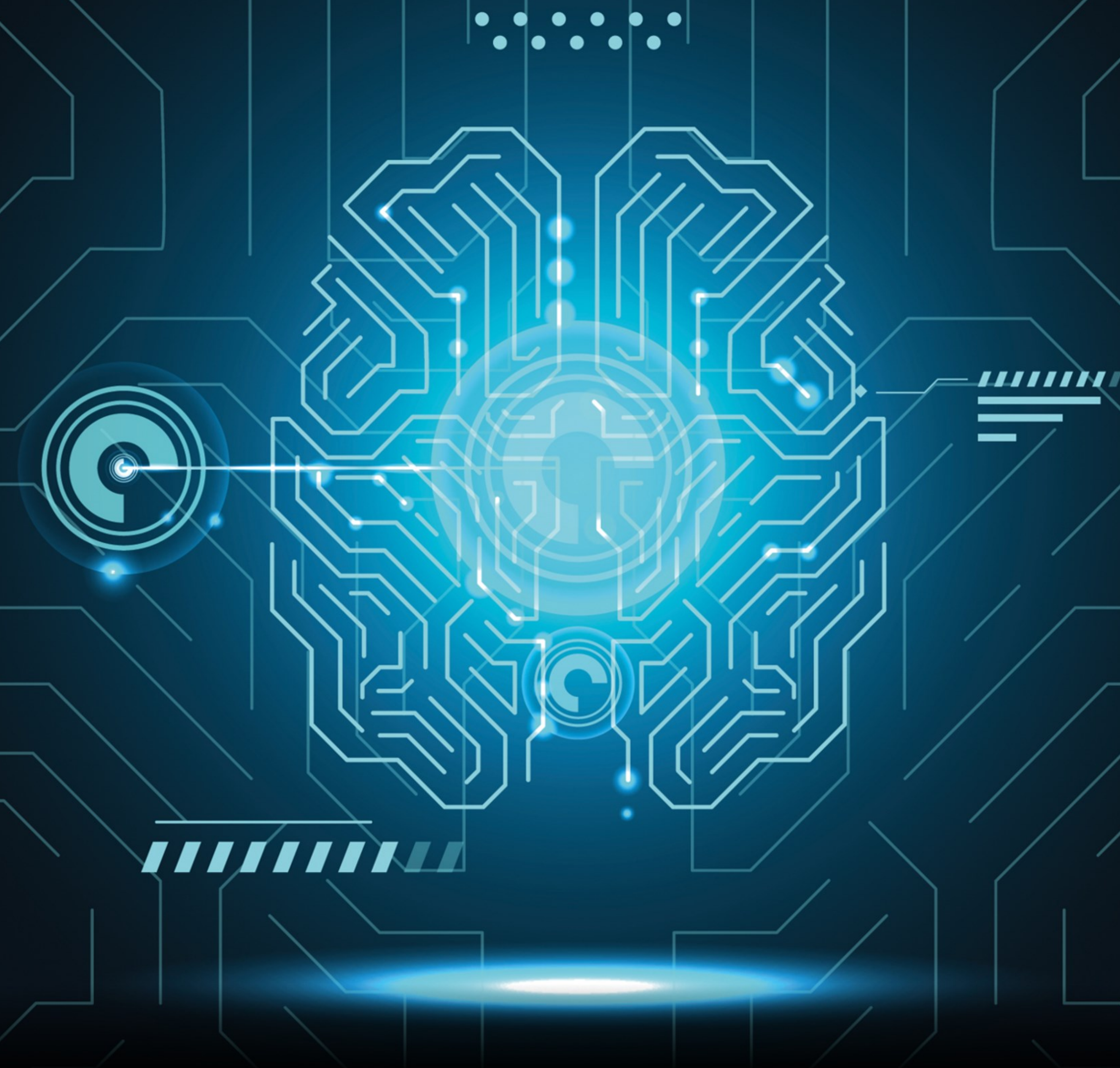
Recommendations for an AI Strategy in Switzerland



satw it's all about technology

Recommendations for an AI Strategy in Switzerland

A white paper organised by
the SATW topical platform on Artificial Intelligence

Table of Content

Contributors

Alessandro Curioni, Editor, *IBM Research – Zurich*
Lukas Czornomaz, Editor, *IBM Research – Zurich*
Joachim Buhmann, *ETH Zurich*
Ernst Hafen, *ETH Zurich, DatenundGesundheit Assoc., MIDATA Genossenschaft*
Manuel Kugler, *SATW*
Hervé Bourlard, *Idiap Research Institute and EPFL*
Jana Koehler, *Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI)*
Matthias Kaiserswerth, *Hasler Stiftung*
Anika Schumann, *IBM Research – Zurich*

This document reflects a consensus between the contributors, but the responsibilities lie with the authors of each part.
The positions and opinions in this document are those of the contributors and don't necessarily represent IBM's positions, strategies or opinions.
This white paper has not undergone SATW's usual review process, and SATW therefore assumes no responsibility for the content or the quality of the document. Responsibility for these lies solely with the authors.

Cover photo: freepik.com

Preamble[1]

Digital transformation is radically reshaping almost every aspect of our society. The explosion of artificial intelligence (AI) and big data analytics applications is enabled by the extreme availability of data in combination with the substantial computing power of modern highly distributed computing infrastructures connected by high-speed networks. Machine learning technologies can be trained to perform specific tasks with an efficiency and an accuracy that can supplement and, in some cases, outperform that of humans. These systems provide deep insights by learning from data and interactions with users, which is already leading to a profound transformation of numerous industries, professions and society at large. The current state of AI is, however, still far from delivering truly intelligent behaviour that is comparable to human intelligence. An AI research strategy should therefore carefully analyse AI's history with its various waves of large promises and conceptual shortcomings.

