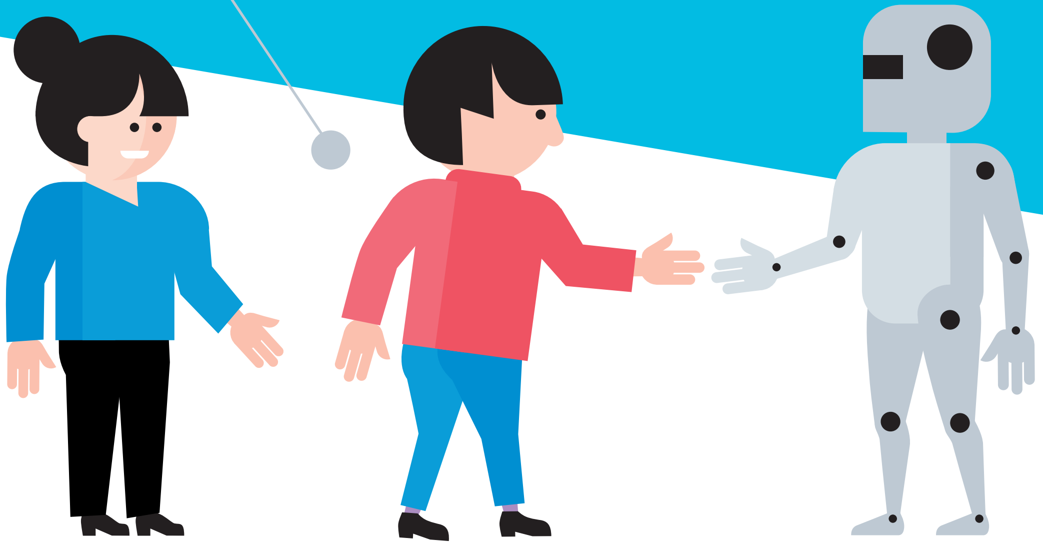
# Brainless but brilliant

**Summary of the «Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz» study conducted by TA-SWISS**



TA-SWISS, 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 led by Markus Christen (Digital Society Initiative University of Zurich), Clemens Mader (Swiss Federal Laboratories for Materials Science and Technology Empa) and Johann Čas (Austrian Academy of Sciences ÖAW). The abridged version presents the most important results and conclusions of the study in condensed form and is aimed at a broad audience.

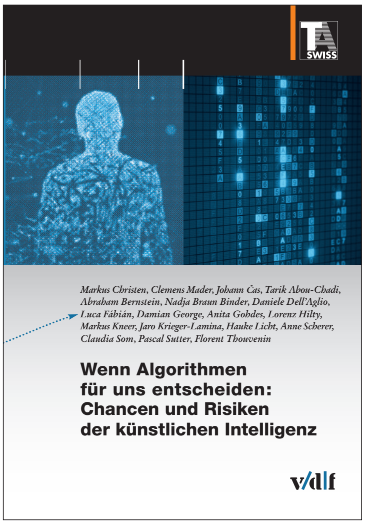
**Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz**

Markus Christen, Clemens Mader, Johann Čas, Tarik Abou-Chadi, Abraham Bernstein, Nadja Braun Binder, Daniele Dell'Aglio, Luca Fábián, Damian George, Anita Gohdes, Lorenz Hilty, Markus Kneer, Jaro Krieger-Lamina, Hauke Licht, Anne Scherer, Claudia Som, Pascal Sutter, Florent Thouvenin

TA-SWISS, Stiftung für Technologiefolgen-Abschätzung (Ed.)
vdf Hochschulverlag an der ETH Zürich, 2020.
ISBN 978-3-7281-4001-2

Also available in open access: www.vdf.ch

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



*Markus Christen, Clemens Mader, Johann Čas, Tarik Abou-Chadi, Abraham Bernstein, Nadja Braun Binder, Daniele Dell'Aglio, Luca Fábián, Damian George, Anita Gohdes, Lorenz Hilty, Markus Kneer, Jaro Krieger-Lamina, Hauke Licht, Anne Scherer, Claudia Som, Pascal Sutter, Florent Thouvenin*

**Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz**

**Artificial intelligence in brief** 4
Its opportunities … 4
… and risks … 4
… and some recommendations 4

**Artificial intelligence, a child of digitalisation** 5
From mathematical problem to complex model 5
Outperforming humans 6
Distortions, insufficient fairness and losses of confidence 7
Many different fields of application 7

**AI and consumption: a personalised shopping spree** 8
Personal preferences aid navigation 8
Human or machine? 8
Digital traps 8
Data protection also protects the person 9
Tricky balance between personalisation and data protection 9

**AI and public communication: opinion making in the media echo chamber** 10
Reinforced in the bubble 10
Conspiracy theories from the troll factory 10
Democracy at risk 11
Sceptical view of social networks 11

**AI in the working world: when machines carry out brain work** 11
AI impacts careers 12
Operating like a machine 12
Measures against the widening gap in the labour market required 12

**AI in education and research: providing support to human intelligence** 14
A superteacher for each child 14
Spotting future high flyers 14
Growing influence of private companies on education 14
AI is a subject – and driver – of research 15
Developing key competences and ensuring transparency 15

**AI in public administration: sovereign powers for artificial intelligence?** 15
Fighting fraud and crime 15
Fundamental rights prevail over governmental arbitrariness 16
Switzerland is cautious 16
No delegation of decision-making powers to machines 17
«Smart» machines – a challenge for today's legislation 17
Keeping AI systems on a tight leash? 17
First and second hand creativity 18
Fairness is difficult to calculate 18
Europe-wide ethical guidelines 18
Distortions and non-transparency are a major concern 19

**Recommendations: over a wide spectrum** 20
Focus on specific applications 20
High demands on public authorities 20
Transparency and manageable information 20
Quality seals instead of general market authorisations 20
Look beyond the technology 20

# Artificial intelligence in brief

**In contrast to conventional computer programs, which carry out clearly specified tasks according to predefined steps, artificial intelligence (AI) operates in a creative way. Trained on huge data sets, many AI systems are able to detect hidden patterns and come up with solutions that users would never have discovered themselves. More advanced AI systems can combine a number of different skills, such as language processing, information organisation and machine learning, and are even capable of performing challenging tasks currently assigned to highly trained professionals.**

## Its opportunities …

AI systems can be applied to a wide range of differing fields. They act much faster and often more accurately than humans and, in complex cases, could soon empower humans to conduct certain activities much more efficiently than ever before.

AI systems often allow services and offerings to be better adapted to the needs and abilities of each individual. It is in this capacity to personalise that AI holds great potential.

