Vereinigung der
Schweizerischen Hochschuldozierenden

VSH
AEU

Association Suisse
des Enseignant-e-s d'Université

Bulletin

Artificial intelligence— Scientific Foundations and Societal Implications

Mit Beiträgen von | Avec les contributions de

Mike S. Schäfer & Julia Metag

Maximilian Janisch

Rico Sennrich

Bastian Grossenbacher-Rieck

Manuel Mondal, Philippe Cudré-Mauroux
& Julien Audiffren

Mauricio Reyes, Manuela Eugster, Stavroula Mougiakakou,
Tobias Nef & Raphael Sznitman

Isabelle Wildhaber & Frédéric Barth

Pierre Dillenbourg

Hans-Georg Fill & Jesús Muñoz Cádiz

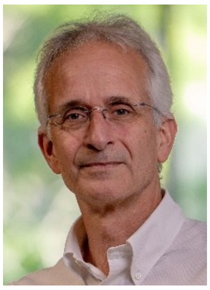
VSH-AEU-Bulletin 1

January 2026

50th Year

ISSN 1663-9898

Visual Design:
Damaris Pfeffinger, Graphic Designer, Basel
damaris.pfeffinger@gmx.ch | @damarispfeffinger



EDITORIAL

Christian Bochet
Président VSH-AEU

Photo credentials : Christian Doninelli

Chèr·e·s membres de la VSH-AEU

Le début d'une nouvelle année est toujours une bonne occasion de s'y projeter et de tenter d'esquisser ce qu'elle va nous apporter. Bien sûr, je vous la souhaite pleine de bonheur, de succès et de bonne santé ! Je pourrais jouer à Cassandre, et m'étendre à nouveau sur les coupes budgétaires de la Confédération, et sur les conditions de plus en plus restrictives d'accès aux subsides du FNRS. Je m'y abstiendrai, mais soyez assuré·e·s que la VSH-AEU travaille activement à défendre les intérêts des enseignant·e·s, en particulier à travers swissfaculty. Ces dernières semaines, lors de plusieurs interviews dans des cadres aussi divers que la chimie, l'enseignement ou l'édition de revues scientifiques, m'a été posée la question : « quel développement majeur voyez-vous dans votre activité ? », et la réponse à chaque fois était : « l'usage de l'IA ». Ce n'est évidemment pas une coïncidence, et vous toutes et tous ont sans doute été confronté-e-s à cette réflexion. Et donc il nous semblait naturel de dédier un bulletin à cette brusque accélération de l'usage de l'IA depuis 2023.

Comme l'expriment un certain nombre des articles qui suivent, « intelligence artificielle » est un terme générique, qui englobe une multitude d'outils distincts, allant des LLMs (*Large Language Models*), tels que Chat-GPT, au MLs (*Machine Learning*), tels que AlphaFold, logiciel de modélisation de la structure des protéines, dont les auteurs ont reçu le prix Nobel de chimie en 2024. Ces divers outils sont très différents, avec leur propre profil d'avantages et d'inconvénients. On a de la peine à imaginer les risques qu'un biologiste ait à utiliser AlphaFold (bien sûr autre que de simplement se tromper !), alors que les bénéfices peuvent être massifs. A l'autre extrémité du spectre, l'usage aveugle d'un LLM par des enfants, sans en comprendre les limitations, peut représenter des risques immenses. L'actualité regorge d'exemple d'abus, tels que les biais, la désinformation ou la pornographie, pour n'en citer que quelques-uns.

Les articles proposés dans ce bulletin couvrent un éventail assez large de perspectives, allant d'aspects techniques aux aspects légaux, sans oublier quelques éléments d'histoire de l'IA. L'éducation étant au cœur de l'activité de notre Association, l'usage, l'utilité et surtout l'impact de l'IA sont également abordés. En particulier, l'un des articles utilise le qualificatif de « prothèse cognitive », terme qui en illustre parfaitement l'utilité et le danger. Dans cette perspective, et les lecteurs me pardonneront les anglicismes, les concepts de *de-skilling, un-skilling* et *up-skilling* sont discutés, se référant respectivement à la perte, au manque d'acquisition et au renforcement des compétences.