Recent advancements in machine learning have enabled AI technologies to become extremely successful. Speech recognition, natural language interaction with machines and facial recognition based on deep learning are now commodities that have changed the way people interact. The machine learning strategy of emulating human performance by learning from human experience promises a solution to the knowledge extraction problem. However, the automated reasoning process is as opaque as human decision making. Evolution has enabled humans to collectively reason and act on our collective experience, though other humans are often black boxes. Today, we are confronted with computational artefacts that are adapted to complex human decision making and, thereby, have inherited a similar "black box" behaviour.

Given the penetration of AI across most industries, its potential impact on GDP promises to be very high[2]. In Switzerland, AI is already reshaping industries such as banking, insurance, pharmaceuticals and manufacturing. Furthermore, Switzerland is the European country that has the highest number of AI start-ups per citizen[3], with more than 100 start-ups[4]. Many leading countries are heavily investing in AI development strategies and the establishment of technology transfer centres in this field[5,6,7,8,9,10,11,12,13]

[1] **Preamble – lead author: Lukas Czornomaz**

[2] Chui, M., et al., (2018). NOTES FROM THE AI FRONTIER –INSIGHTS FROM HUNDREDS OF USE CASES. McKinsey Global Institute. Retrieved from www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning

[3] The European Artificial Intelligence Landscape | More than 400 AI companies built in Europe. medium.com/cityai/the-european-artificial-intelligence-landscape-more-than-400-ai-companies-build-in-europe-bd17a3d499b

[4] 100 start-ups on the Swiss Artificial Intelligence Start-up Map, www.startupticker.ch/en/news/july-2017/100-startups-on-the-swiss-artificial-intelligence-startup-map

[5] Overview of national AI strategies. medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd

[6] US to endorse new OECD principles on artificial intelligence. www.politico.eu/article/u-s-to-endorse-new-oecd-principles-on-artificial-intelligence/amp/

[7] Ethics guidelines for trustworthy AI. ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[8] Eckpunkte der Bundesregierung für eine Strategie Künstliche Intelligenz. www.bmwi.de/Redaktion/DE/Downloads/E/eckpunktepapier-ki.pdf?__blob=publicationFile&v=4

[9] « La stratégie IA, pour faire de la France un acteur majeur de l'intelligence artificielle ». www.enseignementsup-recherche.gouv.fr/cid128618/la-strategie-ia-pour-

To date, Switzerland has not developed a dedicated AI strategy. AI is one of many topics covered in the strategy "Digitale Schweiz". An interdepartmental working group on AI which should ensure knowledge exchange in the domain of AI within the federal administration and coordinate Switzerland's positions in international bodies, is mandated to submit a report to the Federal Council by September 2019. Furthermore, an interdisciplinary study on behalf of TA-SWISS is evaluating the opportunities and risks of AI on the basis of various focal points: work, education, media, consumption and administration. The publication of that study is planned for the end of 2019.

The contributors to this publication are convinced that Switzerland can be a frontrunner in selected fields of AI services and applications. However, to attain this position, a structured approach and common direction for efforts in AI research and technology development are required. This document formulates recommendations for an AI strategy for Switzerland. The goal is to position Switzerland as a leading AI country and to amplify and accelerate the positive impact of AI on the Swiss economy. The recommendations are delivered in the context of five important aspects of AI.

Switzerland should pioneer a model for an open market for data, stimulating the valuation and exchange of data while ensuring privacy, security and trust. Thus, we propose leveraging Switzerland's unique historical geopolitical reputation and trust to promote itself as **a safe harbour for data** storage. Furthermore, Switzerland should take the lead in the definition of the general requirements that AI systems should follow to enhance their acceptability by businesses and society. As AI is being more widely deployed, it becomes crucial to ensure that these models meet high ethical standards and provide safety. **A verification body for AI** could provide requirements for trusted ethical and safe AI systems.

To improve the image of AI, we need to **increase human trust in AI**. A positive message about how AI empowers humans to become more efficient, makes knowledge more accessible and improves quality of life needs to be communicated. The profound impact of data science on society requires additional educational efforts to enable the population to meaningfully use this AI technology. Corresponding challenges lead to a demand for **AI in higher education** but also a profound revision of the topic catalogues in grammar schools. The positive effect of AI on the Swiss economy could be strongly accelerated if we were to create an appropriate framework through which to ease access and **enable AI for SMEs**. AI empowers SMEs to compete better with larger market actors and to accelerate their growth.

faire-de-la-france-un-acteur-majeur-de-l-intelligence-artificielle.html?menu=4

[10] Pan-Canadian Artificial Intelligence Strategy. www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy

[11] United Kingdom AI Sector Deals. www.gov.uk/government/publications/artificial-intelligence-sector-deal

[12] US President's approach to AI. www.whitehouse.gov/wp-content/uploads/2018/05/Summary-Report-of-White-House-AI-Summit.pdf

[13] New Generation Artificial Intelligence Development Plan. www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm

A Safe Harbour for Data

The status quo and Switzerland's future in a global AI-dependent digital society[14]

Individuals are a major source of data for AI solutions in various fields, including healthcare, education, urban planning, energy, and economics. This chapter focuses on how personal data can best be harnessed in a fair and sustainable way that protects individuals' privacy and digital self-determination and how Switzerland can play a leading role in the concomitant democratization of the personal data economy.