## … and risks …

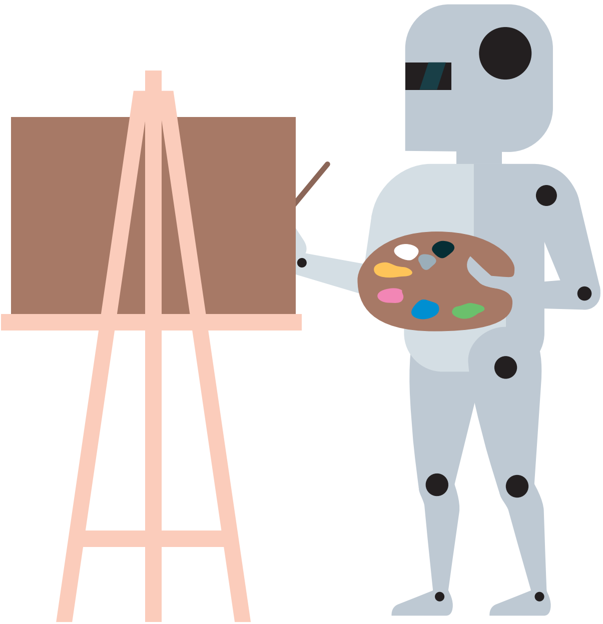
Many AI systems have to be trained on massive data sets to achieve the required skills. Privacy and data protection run the risk of being eroded by the immense amounts of data needed by AI.

If the data sets on which an AI system is based contain errors, results will also be incorrect. Unbalanced data also creates problems, as it leads to results that are mathematically correct but distorted in terms of content, which may systematically penalise certain groups of people.

Self-learning AI systems go on developing and may produce results that are no longer explainable – not only to their creators, but above all to those concerned.

## … and some recommendations

The use of AI systems should be clearly and easily identifiable so that those concerned know that they are interacting with an AI system and not with a human being.



Important decisions affecting people should not be left to an AI system without due consideration of the advantages and disadvantages. As a rule, when it comes to matters that have a relevant impact on people, the result provided by the system must be reviewed and justified by a human.

The use of AI has consequences that go far beyond its technical implications. For this reason, anyone who develops AI, uses AI or works with its results should also have an understanding of ethics and law and, in addition, be willing and able to collaborate on an interdisciplinary basis with representatives from other scientific fields.

> In terms of methodology of the study, the first step was to conduct a wide-reaching literature analysis. This provided the basis for a second phase in which a group of experts were asked in several rounds to provide assessments of the problem areas and theses that had been identified. The findings of both the literature analysis and surveys culminated in recommendations addressed to various target groups.

# Artificial intelligence, a child of digitalisation

**Clearly no longer a figment of science fiction, artificial intelligence (AI) has gradually crept up on us and now forms part of our everyday lives. This is reason enough to not only acknowledge its benefits – greater efficiency, for example – but also to shed light on its darker sides, especially its appetite for data.**

In terms of word origin, «intelligence» stands for the mental freedom of living beings who do not follow any predefined paths, but who can think up – and decide upon – alternatives. But what exactly does AI, the acronym for «artificial intelligence», stand for?

AI began to attract the attention of scientists in 1950. At that time, British computer scientist Alan Turing stated in an article in the magazine Mind that an intelligent machine is one that is capable of imitating a conversation with a human being so perfectly that they feel as if they are conducting a dialogue with a real person. The synonymously named Turing test – which is carried out on the screen with a keyboard and without visual or auditory contact with the interlocutor – has been the litmus test for the intelligence of artificial systems since the first scientific conference on AI held in Dartmouth (USA) in 1956. The term «artificial intelligence» was also coined at this conference.

## From mathematical problem to complex model

As a research subject in computer science, AI has substantially grown in complexity since its conception. Initial efforts focussed on creating solution algorithms for mathematical and geometric problems. It was not until the 1990s that more complex processes started to come within reach thanks to developments in computing power: for in order to decipher meaningful patterns from acoustic signals, images or other inputs, many AI systems have to be trained with huge amounts of data. The rampant rise in digitalisation and the resulting availability of large volumes of data therefore provide an important basis for AI.

Today, AI covers a number of scientific fields of research in which experiments are made to replicate rational or human action using an artefact – such as a computer program. One research subgroup, for example, aims to reproduce human thinking and learning. In addition to procedures that need large amounts of data in order to identify patterns and design models with the help of algorithms, others are also used that rely on expert knowledge or logical thought processes. Advanced AI systems can combine a number of different skills such as language processing, information organisation and machine learning.

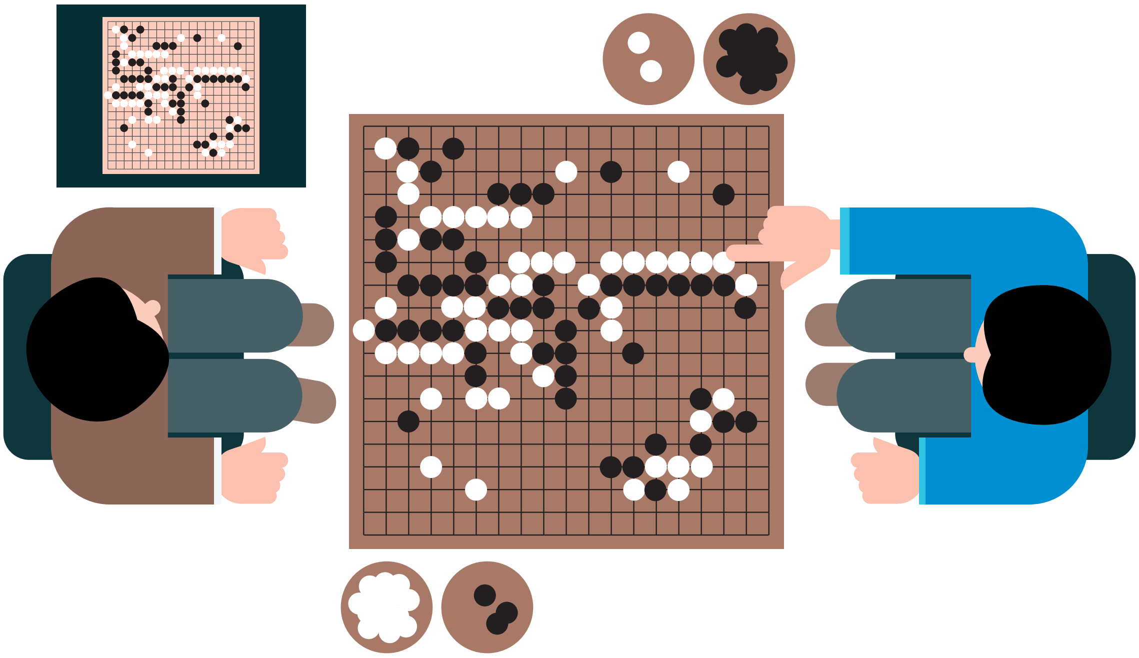
In recent years, AI has done much to increase the comfort and safety of our everyday life: mobility concepts that link public transport with rental cars or private cars are based on AI systems, railway companies raise safety levels with intelligent damage pattern recognition, and industrial manufacturing uses AI for more efficient processes – to name just a few examples. Basic research on the nature of the cosmos, as carried out by CERN, would be impossible without AI-based analysis of the enormous data volumes, and AI is also used for climate change simulations. The use of this kind of factual data in AI systems is generally viewed as unproblematic and often results in substantial benefits to society. It gets more tricky when AI operates with personal data. Applications of this kind are also discussed in the TA-SWISS study.