Conclusion en forme de clin d'œil, cet éditorial n'a bénéficié d'aucune prothèse cognitive, ni dans sa conception, sa rédaction et même sa correction. Comme dans l'alimentation, où certains produits sont garantis bio par un sigle approprié, l'usage d'un formalisme similaire devrait se généraliser.

(Crédits: Vecteezy.com)

Christian Bochet

Artificial Intelligence |
Scientific Foundations and Societal Implications

Increasingly used, Sceptically viewed: The Role of Artificial intelligence in Science Communication in Switzerland

Mike S. Schäfer (University of Zurich) & Julia Metag (University of Münster)

Introduction

Is climate change real? Do vaccinations work? Are there microplastics in drinking water? Questions like these are asked millions of times every day to AI chatbots powered by large language models (LLMs) such as ChatGPT, Google Gemini, or Microsoft Copilot. People use these systems to summarise complex information, get writing support, and answer questions they have. Generative AI has rapidly become embedded in daily life. As AI tools become more frequently used for routine information seeking, their spread raises far-reaching political, economic, and social questions for individuals, organisations, and society – and they increasingly shape how scientific knowledge is accessed, presented, interpreted, and ultimately trusted.[1]

AI is not only becoming increasingly important within science – as an analytical and methodological tool and as an object of research – but also for science communication, as it is turning into a central channel through which many people encounter scientific claims. This creates major opportunities: automated content creation, translation, and multimodal, dialogic formats can increase efficiency, tailor complexity to diverse audiences, and potentially enhance engagement (e.g., Baake *et al*., 2025[2]; Kessler *et al*., 2025[3]). Yet these benefits come with substantial risks, including unreliable or low-quality outputs, biases embedded in training data or processes, large-scale misinformation, as well as new structural pressures on science communicators and science journalism, such as shifting roles, dependencies on intransparent digital infrastructures, and job insecurity (Guenther *et al*., 2025[4]; Schäfer, 2023[5]).

Since recent research has pointed to cross-national differences in the perception and use of AI for science communication (Greussing *et al*., 2025[6]), the Science Barometer 2025 examines how the Swiss population perceives and uses AI, and how it assesses AI's role in science and science communication.

The Science Barometer is based on a representative population survey conducted using online and paper questionnaires, drawing on the Swiss Federal Statistical Office's sampling frame for person and household surveys (SRPH).[7] Every three years, the Barometer covers the Swiss population aged 16 and older. In 2025, 1'548 respondents were questioned, and questions about AI were administered to half of them using a split-ballot design.

---

1 Generative AI was used to improve the language of the article. The authors thank Niels G. Mede for helpful comments as well as Xiran Liu, Gerta Lokaj and Damiano Lombardi who did part of the analysis reported here.

2 Baake, J., Schmitt, J. and Metag, J. (2025). Balancing realism and trust: AI avatars in science communication. *JCOM: Journal of Science Communication*, 24(02), A03. doi.org/10.22323/2.24020203.

3 Kessler, S. H., Mahl, D., Schäfer, M. S., & Volk, S. C. (2025). All eyez on AI: a roadmap for science communication research in the age of artificial intelligence. *JCOM: Journal of Science Communication*, 24(2), Y01. doi.org/10.22323/2.24020401.

4 Guenther, L., Kunert, J. and Goodwin, B. (2025). "Away from this duty of chronicler and towards the unicorn": how German science journalists assess their future with (generative) Artificial Intelligence. *JCOM: Journal of Science Communication*, 24(02), A06. doi.org/10.22323/2.24020206.

5 Schäfer, M. S. (2023). The Notorious GPT: science communication in the age of artificial intelligence. *JCOM: Journal of Science Communication*, 22(2), Y02.