Personal data are a new asset class[15], and its market value is estimated to reach more than one trillion Euros in 2020[16]. Currently, personal data are stored in incompatible silos and are increasingly fuelling the profits of a few multinational data companies that offer free services and smartphone apps in exchange for personal data. This approach has led to a rapid increase in socioeconomic asymmetry in the concentration of and control over personal data[17]. With the increasing use of AI in healthcare, education and other disciplines, this asymmetry will increase dramatically since the companies and institutions with the largest amount of data will be able to develop the best algorithms[18]. This asymmetry threatens to leave little opportunity for smaller institutions and countries to catch up or compete, particularly countries such as Switzerland. Although academic research groups in Switzerland have been at the forefront of cloud computing and machine learning, the technology and knowledge gap between large multinational cloud storage and AI providers and local companies is rapidly expanding.

In the European Union, the rights of individuals to control their personal data have been strengthened with the enactment of the European General Data Protection Regulation (GDPR) in May 2018. In particular, the right to data portability gives EU citizens the opportunity to obtain a digital copy of all their personal data. In the revision of the Swiss data protection law, which is currently being discussed in parliament, the federal council has omitted the data portability provision on the grounds that it decreases the competitiveness of small and medium enterprises, a view that is also shared by Economiesuisse[19]. Thus, in Switzerland, individuals currently possess significantly less autonomy in the control of their own data than EU citizens. In a future digital global society that is heavily dependent on AI solutions involving personal data, individuals will have to play a central role as the aggregators

[14] **A Safe Harbour for Data – lead author: Ernst Hafen**

[15] World Economic Forum. (2011). Personal Data: The Emergence of a New Asset Class (pp. 1–40). World Economic Forum.

[16] The Boston Consulting Group. (2012). The Value of Our Digital Identity (pp. 1–65). Retrieved from www.libertyglobal.com/PDF/public-policy/The-Value-of-Our-Digital-Identity.pdf

[17] Haynes, P., & Nguyen, C. M.-H. (2013). Rebalancing Socioeconomic Asymmetry in a Data-Driven Economy. In B. Bilbao-Osorio, S. Dutta, & B. Lanvin (Eds.), The Global Information Technology Report (pp. 67–72).

[18] Lee, K.-F. (2017, June 24). The real threat of artificial intelligence. New York Times. Retrieved from www.nytimes.com/2017/06/24/opinion/sunday/artificial-intelligence-economic-inequality.html

[19] Djonova, I., & Herzog, E. (n.d.). Eine Datenpolitik des Vertrauens für Fortschritt und Innovation (No. # 03/2018). Retrieved from www.economiesuisse.ch/de/entityprint/node/45425

and access controllers of their personal data. In the current financial economy, most citizens possess a bank account and decide how to invest or spend their money. It is the individual consumer who drives the economy.

The role of citizens[20,21] in a data-driven society and economy is more central than in the current consumer economy, which is still dominated by physical goods, because three unique features of personal data set them apart from goods and financial assets. First, personal data are a nonrivalrous good. In contrast to money and other physical assets, data are not consumed; they can be copied and reused. This fact makes the article on Data Portability in the EU GDPR a true innovation since it legally entitles data subjects to copies of all their personal data[22]. Second, personal data are a new asset class that is equally distributed among individuals. In contrast to the uneven distribution of financial assets, our common biology implies that each individual possesses a genome that contains six billion base pairs, has a heart that beats with a similar frequency, consumes similar numbers of meals (even though caloric intake differs), etc. Third, and most importantly, individuals are the maximal aggregators of their personal data. Only individuals have the potential and the legal right to aggregate medical, social media, consumer, and genome data. AI solutions in areas such as healthcare, education or urban planning are increasingly dependent on the aggregation of such different data types from millions of people. Thus, the individual, as an active producer and aggregator of personal data and consumer of data services, will play an even more important role in a future AI-supported society. For this to happen, however, there is need for a new trust-promoting framework in which citizens play an active role in the personal data economy.

Medical data stored in incompatible silos or on paper has hindered systematic outcomes research and medical research in general. The only datasets that are also in demand internationally are those of cohorts[23] funded by the Swiss National Science Foundation, which provide a longitudinal dataset of AIDS patients and healthy individuals and typically contain data from several thousand participants.

A trust-promoting framework that supports the fair and active participation of citizens relies on several of the qualities for which Switzerland is recognized internationally. The one person one vote principle of direct democracy fits well with the equal distribution of personal data. Historically, Switzerland has been a forerunner in democratically controlled cooperatives. Some of Switzerland's alpine farming cooperatives are several hundred years old[24]. As a repository and platform for sharing personal data, cooperatives offer two advantages over other organizational forms. Cooperatives belong to their members. Much like Swiss citizens who control their government, cooperative members manage their data and control cooperative governance and how revenues are invested.

[20] Toffler, A. (1980). The third wave.

[21] Tapscott, D. (1995). The Digital Economy.