## Outperforming humans

It is the AI systems based on machine learning and neural networks in particular that solve problems at which conventional computer programs falter. With demanding and nuanced tasks such as translating, image recognition and analysis, as well as games, AI today is not only able to keep up with humans, but sometimes even surpass them.

A milestone was set in March 2016 when AI defeated the world's best player in the complex Asian strategy game «Go». But it didn't stop there: as a result, the software company DeepMind – a Google subsidiary – went on to develop a self-learning version that taught itself from the rules of the game alone instead of analysing numerous positions from games played with advanced human players, as earlier versions had done. Starting with randomly played games at beginner's level, the software called Alpha Go Zero reached professional levels within three days and went on to outclass its predecessor, which had succeeded in beating the world's number one.

Today's rapidly developing AI systems are increasingly able to perform tasks in domains previously reserved to humans and have now become a fundamental technology for countless applications. When new forms of AI are deployed, the role of humans in defining and solving problems changes: previously, it was always up to a human being to identify problems and specific tasks, decide upon the action required to solve the problem (and, if necessary, make corrections after initial setbacks) and finally implement the resulting actions in relatively narrowly defined fields of application; if a subsequent evaluation showed that the steps taken to resolve the problem had not led to the desired results, the human being, in turn, made the necessary amendments. In contrast, machine learning is largely carried out without human intervention: large data sets pre-structure the problem, and the action required to solve the problem – the algorithm – is developed on the basis of patterns in the data and by the machine's reinforced learning. The machine also carries out the feedback process itself by tweaking



and refining the algorithm if required on the basis of further, autonomously collected data; human beings remain excluded from large sections of the process, even if they have the right to reject the result of the AI system if necessary.

What may sound rather abstract has tangible consequences on everyday life. Take the example of obtaining a bank loan: in standard practice, a qualified specialist processes a client's request for credit on the basis of various criteria and by drawing on experience. Software, by contrast, examines the data of other clients, analyses their personal details and financial situation, and also knows which of them is punctually repaying the loan. From this data, the software creates a prediction model for the probability that a new client will pay off their loan. Self-learning systems, which are constantly improving through experience, raise a fundamental problem: who is responsible for a decision if the process chain that led to the decision is indecipherable because no-one is able to retrace it? Even if it is ultimately a human who assesses the results provided by the software and takes the decision, the situation gets tricky when the human gradually starts to lose understanding of the tool.

## Distortions, insufficient fairness and losses of confidence

When AI is used to find ways to solve problems, a number of fundamental difficulties arise.

When the fuel to drive an AI system consists of extensive data sets, the system is unwittingly fed any errors or flaws that are hidden in the data. And biased data can affect the behaviour of the algorithm – or even deliberately mislead it if the respective learning data has been manipulated. Nonetheless, any potential imbalance in large data sets – which are compiled, for example, from surveys in search engines – is not necessarily due to fraudulent intent but may instead have grown across time by its reflection of unquestioned value judgements and habits. In these cases, AI can help to detect such distortions – provided it is used with due care.

Another problem is due to AI's increasing independence. To be sure, the software begins from a starting point known to the programmer. However, particularly when artificial neural networks are involved, the strength and weighting of the individual connections change over the course of the numerous training cycles to such an extent that the software ends up proposing solutions whose logical or physical basis is almost impossible to fathom. In practical terms, such non-transparent AI algorithms, which effectively act as a black box, are of limited use when we want to understand how a system draws its conclusions.

Not only can the data be distorted, but also the algorithms as such. This is because the models are set by the developers, who may give priority to certain values and interests over others. Such «unfair» algorithms have already led to unfair results by assuming, for example, that people from poorer residential areas have a greater propensity for delinquency.

Empirical studies show that many people have more confidence in a human decision even when they know that AI has been proven to make more objective decisions. But there are also studies that point to the opposite, showing that people sometimes rely too much on the results of automated decision-making processes. It is difficult to assess the level of confidence that AI deserves.

Finally, there are signs that AI promotes corporate domination: because many new forms of AI depend on huge data sets, companies that have access to large amounts of data enjoy a competitive advantage. This explains why leading tech companies from the USA and China, which can use the information provided by their customers, are also involved in the development of AI. The phenomenon already known in the internet industry – whereby those who are already well positioned rapidly become more powerful, thus creating oligopolies – is likely to get stronger.

## Many different fields of application

AI is already being applied in a wide variety of areas. One of the strengths of AI systems lies in being able to tailor services and offers to the individual person.

AI is increasingly being implemented in the field of consumption, classical and new media, the working world, education and research, and administration. The TA-SWISS study takes a close look at these five areas of application. The order chosen in this brochure differs from that in the detailed study report and reflects the everyday experience of most users who are more likely to shop online and consult a screen for their news than to contact public authorities. However, this bears no relation to the significance of each field of application.

# AI and consumption: a personalised shopping spree

**When ordering goods online, close contact with AI is inevitable. The personalised shopping experience has become standard practice, with information technology processes working quietly in the background – and hardly ever being questioned.**

Messages like: «You may also like», «order too» or «matches your order» are sent to online shoppers from relevant shops directing them to other products that they might find attractive or practical. Recommendation systems such as these exploit customers' «digital footprints»: from an AI-based analysis of all purchases made on the corresponding platform and customers' personal details, as well as other behaviour such as time spent on the site or mouse movements, the system draws conclusions about the preferences of certain customer groups.

Advertising is also becoming increasingly personalised. Google and major providers such as Alibaba and Amazon are active in this field, as are social networks like Facebook: they advertise a whole variety of goods and services to specific target groups, often on behalf of other companies. The technology working in the background is comparable to that of recommendation systems.

## Personal preferences aid navigation

To start with, AI-based recommendation systems offer consumers the advantage of making it easier to navigate the consumer cosmos. And instead of advertising being randomly scattered, only advertisements for products tailored to the interests of the respective person appear on the screen.

Anything that helps customers to find their way around the online shopping complex helps suppliers to reduce their expenditure. Processing costs can be cut by up to 20 percent thanks to AI, and advertising has fewer losses when it can be focussed on a specific target group. Customer service also reaches a new level with the ability to be increasingly personalised. In addition, profits can be maximised – through dynamic pricing, for example: with AI, providers can determine their customers' willingness to spend and calculate the value of goods automatically.

## Human or machine?

AI is increasingly being used not only for the purchase of goods but also services. In May 2018, for example, Google introduced the Duplex digital personal assistant, which uses a deceptively realistic-sounding female voice to arrange a hairdresser appointment or reserve a table in a restaurant on behalf of its owner. Because Duplex seemed to intimidate many people, even in a country open to technology like the USA, Google reset the assistant to introduce itself at the beginning of a conversation. In its blog, the Group points to other potential applications of this technology, such as various customer or public authority information services: instead of having to put up with long waits until a personal advisor is available, digital assistants could increasingly be used to answer the most frequently asked questions.

