speaker and facial recognition technologies, on where they are to be permitted and where their use should be limited. In addition, the general public should be encouraged to reflect on the significance that speech, speaker and facial recognition systems have for the way we live together – not least because the use of these technologies increases social pressure and can foster the tendency to adopt a moralistic view of others and their actions. The present TA-SWISS study is an invitation to this kind of discourse.




Shanthini
Kevin
Fabian

Advisory group

- Dr Bruno Baeriswyl, data protection expert member of the TA-SWISS Steering Committee, president of the advisory group
- Dominik Brumm, Head of Development, Cubera
- Dr Volker Dellwo, Institut für Computerlinguistik, Universität Zürich
- Dr Jean Hennebert, informatique et systèmes de communication, Université de Fribourg
- Dr Anna Jobin, sociologist, Alexander von Humboldt Institut für Internet und Gesellschaft
- Dr Annett Laube, Technik und Informatik, Berner Fachhochschule
- Dr Klaus Scherer, Swiss Center for Affective Sciences, Université de Genève
- Remo Schmidlin, lawyer, Lenz & Staehelin
- Dr Thomas Vetter, Departement Mathematik und Informatik, Universität Basel
- Patrick Walder, Amnesty International Switzerland

Project management at TA-SWISS

- Dr Elisabeth Ehrensperger, Managing director
- Dr Christina Tobler, Project manager (2020 – 2021)
- Dr Laetitia Ramelet, Project manager (2022)

Impressum

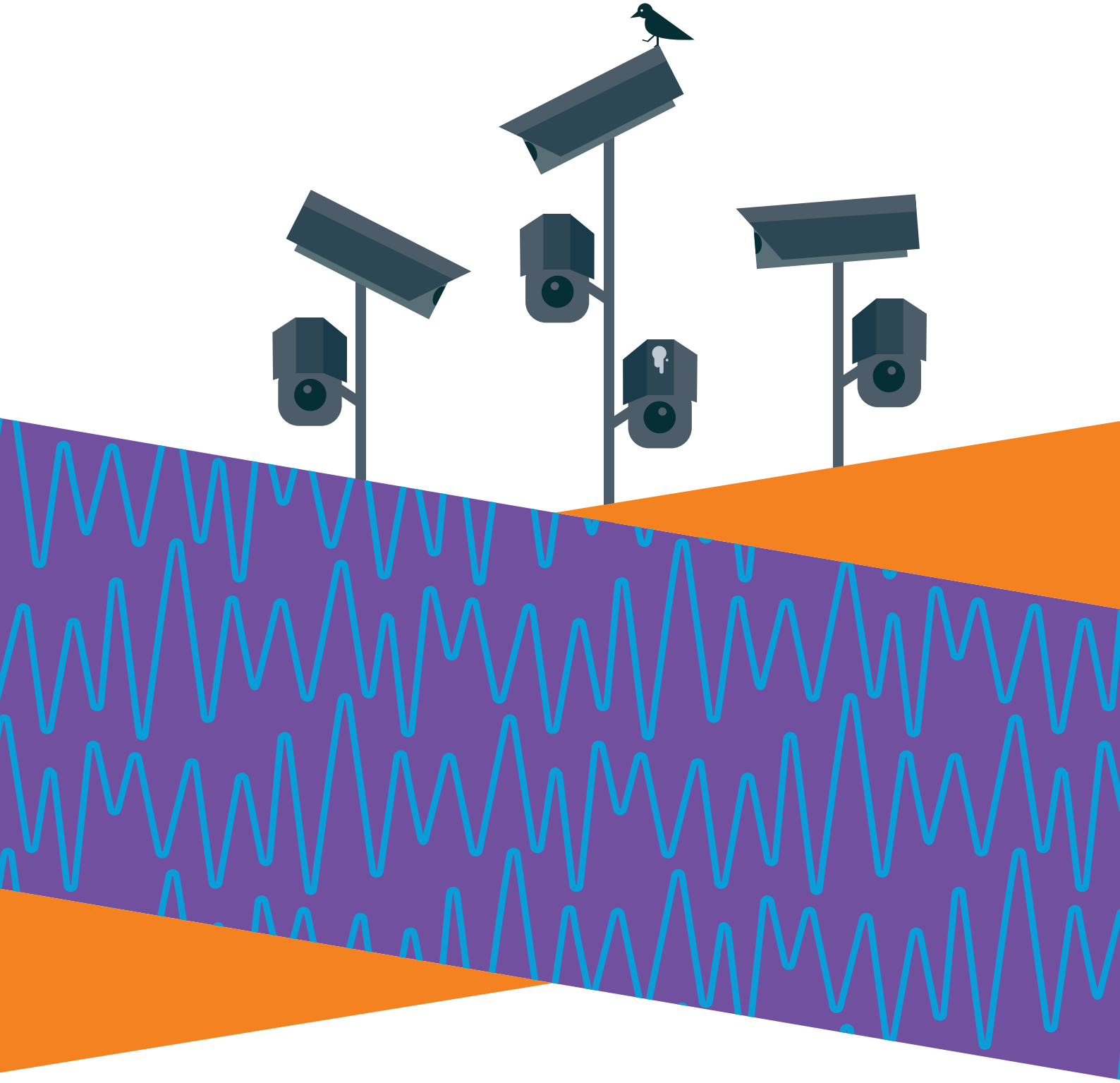
Observed at every turn
Abridged version of the study «Automatisierte Erkennung von Stimme, Sprache und Gesicht: Technische, rechtliche und gesellschaftliche Herausforderungen»
TA-SWISS, Bern 2022
TA 79A/2022

Author: Dr. Lucienne Rey, TA-SWISS, Bern
Translation: pro-verbial, Zurich
Production: Dr. Laetitia Ramelet and Fabian Schluep, TA-SWISS, Bern
Layout and graphics: Hannes Saxer, Bern
Printed by: Jordi AG – Das Medienhaus, Belp

TA-SWISS – Foundation for Technology Assessment

New technology often leads to decisive improvements in the quality of our lives. At the same time, however, it involves new types of risks whose consequences are not always predictable. The Foundation for Technology Assessment TA-SWISS examines the potential advantages and risks of new technological developments in the fields of life sciences and medicine, information society as well as mobility, energy and climate. The studies carried out by the Foundation are aimed at the decision-making bodies in politics and the economy, as well as at the general public. In addition, TA-SWISS promotes the exchange of information and opinions between specialists in science, economics and politics and the public at large through participatory processes. Studies conducted and commissioned by the Foundation are aimed at providing objective, independent, and broad-based information on the advantages and risks of new technologies. To this purpose the studies are conducted in collaboration with groups comprised of experts in the relevant fields. The professional expertise of the supervisory groups covers a broad range of aspects of the issue under study.

The Fondation TA-SWISS is a centre for excellence of the Swiss Academies of Arts and Sciences.



TA-SWISS
Foundation for Technology Assessment
Brunngasse 36
CH-3011 Bern
info@ta-swiss.ch
www.ta-swiss.ch

member of the
swiss academies
of arts and sciences