[22] De Hert, P., Papakonstantinou, V., Malgieri, G., Beslay, L., & Sanchez, I. (2017). The right to data portability in the GDPR: Towards user-centric interoperability of digital services. Computer Law & Security Review: The International Journal of Technology Law and Practice, 34(2), 1–11. doi.org/10.1016/j.clsr.2017.10.003

[23] Example 1: The Swiss Digital Health Cohort, https://www.satw.ch/fileadmin/user_upload/documents/02_Themen/08_Kuenstliche-Intelligenz/SATW-Swiss_AI_Strategy-Example1.pdf

[24] Ostrom, E. (2015). Governing the Commons (Reissued edition). Cambridge.

Recommendations for networked personal data cooperatives

The Swiss MIDATA cooperative[25], with its non-profit and ethical governance principle, aims to be the founding example of networked personal data cooperatives in other countries, thus promoting the democratization of the personal data economy. Moreover, Switzerland's trusted role as a safe harbour and provider of financial services could be extended to the management of personal data. Finally, Switzerland's neutrality and its credibility in international organizations will help promote the democratization of the personal data economy.

Given that personal data is a non-rivalrous good, cooperatives in which citizens control the aggregation of and access to their personal data do not replace but extend the existing global personal data ecosystem. All existing providers of data and AI solutions will benefit from such an extension, since they obtain access to new data aggregates. Democratically governed cooperatives acting as the fiduciaries of their member's data can ensure that the socioeconomic asymmetry in a global data and AI-dependent society is rebalanced and that the economic benefits of these data are also returned to society at large.

Switzerland possesses an active and internationally competitive academic research community in machine learning, AI, cloud computing and data security. Recent investments in the Swiss Data Science Center and the Swiss Personal Health Network initiative are further steps in strengthening academic research in the areas of AI. As outlined above, it is essential to **empower and enable citizens to become active data aggregators** and participants in all areas that rely on the availability of personal data, including healthcare and education. Early projects and platforms by CERN and the University of Geneva[26] as well as ETH Zurich and the University of Zurich[27] to promote and strengthen the participation of citizens in science are encouraging.

Two problems remain that the Swiss national parliament must address. First and foremost, it is essential that Switzerland adopt the **data portability framework** of the EU GDPR. Switzerland should do so, not only because it is the only country in Europe that does not grant its citizens the fundamental right to a digital copy of their personal data but more importantly because it lays the foundation for a citizen-controlled extension of personal data and AI ecosystem.

Second, the rapid introduction of an **electronic identity** (eID), which is currently being discussed in parliament, is essential. Countries, such as Estonia, that possess such an eID have seen a transformative shift towards a digital society, with many start-ups and international companies offering new services. Moreover, politicians and the media need to realize that, while it is important to talk about privacy and data protection, it is equally important to talk about the po-

[25] Example 2: Box – MIDATA Personal Data Cooperative, https://www.satw.ch/fileadmin/user_upload/documents/02_Themen/08_Kuenstliche-Intelligenz/SATW-Swiss_AI_Strategy-Example2.pdf

[26] The Citizen Cyberlab, www.citizencyberlab.org

[27] The Citizen Science Center Zürich, citizen-science.ch

tential benefit of open, transparent and fair data sharing. Many studies have shown that people are willing to share data for the benefit of medical research and for personalized information[28],[29].

Finally, financial support for the effective initiation of the transformation to a **fair democratically controlled personal data ecosystem** is essential. Once established, a new democratically controlled personal data economy that involves hundreds of millions of people across the globe will generate substantial benefits because of the data aggregation power of citizens and because they will not only contribute data but also their human intelligence. Run by citizen-owned non-profit cooperatives, these benefits will be returned to society. Examples include real-world patient-reported outcomes via smartphone apps and sensors from medical treatments, drug efficacy or the active recruitment of people for clinical trials. Today, pharmaceutical companies are making deals with data companies for hundreds of millions to billions of dollars to access digital health records[30,31]. The market value of real-time health data from smartphone sensors and medical records willingly and knowingly provided directly by people will be orders of magnitude greater.

However, the path to establishing such a democratically controlled data economy cannot be financed by classical investment strategies that rely on venture capital equity investments. Cooperatives belong to their members, one vote at a time, and thus cannot accept equity investors. Therefore, there is a need for foundations, philanthropists and crowd funding to bridge this initial gap to achieve the breakeven point of data cooperatives.

[28] Vayena, E., Ineichen, C., Stoupka, E., & Hafen, E. (2014). Playing a part in research? University students' attitudes to direct-to-consumer genomics. Public Health Genomics, 17(3), 158–168. doi.org/10.1159/000360257

[29] Mooser V, Currat C. The Lausanne Institutional Biobank: a new resource to catalyse research in personalised medicine and pharmaceutical sciences. Swiss Med Wkly. 2014;. doi.org/10.4414/smw.2014.14033

[30] Roche and Foundation Medicine reach definitive merger agreement to accelerate broad availability of comprehensive genomic profiling in oncology. (2018, June 19). Roche and Foundation Medicine reach definitive merger agreement to accelerate broad availability of comprehensive genomic profiling in oncology.

[31] GSK and 23andMe sign agreement to leverage genetic insights for the development of novel medicines | GSK. (2018). Retrieved from www.gsk.com/en-gb/media/press-releases/gsk-and-23andme-sign-agreement-to-leverage-genetic-insights-for-the-development-of-novel-medicines/

A Verification Body for AI

Towards ethical and safe AI[32]

New risks arise with the use of AI and decision-making algorithms. Aspects such as bias or interpretability are nontrivial and require a thorough examination. AI systems should follow general ethical and safety requirements that enhance their acceptability by businesses and society. To date, no clear generally accepted guidelines for the implementation of AI systems exist. Therefore, there is a need to define the essential requirements that must be fulfilled by an AI system and the associated verification processes. The definition of such ethical and safety requirements and their verification implies a profound technical expertise but also a more societal, political and philosophical point of view regarding how democracies can adapt to coexist with AI and benefit most from it.