The Duplex example is a palpable illustration of a basic problem associated with AI: when ordering goods and services via the internet or mobile phone, it is becoming more and more difficult for customers to know whether a flesh-and-blood person is answering or whether they are interacting with an algorithm. Moreover, it is not only difficult for customers to see if and when AI systems are being used, but also how and for what purpose. Even if consumers are told that their conversations or personal data are being recorded, they don't know which conclusions the algorithms will later draw and with what accuracy.

## Digital traps

An account in which the information about all the previous orders a person has made, and also where their individual data is stored, is practical: a personalised entrance portal displaying images of their last purchases creates a feeling of familiarity, and since the person's address and credit card number are also saved, there's no need for them to waste time entering the information again every time they place a new order. Over time – i.e. with each transaction – the system gets to know its customers better and better, so that its recommendations grow ever closer to their tastes. There is however another side to the coin: convenience tends to make us stick with one supplier, even though another may offer a more attractive range of products or services. Marketers refer to this effect as «stickiness».

In addition, the sheer size of a network can also have the effect of binding individuals. After all, it is useful to be part of a highly extensive network through which you can reach many other people. It is also practical to be able to perform different activities on a single platform without having to log in separately for every transaction or interaction. As a result, the data is ultimately concentrated in a few large corporations holding considerable market power.

## Data protection also protects the person

Our online tracks are extremely revealing. The social network, Facebook, published a study with over 8000 voluntary users. Working on the basis of the users' «likes», their personal characteristics were assessed more accurately by the AI system than by their personal friends. Ten likes were enough for the machine to deliver better results than work colleagues; from seventy likes upwards, the machine's assessment exceeded that of friends. Other studies have shown that AI systems need only a profile picture to detect a person's individual characteristics – including sexual orientation – with astonishing accuracy.

AI is particularly effective when it comes to combining personal data from different sources, an advantage that companies like Facebook and Google know well how to exploit, having amassed a range of popular services: the photo service Instagram and WhatsApp messenger service belong to Facebook, while Google owns – amongst many others – the video portal YouTube and the fitness tracker company Fitbit. Other technology giants such as Amazon and Alibaba have also purchased a variety of web platforms and data sets, enabling them to link data from different sources and thus increase their competitive advantage over smaller companies. However, not all specialists see a danger in the concentrated data sets; after all, users can supply their data to a number of different providers. Simply having access to certain information does not guarantee a company's lead over its rivals.

The solution cannot be to saddle the customers alone with the responsibility of protecting their own data. Admittedly, a company's General Terms and Conditions (GTC) is obliged to give details on what information they gather and store. However, even if someone agrees to the terms and conditions, it cannot then be assumed that this person thereby allows access to their data, because they are unlikely to ever be aware of the extent of the information collected and how it is used: just to read the GTC of their most frequently used services in full would take an average user several weeks a year.

## Tricky balance between personalisation and data protection

According to the experts surveyed for the TA study, it is a matter of urgency that users become more aware of the value of their own data. However, it is difficult for companies to reconcile the process of personalisation with data protection. At least this is how the surveyed experts see it. The threat of customers losing confidence is also a problem that online providers need to address. In the eyes of the experts, effective measures to enhance customer confidence would be: open data with control options for customers, the transparent disclosure of the use of AI, as well as the right to delete data.

The emergence of oligopolies in which a few giants dominate the market is cited by the respondents as another significant risk associated with the use of AI systems. After all, structures of this kind could ultimately lead to a reduction in the range of products and services on offer and also restrict customers' ability to switch providers. It is therefore important to ensure that customers can take their personal data with them when they decide to switch providers. In addition, any shifts in the market must be closely monitored and control and intervention mechanisms must be put in place to safeguard competition.

# AI and public communication: opinion making in the media echo chamber

**With more and more people using social media such as Facebook or Twitter to stay up to date, the press, radio and television have lost their prominence over the last ten years. Professionally researched news set in a broader context must now compete against blog articles and social network posts.**

In 2017, 83 percent of the population in Switzerland consumed their news online, 45 percent thereof through social media. However, in addition to Facebook and Twitter, more people are now turning to platforms such as Google News and Apple News which bundle content from other sources. As a result, more and more users are sharing articles and information directly with each other, often also sharing news from traditional providers. It is becoming increasingly difficult to clarify the origin of the news being circulated.

## Reinforced in the bubble

One of the responsibilities of conventional media is to provide balanced information on a wide range of topics and to reflect the prevailing diversity of opinion. People who get their news from the press or radio and television are automatically also confronted by reports on issues of little interest to them.

This is different when news is shared via social networks. If several news feeds are sent to a person's account, the algorithm determines which content is shown first – based on the information collected about the account holder, which the holder has either voluntarily disclosed or which has been revealed on the basis of their previous online behaviour. In doing so, AI systems tend to offer information to users that corresponds to their content preferences; as a result, users then navigate primarily within their filter bubble. If, on top of this, people only read articles recommended to them by their friends, they run a greater risk of being trapped in an echo chamber, only encountering content that corresponds to their interests and reflects their own opinion, while any deviating stances or conflicting information is eliminated.

Although digital filter bubbles and echo chambers are attracting considerable public attention and are being discussed as a threat to democracy, there is as yet no scientifically substantiated evidence that the diversity of opinion online is more limited than in conventional circles of like-minded people. However, the developments described above are still in their early stages and could very well prove to be a problem for public opinion formation in democratic societies.

## Conspiracy theories from the troll factory

Inaccurate media reports are no new phenomenon, with the «canard» (French for duck) having enriched the world of animal-related metaphors since the early 19th century. A canard usually appears involuntarily and damages the reputation of the newspaper that published it. What's new, however, is the deliberate act of spreading misinformation via social media on a massive scale.

The process of distributing fake news is now already largely automated and highly efficient. Perpetrators are using social bots – simple programs written specifically to spread news in a targeted manner via social media. During the 2016 US election campaign between Hillary Clinton and Donald Trump, around one fifth of the tweets sent out were from social bots; of the automated Twitter accounts, the vast majority (i.e. up to three quarters) sided with Trump.

In the future, AI could have the capability of imitating trolls, i.e. users who write emotional blog and forum posts, often containing false information, with the aim of provoking violent responses. With this strategy, the trolls – who are often employed in actual «factories» – seek to raise the visibility of certain groups or even circulate conspiracy theories. Social bots as well as trolls often refer to websites that are specifically designed to spread misinformation and increase the credibility of their digital henchmen. AI systems are already maximising the efficiency with which trolls deliver their messages.

### Democracy at risk

One strategy adopted by disseminators of fake news is to infiltrate established news channels. In the USA it has been shown that false information is first likely to be sent to local media stations, as they have fewer opportunities to verify the facts. Once included in the local news, fake news quickly reach the national media. However, AI not only helps to spread false information, but also helps to enhance it. For example, AI can be used to manipulate images, audio files and even videos, or to create fictional footage from scratch to discredit or blackmail certain individuals. Specialists call these technologically generated fake videos «deep fakes».

Under authoritarian regimes that crush any undesirable reporting and control the media, social media is sometimes the last remaining outlet for independent information. In healthy democracies, however, influencing public opinion with fake news distributed through social media is seen as a serious threat to the political system. Firstly, because the public loses confidence in established news channels: recent surveys reveal that in almost all countries only less than half of the respondents trust the ability of conventional media to distinguish fact from fiction. Secondly, various high-profile analyses have shown that fake news spreads even more quickly than true information. In Germany, for example, it was shown that of the ten most widely publicised news stories about Angela Merkel, seven were classified as fake news.

### Sceptical view of social networks

The surveys conducted by TA-SWISS as part of the study revealed a sceptical attitude towards social media: respondents mainly fear that social networks will have a damaging effect on the quality of journalistic content. The act of circulating news online is viewed upon with slightly less severity. Personalised content is also favoured by the experts; the key factor seems to be that users themselves must be able to decide whether, how and to what extent the news stories are personalised for them.

Fake news, on the other hand, is seen as a serious threat by the survey participants. Only a minority is of the opinion that technical progress will lead to improvements in exposing false information. On the contrary, the majority fear that advanced technology will only generate fake news of an even more sophisticated nature. However, the results of the survey also made clear that in the fight against the circulation of misinformation, attention must be paid to maintaining the right to freedom of expression. Rather a critical view is taken of the fact that bodies of authority are being set up with the specific task of checking content. Nonetheless, respondents are happy when platform operators delete fake news and block its authors. Automated – i.e. AI-based – fact-checking is considered more beneficial than manual checking procedures.

## AI in the working world: when machines carry out brain work

**For a long time, only simple and repetitive activities were considered at risk of being sacrificed to rationalisation. But since AI has succeeded in penetrating into high-profile professions requiring good qualifications, the fear of job losses has also spread among white-collar workers.**

Around the turn of the millennium, various newspaper columns entertained readers with translations of short texts produced by machines. The programs failed at even the simplest examples of language imagery and came up with the most peculiar and often comical formulations. Until recently, translation was considered a job that could not be done by a machine. But software such as DeepL, which works with artificial neural networks, can now provide perfectly adequate translations, while other programs can even generate simple journalistic texts on their own.

It is therefore no longer just simple routine tasks that are affected by automation: AI is used in the medical field, where it provides more accurate results than most medical professionals when it comes to evaluating ultrasound, magnetic resonance or X-ray images, or in the legal field, where it helps lawyers search thousands of contract pages for specific clauses.

Figures on the threat of job losses – or, depending on the point of view, the potential for rationalisation – are provided by a number of studies. Estimates for possible job losses range between 6 and 57 percent, depending on the basic assumptions and methods applied. Specialists agree that – in contrast to previous economic upheavals caused by innovation – it will probably be difficult to make up for AI-induced job losses with new products and services.

## AI impacts careers

AI now has an influence not only on the quantity but also on the quality of work. For example, HR managers are increasingly using AI systems to pre-select applications when recruiting new staff. Applicants must first complete standardised questionnaires online. HR managers then resume the process using the AI-generated shortlist. This doesn't necessarily work against the candidates: if a human first examines the submitted documents, unobjective decision criteria such as a person's photogenic appearance may be disproportionately weighted. Using an AI system therefore provides the opportunity to expose and change existing discriminatory practices.

Large corporations are increasingly using AI to draw up skills profiles for their employees and to propose further training courses tailored to the individual. The problem with this is that it curtails personal autonomy for it is no longer the individual who is deciding upon their own further training but a system that is aiming to optimise the human (and financial) resources available. Even if it isn't the system that ultimately makes the decision, its recommendations can develop a momentum of their own within the company, when employees find it difficult to justify why they prefer not to follow a particular recommendation. Overall, the protection of personal data poses a fundamental problem to the use of AI in the working world.

## Operating like a machine

In the future, AI will not necessarily lead to the disappearance of entire jobs, but rather of individual activities. Employees will have to learn to collaborate with «smart» machines and to deal with a faster pace and increased density of work. When it comes to the organisation of work, specialists and institutions such as the International Labour Organization (ILO) stress the need to put the interests of people first in order to prevent them being worse off than before the introduction of AI systems. And especially in cases where AI behaviour is unexplainable, it will be important that employees are able to assess the plausibility of the results to avoid any gross errors.

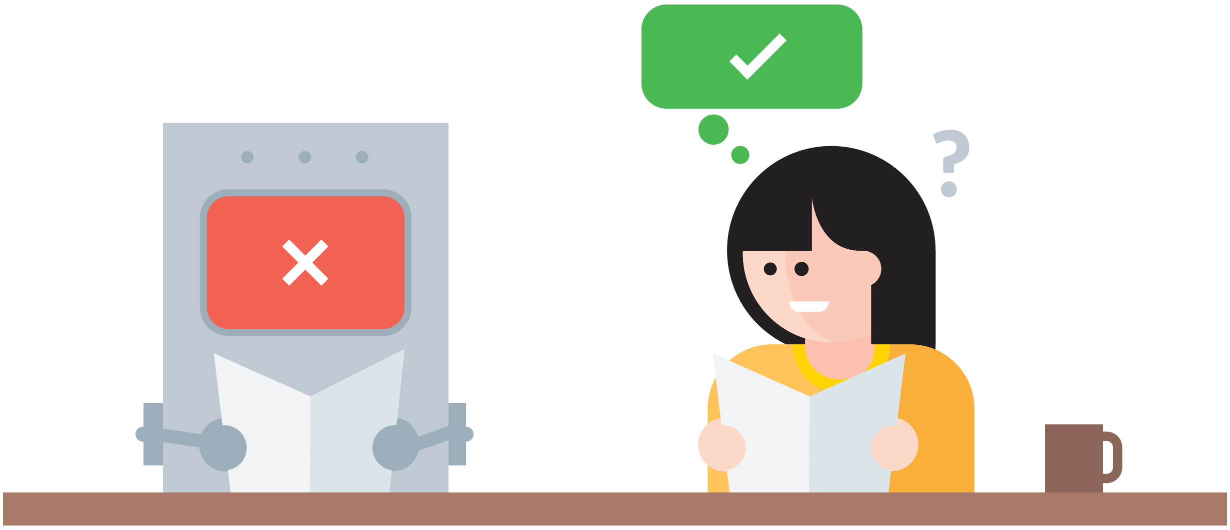
The influence of AI on the organisation of work is difficult to separate from the effects of digitalisation in general. While the TA-SWISS study on flexible work demonstrated that digitalisation can help increase people's levels of personal freedom and their ability to manage their time autonomously, it also puts our social security system into question. Take all the online platforms that break down large orders and pass them on in packages to «clickworkers»: although these platforms can help by, for example, enabling students to earn extra pocket money in the evenings, they actually raise the pressure and existential concerns of workers like self-employed creatives who use these platforms to acquire business in the international market.