The topic of the requirements, validation and verification of AI systems is related to software liability, where providers typically are not liable for bugs in their systems or consequential damage caused by a failure of their systems. Medical software tools or apps also have similar challenges, and their effectiveness must be provable to be accredited as a medical product. Providing broadly applicable requirements and verification processes could advance the development and dissemination of AI systems in Switzerland and thereby support and strengthen the local economy. The global trust in its institutions and democratic system could help to position Switzerland as a global role model that defines widely acceptable guidelines and procedures for AI systems.

There are many activities in the field of trust in AI; increasing the reliability of algorithms is a prerequisite for their deployment in critical applications. Various experts and initiatives aim to foster ethics in AI[33,34,35,36,37]. However, there are few concrete projects establishing a holistic view for addressing general ethics and safety requirements across industries and applications other than some limited projects with specific applications, e.g., autonomous cars[38].

In different countries and cities, laws are implemented to guide the application of machine learning and the accountability of algorithms[39,40]. Currently, we do not know of any concrete measures for the implementation of an interdisciplinary approach to specify the sufficient and practical requirements for the safety of self-learning systems and to address the validation and verification of AI systems in Switzerland.

[32] **A Verification Body for AI – lead author: Manuel Kugler**

The authors acknowledge **Markus Christen** (Managing Director Digital Society Initiative, University of Zurich) and **Christian Westermann** (Leader Data & Analytics, PwC Switzerland) for their valuable inputs to this section.

[33] The partnership on AI is a multi-stakeholder organization to better understand AI's impacts. www.partnershiponai.org

[34] Informatics Europe recommends to establish means, measures and standards to assure fair ADM systems. www.informatics-europe.org

[35] High-Level Expert Group on Artificial Intelligence supports implementation of European AI strategy. ec.europa.eu/digital-single-market/en/high-level-group-artificial-intelligence

[36] MIT Turing Box as a certification environment. turingbox.mit.edu

[37] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. standards.ieee.org/industry-connections/ec/autonomous-systems.html

[38] TÜV SÜD and DFKI to develop "TÜV for Artificial Intelligence". www.tuv-sud.com

[39] European General Data Protection Regulation. eugdpr.org

[40] New York City Council, Fairness in computing. legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0

Essential general requirements of AI systems

General requirements and associated verification and validation procedures should cover cross-sectorial industries, be technologically independent and apply to systems of different providers. They should differentiate AI systems based on whether they are truly self-learning, adaptive or fixed. The general guidelines should also provide a framework to determine whether or not a human is required in the loop depending on the criticality of the application and severity of the consequences of a failure. Decision-making systems that are truly self-learning pose the greatest challenges. In such cases, a decoupling of decision making and autonomous action-taking might be required.

Another key requirement is to demonstrate the system's flexibility in reacting to new situations. The algorithms should be robust and reliable when facing shifts in data or manipulations of processed data, i.e., the algorithms have to generalise. The algorithms should be able to detect when the input data are outside of the trained capabilities and defined boundaries. Learning algorithms must be enhanced by an automonitoring capability to assess the uncertainty in the algorithm's knowledge of the domain/world. Often, it is essential to know what one does not know. To minimize the risk of a system failure in such an event, a domain-specific process is required, e.g., a feedback loop so humans can remain in control of the system's response.

The control of bias in data or AI systems must be considered. AI systems should be fair, and they should not fall prey to unintended discrimination. There must be a strong ethical definition of what an AI system should and should not decide. In particular, the responsibility for a system failure needs to be clarified. Along with such design principles, all AI systems should be able to explain the reasons, at an adequate level of detail, of why they made certain decision and what the determining factors were in the input and training data. Bias and explainability are essential requirements linked to ethics and liability.

Security aspects are crucial, particularly in cybersecurity. A successfully tested AI system should be verified as safe against known adversarial attacks. The performance of the algorithms and their adaptation needs to be verifiable while maintaining individuals' data privacy and the intellectual property rights of the system developers. The verification process should be as transparent and explainable as possible. Thus, the key components of the verification and validation procedures should be based on open source systems.

There are numerous incentives for Switzerland to establish a place to define requirements and verify AI systems. Given its neutrality and reputation as a mediator, Switzerland is in a very favourable position. Being known worldwide for excellence in product quality and services, trust in Switzerland as a reliable state with profound ethical standards is high and could also apply to Swiss-verified AI systems. Being small and independent, Switzerland is more agile than other European countries. In addition, Switzerland has internationally known research institutes in the field of AI, and various large technology companies that are leaders in AI have research facilities here. Bringing together industry, research institutes and technology companies as well as governmental institutions to set up a requirements and verification process is demanding but crucial and in the interest of all the involved parties.

Based on the statements above, we recommend defining an **institution at the governmental level** in Switzerland to assume responsibility for the requirements and verification of AI systems.

- The institution should provide clear but flexible **general guidelines** for testing algorithms.
- The guidelines should be extended by **industry-specific standards** and include different levels of requirements depending on the criticality of the use cases and the severity of the consequences in case of failure.
- Standards and guidelines can be verified by existing industry-specific service providers and testing bodies.
- The proposed body could be organised as an extra-parliamentary committee[41] or existing verification bodies[42].