## Measures against the widening gap in the labour market required

The experts surveyed in the TA-SWISS study consider it likely that AI will accelerate the division on the labour market. They expect the gap between highly-qualified employees with corresponding salaries and low-skilled workers on small incomes to continue to widen. The experts also predict more control over employees and a growing instability in employment relationships. By contrast, they consider AI unlikely to have any effect on reducing workers' hours and workload. All in all, the majority of the experts anticipate that large companies will be the ones drawing the greatest benefits from these changes in the working world.

In the eyes of the experts surveyed, continuing education is a particularly effective measure against these polarising developments. The question of which are the most suitable actions to combat the growing divisions in the professional world seems to polarise the labour market just as much as AI does. A relatively small majority of the respondents support the idea of easing social inequality by expanding the social system; an equally narrow majority is opposed to the idea of reducing working hours in order to distribute the diminishing number of jobs among more people. Again, it is only a relatively small majority that supports the introduction of an unconditional basic income. In contrast, the authors of the study themselves consider a better and more flexible distribution of work and income to be the preferred alternative in the short to medium term. Furthermore, new types of employment opportunities such as creative start-ups need to be supported and innovative forms of taxation developed in order to ensure funding for the national budget and social system.

One concern the respondents have is the need to protect employees from excessive control by their employers. Measures against workplace surveillance coupled with better data protection in general are considered desirable and effective, although employees contracted under employment relationships that extend beyond national borders must also be protected. However, various experts also emphasise the positive potential of AI when it is used to automate monotonous work and in applications that increase both productivity and competitive edge. Provided the advantages gained from AI are distributed fairly, society as a whole will benefit.



# AI in education and research: providing support to human intelligence

**AI is a product of scientific research – which, in turn, helps to advance scientific development. AI systems can be used in schools to provide individual support to pupils.**

The idea of supporting pupils on an individual basis is already part of the education policy agenda. AI raises the prospect of providing the required technology.

## A superteacher for each child

Numerous start-ups have already developed tools to help schoolchildren meet their educational goals. Take tools such as chatbots, for example: these are programs that automatically respond to learners' questions and messages by giving answers based on keywords and rules that are either pre-programmed or individually generated by the AI system.

Other programs guide the pupils step by step through tutorials that teach them what they need to know in mathematics, science or physics. First a skills and interest profile is drawn up by the AI system in order to identify the child's strengths, weaknesses and existing knowledge; after this the system analyses the child's learning path and suggests the best step to take next.

Another approach is to have the AI system monitor a learner's writing process and provide ongoing feedback. Other systems include those that organise learning in groups by networking schoolchildren with matching profiles or by putting pupils who fail at mathematics in contact with maths tutors from other classes, schools or even countries.

AI also assists teachers – not only in assessing their students' performance, but also in providing personalised support to individual learners. In administrative tasks alone, teachers can save several hours a week with AI applications.

## Spotting future high flyers

With more and more learners doing their schoolwork online in dialogue with an AI system, the provider's servers are flooded with personal data. As a result, the provider not only learns a lot about the young people's learning strategies and skills, but also about their preferences and leisure activities. Data of this kind is interesting as it allows for the early recognition of potential high flyers. Data on underachievers is also useful as these children may be open to the idea of a tutoring course offered by the same company.

When a Swiss school uses AI software from private providers, a number of particularly critical data protection issues arise: who is allowed to store, own or access which personal data and for how long? And how will it be ensured that the data can be deleted if necessary? Can pupils take their data with them when they change schools – and are other schools allowed to view the data when they admit new students? How can it be guaranteed that data on examination and learning performance will not affect competition for jobs in the future if some candidates can present good results, while others have no data at all or only data showing inferior results. In order to counter key risks that arise from problematic handling of personal data, the Swiss Conference of Cantonal Ministers of Education EDK has launched the edulog.ch platform, allowing Swiss pupils to log into educational programs anonymously. The cantons are participating voluntarily, and it remains to be seen how much use will be made of the service in practice.

## Growing influence of private companies on education

A majority of the experts surveyed by TA-SWISS for the study welcome the chance to tailor teaching content to the individual student. This personalisation allows each learner's specific problems to be addressed. However, the experts are concerned that private companies will exert more and more influence on schools, and that schools could consequently become dependent on one provider.

Nonetheless, banning the use of private AI systems is not considered a viable option. Instead, precise and binding agreements should be negotiated with the providers, including clear rules on data protection. The experts diverged on the question of whether responsibility for developing the AI systems for schools should lie with the public authorities but agreed that collaboration between the authorities and private providers would be highly desirable. The majority of the experts also recommend an official review of AI applications designed for use in schools.

### AI is a subject – and driver – of research

In absolute figures, the highest number of publications on AI are issued in China and the USA. But Switzerland's research activities are also quite remarkable: in terms of population size, more scientific articles on AI are published here than in Germany or France; weighted according to the impact of the citations, Switzerland is one of the world leaders.

AI is applied to the most diverse scientific disciplines, and especially to those in which huge amounts of data are analysed: in materials science, for example, as well as in biochemistry and astrophysics. It is also useful in teaching – when it comes to exposing fraud in academic work, for example. AI-based tools can detect similar sounding passages in plagiarised texts as well as fake statistics. AI is also able to recognise a change in writing style, which is an indication of ghostwriting.

### Developing key competences and ensuring transparency

AI is becoming increasingly valuable for research; indeed, major leaps in development often depend on it. The majority of the experts surveyed therefore believe that scientific institutions should set up contact points to promote continuing education and the use of AI in research, which could also serve as a hub for interdisciplinary exploration of the technology.

There was general consensus among the respondents that data generated by AI must remain the property of the respective research institution or be made publicly available in an open access format. Finally, all knowledge generated by AI systems must be as transparent and reproducible as possible. In order to fulfil these requirements, corresponding transparency guidelines must be drawn up.

## AI in public administration: sovereign powers for artificial intelligence?

**As a rule, public authorities work on the basis of forms to be filled out and standardised procedures: ideal conditions for applying AI. But greater efficiency is countered by the risk of curtailing the fundamental rights of the citizen.**

In September 2018, a Swiss National Councillor from the Social Democratic Party submitted a motion requesting that the Federal Council establish a centre of competence in public administration. The idea of this centre was to promote the use of AI and machine learning in public administration in order to increase efficiency. The Federal Council recommended the motion be rejected.

### Fighting fraud and crime

AI has managed to penetrate into official functions. Australian tax authorities, for example, implement AI to automatically collect money owed to the government. The system has been criticised for sometimes drawing incorrect conclusions and for leaving it to the taxpayers to protect themselves from unduly high taxes. In Germany's tax offices, AI is used to detect fraud in tax returns.

Although the use of AI in the taxation process has not (yet) been discussed in Switzerland, digitalisation is making advances here too: customs duties are increasingly being recorded and processed digitally, and various excise duties – on alcohol, mineral oil and tobacco, for example – could also be levied automatically in future. The corresponding legal bases are provided in the annex to the draft of the new Data Protection Act adopted in September 2015.

In terms of police operations, one use of AI is to estimate in which neighbourhood a crime might soon occur. Based on these predictions, task forces can conduct more intensive patrols of areas where the probability of break-ins or assaults are particularly high. Predictive police work is put to frequent use in several US cities. Another approach focusses not on location, but on people and their social network: using data from social contacts in specific circles, the AI system calculates the probability of a person being a member of a criminal gang.

In another instance, AI helps to assess the risk of criminals re-offending. In the USA, an AI system is used to calculate the probability of a criminal committing another offence on the basis of 137 characteristics. The system came under criticism because it systematically assumed that coloured offenders had a higher risk of recidivism than white offenders. Although software is also used in Switzerland for criminological assessments – in particular to determine whether a prison sentence can be reduced for a particular person – this is not based on advanced AI technologies.

## Fundamental rights prevail over governmental arbitrariness

In performing their sovereign duties, public authorities process a large amount of personal data. Using AI systems therefore promises a substantial increase in efficiency in this area: in the future, documents and files could be created automatically, while voice-based digital office assistants would answer routine inquiries from the public.

However, the authorities must be sure to exercise great care in whichever actions they take and not make their actions subject to efficiency alone. The fundamental rights guaranteed to the individual vis-à-vis the state stipulate that each individual is to be protected from any acts that might encroach upon positions protected by their fundamental rights.

The protection of personal data is therefore a crucial requirement to this end; all data must be stored securely and protected from misuse. All persons should also be able to assume that the data collected about them and subsequently processed is accurate, complete and balanced. In reality, however, many people are not aware of which data is actually stored and used in the context of AI-based processing; consequently, they are hardly in a position to investigate the data sets compiled on them and correct any errors.

In view of the distortions and imbalances that can result from AI-based data processing, automated procedures are particularly problematic since they could lead to discrimination against certain groups and individuals. The revised Data Protection Act now stipulates that persons must be informed in the case of fully automated decision processes.

Leaving the decision to a fully automated AI system ultimately contravenes the constitutional requirement that any official directive or ruling imposed on a person must be justifiable. Every official legal act has to be presented in a transparent manner so that the persons concerned can object if necessary. However, if an AI-based, self-learning machine automatically makes a decision or performs substantial preparatory work for the decision, it may no longer be possible to trace the AI system's reasoning. This also makes it impossible to justify the decision in a way that meets the requirements of the law.

## Switzerland is cautious

In the survey carried out as part of the TA-SWISS study, the majority of the experts were of the opinion that Switzerland was less likely than other countries to deploy AI systems in public administration activities and the chance of complex official procedures becoming fully automated was very low. In Switzerland, AI systems could at best be implemented for issues of a more simple nature and in partially automated processes.

However, the majority of the respondents see a considerable risk in Swiss authorities becoming dependent on technology providers from abroad. The experts also point to the threat of a certain servility to machines, in the sense that administrative staff might find it difficult to justify a position that differs from that of the AI system. Finally, the experts warn of a weakening in data protection and of the fact that automated processes cannot be adequately controlled.

## No delegation of decision-making powers to machines

A clear majority of the surveyed experts agree that when it comes to predictive analyses or drawing up a personality profile («profiling»), AI systems may only make recommendations; the decision itself must be left to humans. In addition, a slight majority of the respondents believe that AI should only be used to prepare spatial or object-related decisions, but not decisions where people are involved.

The majority favours an obligation by the government to publicly declare in which procedures it deploys AI. In addition, the authorities should clearly explain the fundamental elements of the processes in which AI is used. Finally, the government should be required to justify each individual decision that is reached on the basis of AI.

## «Smart» machines – a challenge for today's legislation

If an engineer miscalculates the bearing capacity of a wall or a translator makes a translation mistake, it's generally clear who is liable for the damage. But what happens if the errors are caused by an AI system? AI raises new legal and ethical questions, some of which are briefly outlined below.

Until now, the legal system has assumed that at the beginning of an action a decision is taken by a subject, and everything that happens afterwards is linked to this decision – including the legal consequences. The use of AI puts this concept for attributing damage to its perpetrator into question. The problem is made worse by the fact that AI systems are increasingly making independent and unpredictable decisions.

## Keeping AI systems on a tight leash?

The idea of taking responsibility for damage not directly caused by oneself is every dog owner's nightmare: whether the four-legged friend has dug holes in the neighbour's garden or snapped at the postman's hand, it is the owner who must pay compensation or face other consequences. Owners can only exempt themselves from liability if they can prove that they took the necessary care in handling their animal under the circumstances in question.

Various legal scholars believe that pet owner liability could serve as a model for a specific form of liability for the actions of autonomous AI systems: these too are sometimes unpredictable and have the potential to cause harm. Critics of this analogy point out that the action undertaken by AI that caused the damage was delegated to an AI system, therefore the origin of the whole chain of actions lies with the person who implemented the AI system in the first place.

A more extensive concept could lead to AI systems having their own legal status. Since 2017, the EU Commission has been considering creating an «electronic person', which would make it possible to hold advanced robot systems liable for the damage they cause; a similar approach would also be conceivable for AI systems. However, according to specialists, this should clearly be rejected – at least for the time being.

## First and second hand creativity

Under current law, mathematical methods are excluded from patentability. The situation is less clear regarding the patenting of computer programs. Based on the practice of the European Patent Office, patentability can be granted in cases where machine learning makes a «technical contribution» to an invention – by using neural networks to classify digital images and videos, for example. In other words, the decision as to whether an AI system can be patented has to be reviewed on a case-by-case basis.

And questions are still open on the subject of copyright issues. Although computer programs are deemed to be works worthy of copyright protection, problems arise with systems that learn by themselves and are therefore constantly developing. The result is that the object of copyright protection also changes – because at a later date the AI system is no longer what it was at the beginning.

A further difficulty emerges in the case of intangible assets produced by AI systems. It is unclear whether such AI-generated assets are protectable at all and, if so, who from the point of view of intellectual property law is to be regarded as the creator: the programmer who developed the system in question, the person who uses it – or the system itself? While the latter possibility is already ruled out since AI systems have no legal capacity, distinctions must be made if such property rights are to be granted to a natural person: Under copyright law, the protectability of works generated fully autonomously by AI is likely to be rejected on the grounds that no human being has contributed intellectually to their creation – which, under copyright law, is a basic prerequisite for the protectability of a work. In practice, however, problems arise in making a distinction between such works and those in which AI was merely involved as an aid.

From the point of view of patent law, on the other hand, it can be assumed that inventions generated by AI will enjoy basic protectability. In this context, the natural person who first takes note of the result and who understands it as a solution to a technical problem is legally classified as an inventor.