An **expert committee** consisting of academic and industry specialists in AI core technology, industry use cases, security, societal implications and ethics should be installed to develop an **implementation plan** for such a requirements and verification institution. SATW would be willing to organise and coordinate such a committee in collaboration with a government body.

[41] www.admin.ch/gov/de/start/bundesrecht/ausserparlamentarische-kommissionen.html

[42] National Institute of Standards and Technology, U.S. Department of Commerce. www.nist.gov/topics/artificial-intelligence

Increase Human Trust in AI

An ambitious but necessary objective[43]

The availability of affordable, highly parallel computing, vast amounts of training data, and economic interest from large corporate actors in "big data" has led to large-scale AI systems capable of solving problems that were, until recently, considered unsolvable for at least a decade, if ever.

It is predicted that these novel and complex AI technologies will impact virtually every human activity, from health and education to manufacturing and warfare, and will become central in the lives of citizens and consumers.

In this context, it will be critical to improve the AI image in public discourse; hence, we are also focusing on "human trust in AI" by studying how to improve the multifaceted trust between human users and "AI systems" by making the systems more robust, more intelligible to human minds, and more aligned with society's needs. This ambitious goal should develop along three complementary axes:

Trust in AI as an abstract object. A clear quantification of AI's technical performance is fundamental to allowing humans to anticipate its failures and conversely to trust it when it operates in a properly functioning regime, just as one constantly infers the behaviour of other humans to safely interact with them. Research and development should place particular emphasis on the challenge of determining the uncertainty in the learned knowledge of AI agents. How can we specify AI's expected performance? How can we ensure that this level of performance is reached? Additionally, at a more fundamental level, how can we create a common understanding between humans and AI?

Trust in AI as an operational system. Can we trust that there will be proper handling of the required mass of data, particularly regarding privacy laws and intellectual property, or that the resulting systems will be robust to adversarial attacks, as is expected from any critical software? Security is a central topic in practice, as an AI system remains a complex piece of software running in a failure-prone and vulnerable physical device.

Trust in the impact of AI on society. Our common sense has been shaped over thousands of years to handle the risks and benefits of dealing with either objects devoid of cognition or intelligent human beings. AI creates a new category to which our most fundamental behaviours have not (yet) been adapted. How can we prepare citizens for pervasive intelligent systems that will influence them in all their decisions, including personal and societal decisions? How can we ensure that AI follows basic ethical principles and adapts to societal changes? How can we forecast the disruptions AI may cause to the economy?

Additionally, we believe that reproducibility and openness in the context of extremely complex AI models and large-scale data corpora are critical and require a strong commitment to and investment in "open science" tools and practices that will be central to the project.

[43] **Raise Human Trust in AI – lead author: Hervé Bourlard**

The author acknowledges Dr. François Fleuret from the Idiap Research Institute for his inputs to this section.

How to improve human trust and AI in publics

It is clear that Switzerland has reached a critical mass in regard to the technical expertise required to improve AI and make it trustworthy and safe. However, at the national level, there is no initiative that brings the different players together in a common forum that could act as a force multiplier. Even more problematic is the scattered expertise and the lack of coordinated research on the implications that AI will have for various aspects of our society, ranging from the economy to law to ethics, which is a situation that, nevertheless, reflects what is typical of AI research worldwide.

In this context, the expert committee proposed for the verification of AI systems would have the right to take a leading role in increasing human trust in and public awareness of AI. Efforts should be focused on sharing positive messages about the impact of AI on society, increasing the transparency of AI solutions by generally requiring explainability in software, and finally encouraging societal debates about wider social consequences of AI technologies.

Share positive AI messages[44]:

Empowering humans: AI will empower humans by realizing technologies that benefit humanity instead of destroying and intruding on the human rights of privacy and freedom to access information.

Open source information[45]: Open source information and AI collections will provide opportunities for new technological progress and global technological parity.

Content creation and information access: AI will play a fundamental role in content creation, enabling easier access to all kinds of data for everyone. As a consequence, knowledge will be "democratized", potentially also improving the enforcement of data privacy.

Positive changes in the quality of life: There are a growing number of AI applications that actively improve people's lives and create positive changes in the world. For instance, AI systems will help improve clinical diagnoses (e.g., applying AI to various types of healthcare data, cancer detection), medical prevention, or the diagnosis of faults in physical systems. AI also has the potential to improve farming and food production processes. Finally, the security of people (including children's safety), systems (communication), and criminal activity prevention will be improved.

More efficient society: Through AI, society will have the potential to become more efficient, for instance, by optimizing traffic and transportation systems, reducing waste, and improving ecological behaviours. AI has great potential to support human society in facing the challenge of climate change and other global threats.

[44] 14 Ways AI Will Benefit or Harm Society, Forbes Technology Council, www.forbes.com/sites/forbestechcouncil/2018/03/01/14-ways-ai-will-benefit-or-harm-society

[45] Top 8 Open Source AI Technologies in Machine Learning, opensource.com, opensource.com/article/18/5/top-8-open-source-ai-technologies-machine-learning

Increase transparency and encourage societal debates[46]:

Embed transparency in software: Sophisticated AI often takes complex and sometimes obscure routes (software, data, etc.) in its methodologies, data mining, and algorithms. We should encourage ways to embed transparency and clarity into the software (as well as hardware). For instance, when AI makes life-changing judgments around sentencing, welfare, and medical decisions, it is fair to assume that those affected need to know how these decisions were made.