## Fairness is difficult to calculate

AI raises not only legal but also ethical problems that are almost impossible to solve. This can be illustrated by the US-American AI system «COMPAS», which relies on 137 characteristics to estimate whether criminals will become repeat offenders. The system came in for criticism when it became apparent that it wrongly calculated a high risk of recidivism almost twice as frequently for African-Americans as for Whites. In turn, the recidivism rates for Whites was almost twice as often not predicted as it was for Blacks. The algorithm had not received any information about the ethnicity of the defendants.

The company providing the test defended itself against the accusation of racism with the comment that it had developed a test to ensure that the precision rate was the same for dark- and fair-skinned people – and this is guaranteed by unbiased software. Mathematicians have subsequently looked into the question of whether a test can be both fair and unfair at the same time, and conclude that this is very possible – namely when a specific characteristic (in this case recidivism) is displayed with different frequency in a community by different subgroups. Since African-Americans reoffend more often than Whites, it is not possible to deliver equally precise predictions for both groups and at the same time achieve equal levels of error rates for both ethnic groups. Specialists therefore maintain that justice cannot be established on the basis of statistics or mathematics. The responsibility falls rather to ethics to come up with procedures and criteria that will ensure or at least increase the fairness of algorithms.

## Europe-wide ethical guidelines

In spring 2019, a group of AI specialists contracted by the European Commission published a set of ethical guidelines to ensure trustworthy AI practices. The guidelines are designed for both AI developers and users as well as people who work with its results. They aim to ensure that basic values and current regulations are respected and that commitment is shown to only using AI for ethically desirable objectives.

The guidelines place particular emphasis on ensuring that control over decisions and processes remains with a human being including when AI is applied. In addition, the technical systems should be robust and provide consistent reproducible results. Data and privacy must be protected, and AI systems should be used transparently in order that the steps leading up to a decision remain trackable. Finally, when using AI systems, unfairly distorted results must be prevented and social and environmental aspects given due consideration too.

## Distortions and non-transparency are a major concern

The majority of the experts surveyed share the opinion that unbalanced data sets and non-transparency in the use of AI pose considerable risks. Nevertheless, the experts are fundamentally in favour of the use of AI and acknowledge its positive contribution to achieving commonly accepted societal goals such as reducing poverty or improving health.

However, all agree that the responsibility for AI decisions must clearly lie with humans. Users need to be informed when they are interacting with AI, and should also be enabled to understand how the AI system reached its decision.

# Recommendations: over a wide spectrum

**The range of application fields in which AI systems can be deployed is wide. The recommendations for how to manage this technology are therefore equally extensive.**

AI systems are used for the most diverse tasks across a wide spectrum of fields, each of which also raises its own legal questions. For this reason, the idea of introducing an «AI law» to provide standardised regulation over the use of AI is not appropriate and is to be rejected.

## Focus on specific applications

Who uses an AI system, for what purposes exactly, what data is it based on, and what legal requirements are to be heeded? These are the questions that need answering before the opportunities and risks of the deployment of AI can be assessed. For this reason, regulations covering the use of AI must always be drawn up in accordance with the respective context of application. Concrete problems and undesirable developments must also be identified and resolved with appropriate precautions; in areas where the risks are unclear, research should be stepped up to determine potential dangers.

## High demands on public authorities

When the government performs its sovereign duties, it acts from a superior position vis-à-vis the individual. For this reason, it is all the more important that the legality of any governmental action can be verified by those concerned, and that AI systems are implemented with transparency. Thus, when the government uses an AI system, it should actually meet higher specifications than users from the private sector.

## Transparency and manageable information

If persons are relevantly affected by the use of AI, they must be informed in a simple and easily comprehensible way that they are interacting with an AI system. However, it should be noted that an excess of information can lead to an attitude of resignation on the part of the persons concerned and ultimately to negligence. It must therefore be ensured that they, as well as organisations that represent injured parties in legal proceedings, are given, upon request, all the information they require to identify and evaluate erroneous results.

## Quality seals instead of general market authorisations

It would not be appropriate to make the use of AI systems universally subject to one market authorisation, given the systems' diverse and varying degrees of application. In cases like medical products, where a market authorisation is already required, the assessment process includes AI. But encouragement and support should be given to any private bodies who develop quality seals for AI. Organisations such as those engaged in consumer protection should be better enabled to assess the quality of such certificates.

## Look beyond the technology

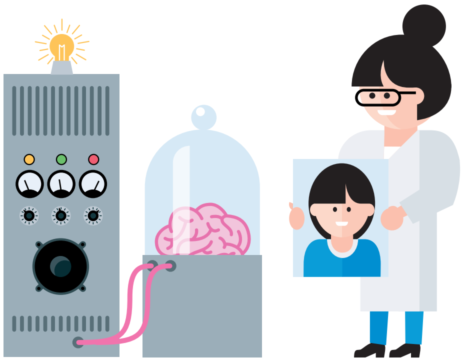
The conception and development of AI goes far beyond the technology itself. In order to be able to assess the risks involved, it is essential that specialists who develop AI also have an understanding of ethics and basic legal expertise, and are also willing and able to collaborate on an interdisciplinary basis. Relevant expertise is also indispensable for all those who use AI systems or work with their results.

## Advisory group

- Prof. Dr. Jean Hennebert, Leitungsausschuss TA-SWISS, Département d'informatique de l'Université de Fribourg, Präsident der Begleitgruppe
- Benjamin Bosshard, Eidgenössische Kommission für Kinder- und Jugendfragen
- Sabine Brenner, Geschäftsstelle Digitale Schweiz, Bundesamt für Kommunikation (BAKOM)
- Dr. Christian Busch, Staatssekretariat für Bildung, Forschung und Innovation (SBFI)
- Dr. Christine Clavien, Institut Ethique Histoire Humanités, Université de Genève
- Daniel Egloff, Staatssekretariat für Bildung, Forschung und Innovation SBFI
- Andy Fitze, SwissCognitive – The Global AI Hub.
- Matthias Holenstein, Stiftung Risiko-Dialog
- Dr. Marjory Hunt, Fonds national suisse de la recherche scientifique (FNS)
- Manuel Kugler, Schweizerische Akademie der Technischen Wissenschaften (SATW)
- Thomas Müller, TA-SWISS Leitungsausschuss, Redaktor Schweizer Radio SRF
- Katharina Prelicz-Huber, TA-SWISS Leitungsausschuss (bis 2019), Präsidentin Gewerkschaft VPOD/SSP, Nationalrätin
- Prof. Ursula Sury, Rechtsanwältin und Professorin, Hochschule Luzern (HSLU)
- Dr. Stefan Vannoni, TA-SWISS Leitungsausschuss, cemsuisse

## Project management at TA-SWISS

- Dr. rer. soc. Elisabeth Ehrensperger, Geschäftsführerin
- Dr. Catherine Pugin, Projektleiterin



## Impressum

Brainless but brilliant
Abridged version of the study «Wenn Algorithmen für uns entscheiden:
Chancen und Risiken der künstlichen Intelligenz»
TA-SWISS, Bern 2020
TA 72A/2020

Editing: Lucienne Rey, TA-SWISS, Bern
Translation: pro-verbial, Zürich
Production: 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 Foundation 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