Address multifaceted AI concerns: Encourage debates across society to ensure that we use the various AI technologies (which are changing the way we behave, work and interact with others) in ways that are ultimately beneficial and that consider many kinds of concerns.

Clearly identify long-term impacts of AI developments: The technology industry must recognise the long-term impacts of AI developments and assume responsibility for the wider social consequences of its work. This goal requires policy makers to start thinking seriously about the challenges that may arise from the advancement and increasing application of AI across all industrial sectors.

Engage business and citizens in discussions: Businesses and citizens need to become more broadly engaged in the discussions, recognising their common stakes in defining the future. Thanks to AI, we have the opportunity to redesign many aspects of our world and to make the *most* of the *opportunities* provided *by new technologies.* We must focus on making these changes in ways that we all trust and that enhance the quality of life and wellbeing of individuals and society rather than in ways that damage lives.

Focused studies, monitoring and analysis[47]: AI guidance should be developed through focused and multidisciplinary studies, monitoring, and analysis. Maximizing the impact of AI developments and their acceptability by society will indeed require engagement with cross-disciplinary groups, including computer scientists, social scientists, psychologists, economists, and lawyer.

Increase transparency and data privacy: In that direction, the EU General Data Protection Regulation (GDPR)[48] is probably the most important change in data privacy regulation in 20 years.

[46] Could transparency make AI safe and reduce public fears ? TechWorld, October 2017, www.techworld.com/data/could-transparency-make-ai-safe-reduce-public-fears-3665487/

[47] AI, People, and Society, Eric Horvitz, *Science,* July 2017: Vol. 357, Issue 6346, science.sciencemag.org/content/357/6346/7

[48] eugdpr.org

AI in Higher Education

A core research topic[49]

Machine learning and AI are currently fundamentally changing the way we think about computing at large, especially how we approach algorithm design, software engineering and computer systems engineering. Modern intelligent decision-making systems process and interpret high volume data streams for computer vision, natural language and speech processing, robotics, and health data analysis and, more generally, for flexible data-centric model building in engineering, natural science and, for the first time, in the humanities. Such systems commonly process heterogeneous data with substantial uncertainty caused by measurements. Classical concepts of how to establish the correctness of planning and decision-making algorithms when they process uncertain data often rely on worst-case guarantees that do not seem to be adequate and predictive enough for many real-world applications. Self-adapting AI algorithms must generalize well over fluctuations and model misspecifications; they should be resilient by proper regularization to avoid overfitting, i.e., "reading the tea leaves". On the other hand, versatile intelligent systems should also be sufficiently flexible to exploit the signal in data to its full extent for prediction. Intelligent decision-making systems should provide (provable) guarantees to deliver typical answers with a high predictive value; however, scientifically, it is still mostly unknown what that means for algorithm design, and we need large research programs to make progress on these questions beyond the standard paradigm of supervised machine learning. Research on these fundamental questions will also strengthen public trust in this technology that is about to be deployed for autonomous transport, for health care applications, for societal infrastructure management and for public and private services, including entertainment.

[49] **AI in Higher Education – lead author: Joachim Buhmann**

AI education should be modernized

Where does higher and continuous education stand in this scientific revolution? Computational thinking has clearly amended the scientific method, the unforeseeably successful methodology of industrialized societies, by adding a third pillar to empiricism – experimentation, theory building and computational modelling. In light of the big data challenge, we must revisit all aspects of computational thinking and fuse it with probabilistic reasoning. Lessons from the history of AI indicate that the reduction of intelligent behaviour to logical calculus delivers successful expert systems but can also be interpreted as one of the causes of the AI winter since neither the scientific community nor the general public has seen its high expectations of AI fulfilled. Human intelligence clearly employs rational logical calculus, but it seems to be even more governed by our subconscious abilities to imitate intelligent behaviours and, more importantly, by our capacity to teach such behaviours to other human beings across generations. Deep learning, with its inherent connectionism, emulates this style of thinking as opposed to the symbolic reasoning strategy and connects to the age of cybernetics before symbolic AI. We see predictive patterns in data apparently without using a consciously accessible, "rationally accessible" theory; for example, medical doctors are top experts in using their experience to achieve successful benefits in health care with often very little theoretical understanding that would explain the success in diagnosis, prognosis and therapy.

Current educational programs in higher education have to evolve to strengthen probabilistic computational thinking. Such a mindset will be vital for our future society since it will largely determine the value generation of future products and service. Are we prepared for these challenges in higher education? Current curricula in computer science emphasize a scientific methodology for algorithms and systems design and their analysis, which mostly focuses on the efficient usage of computational resources such as memory requirements and processing time. Data science, as a necessary prerequisite for intelligent reasoning, planning and acting in the real world, demands resilient algorithms and systems design in computation. These two scientific cultures must be fused into a coherent picture of computational reasoning and inference. Resource efficient computation should be balanced with the predictive power of computational solutions. In many applications, this trade-off leads to a win-win situation in which superior predictions can be computed in a more efficient way than with traditional algorithmic approaches.

Getting our students and workforce ready for AI

Are our students in higher education well prepared for this epochal paradigm shift? The answers to this question depend on their specialization of studies and the role of model building by computation in the respective field. Required knowledge and skills could range from the developer level, where novel AI solutions are created and deployed, to various grades of the expert user level, where there is creative usage of AI tools in various application domains. Science, technology, engineering and mathematics (STEM) students mostly acquire a solid basis for this paradigm shift due to rigorous mathematical training.

Future curricula reforms should be assessed according to criteria that strengthen these educational goals. The long tradition of deterministic computational thinking has to be complemented by a rigorous education in probabilistic modelling and the design of probabilistic algorithms that can deliver typical solutions and readily adapt to changing input data properties, demonstrating at least traces of intelligence. In addition, a knowledge of systems theory and (optimal) control should also be substantially required in disciplines outside of engineering.

The evolution of the scientific method by probabilistic computational thinking has to influence curricula in all scientific disciplines, ranging from physics to theology. Digitalization and intelligent information processing affect nearly all areas of human thought and will leave a profound footprint on all disciplines. Novel ways to conduct research are expected to emerge in the humanities when intelligent search algorithms can scan large digital libraries orders of magnitude faster than human investigators; sociological experiments can be conducted at the scale of societies. Research strategies from the natural sciences and engineering might substantially enrich the method catalogue of the humanities, and conversely, a novel common language will also enrich the STEM disciplines with advances in computational humanities. As a society, we should make particular efforts to empower all students in higher education to participate in these societal transformations.

What are the needs of the workforce in the private and public sectors? AI technologies have enriched the start-up culture with novel models of collaboration and communication. Work teams are now forming over virtual platforms, and continuous education might be provided through such platforms. AI tools affect the job profile of intellectuals and highly skilled brain workers. The critical thinking skills of such employees will be essential in this transformation when algorithms start to replace experts in their decision-making roles. Imagine a radiologist interacting with an automated radiology scoring program: for various tasks, the algorithm already delivers superior performance in scoring radiology images due to its speed and enormous storage capacity. The humans in the loop have to be able to evaluate the machine’s decisions, and such a competence for quality assurance requires profound computational training. Continuous educational programs will serve such an educational need. In particular, aging societies should allocate the necessary resources for continuous education to remain competitive with societies of developing countries that are usually characterized by much younger populations than the highly industrialized world.

Enable AI for SMEs

Create an appropriate environment[50]

AI has not been an important subject in higher education in Switzerland. Few educational courses have been available in the last 20-30 years, and those only showed single aspects of AI in light of a particular approach. Furthermore, computer science has not been viewed as a strong and important field of its own but often been subsumed under other sciences. Only very recently has this approach started to change, for example in the "Lehrplan 21".

Consequently, we do not find much representation of computer science skills among the management of SMEs in Switzerland. Notably, a deep understanding of algorithms, modelling, or software architecture is a rare skill that is highly in demand. This lack of skills makes it very difficult for SMEs to assess what AI can truly do to bring innovation to their businesses and what type of problems can be solved by the current state of AI technology. Many industry representatives are deeply influenced by the deep learning hype and believe that this technology is all that AI is about and that it will solve all problems. Furthermore, the hysteria about AI as a job killer encourages many companies to prefer to look into optimization and automation scenarios but less into innovation. Becoming more efficient and flexible is important; however, such a mindset does not necessarily lead to innovative products and services.

Numerous companies struggle to understand the impact of digitalization and novel technologies on their business. These companies need to link technology (computer science, AI and others) to the transformation scenarios that may turn out to be highly relevant for them. For example, consider booking.com (or uber.com), where brokers have taken over a market and created a difficult situation for the local Swiss hotel industry. A similar effort is currently underway in Switzerland to establish a brokering service for access to AI-related technological know-how. Such efforts are rarely beneficial to SMEs, which have limited budgets.

[50] **Enable AI for SMEs – lead author: Jana Koehler**

Concrete recommendations to enable AI for SMEs

- Offer courses for postgraduate education that teach AI as part of computer science and offer a broad and adequate picture of the technology.

- Adequately inform SMEs about AI safety, technological limitations and risks.

- Show examples[51] of innovative and deployed applications of AI to broaden the understanding of the opportunities that AI can offer (do not oversell current laboratory prototypes that are present in the media, e.g., in medical diagnosis).

- Understand how AI, digitalization, industry 4.0, and novel business models interact for SMEs[52,53].

- Link AI with economics and show transformation scenarios that help SMEs to reposition themselves if necessary.

- Describe and illustrate how AI and other technologies will impact typical professions in Switzerland and show what new professions will be created; link this information to educational offers.

- Start with key industries in Switzerland to drive the transformation of the Swiss economy through AI.

- Help SMEs look beyond the hype and provide methods that enable them to evaluate technology based on their needs and opportunities.

- Link to international sources to make it easier to learn about AI applications created by SMEs in other countries.

- Create a network of experts and knowledge resources that SMEs can access at low cost without paying high membership or brokering fees.

- Create one strong industrial representation in IT to support the exchange of knowledge and networking as opposed to the current situation with scattered IT industry associations.

[51] Industry 4.0 – Germany Market Report and Outlook. www.gtai.de/GTAI/Navigation/EN/Invest/Service/Publications/business-information,t=industrie-40--germany-market-report-and-outlook,did=917080.html

[52] Statusreport "Arbeitswelt Industrie 4.0". shop.vde.com/de/statusreport-arbeitswelt-industrie-40-download

[53] Smart Service Welt – Internetbasierte Dienste für die Wirtschaft. Deutsche Akademie Der Technikwissenschaften. www.acatech.de/Projekt/smart-service-welt/

satw it's all about technology

Schweizerische Akademie der Technischen Wissenschaften SATW
St. Annagasse 18 | 8001 Zürich | 044 226 50 11 | info@satw.ch | www.satw.ch