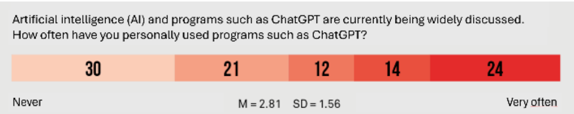
6 Greussing, E., Guenther, L., Baram-Tsabari, A., Dabran-Zivan, S., Jonas, E., Klein-Avraham, I., Taddicken, M., Agergaard, T. E., Beets, B., Brossard, D., Chakraborty, A., Fage-Butler, A., Huang, C.-J., Kankaria, S., Lo, Y.-Y., Nielsen, K. H., Riedlinger, M., & Song, H. (2025). The perception and use of generative AI for science-related information search: Insights from a cross-national study. *Public Understanding of Science*, 34(5), 599-615. doi.org/10.1177/09636625241308493.

7 For more information about the Science Barometer's background, methods, questionnaires and results see www.wissenschaftsbarometer.ch.

The following section presents key data on AI use among the Swiss population and on attitudes toward AI-supported science communication.

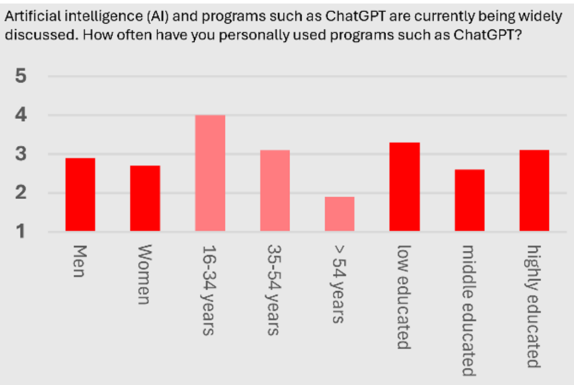
How frequently does the Swiss Population use AI in general?

AI is used by a considerable portion of the Swiss population. When respondents were asked about their general use of AI tools such as ChatGPT, i.e. not limited to retrieving information about science, 70% reported using AI tools at least sometimes, albeit only 38% reported using them frequently or very frequently. This pattern is consistent with findings from other surveys, which show that in many countries, considerable parts of the population have tried these tools out, but that heavy users still represent a minority, though a growing one (Microsoft, 2025[8]; Greussing *et al.*, 2025, p. 604).



Examining sociodemographic differences in AI use reveals patterns in Switzerland that mirror those observed in other countries:

- Male respondents (Mean = 2.9) use AI tools more than female respondents (2.7).
- Younger individuals aged 16 to 34 years (4.0) use AI significantly more frequently than those aged 35 to 54 (3.1) and 55 and older (1.9).



By contrast, patterns related to education are less straightforward:

- Respondents with lower levels of formal education (obligatory school or no formal school qualification) report the most frequent AI use (3.3).
- Those with medium levels of education (secondary education) report the lowest use (2.6).
- Highly educated respondents (tertiary education) fall in between (3.1).

How competent do Swiss people feel and what do they know about AI?

In addition to use, the Science Barometer also measured the Swiss population's knowledge about AI using two complementary approaches that had been tested in other studies (e.g. Wang et al., 2023[9]): a self-assessment and knowledge-based quiz questions.

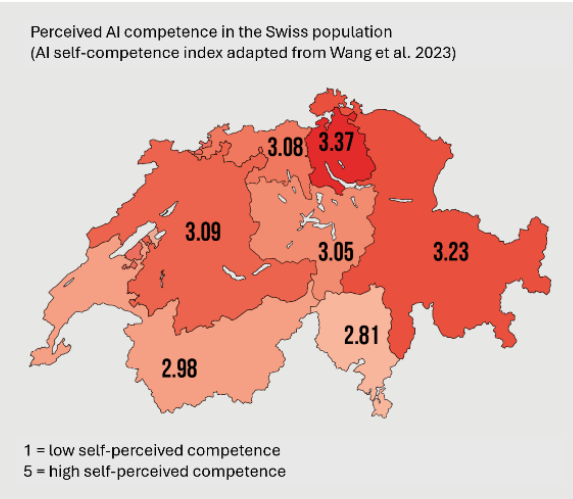
First, we assessed respondents' self-perceived AI competence. To this end, the survey included ten items (e.g., "I understand how applications like ChatGPT work" or "I can select the most suitable tools for a specific task from a range of AI applications") capturing how respondents evaluate their own competence with regard to AI. Responses to these items were aggregated into an index of perceived AI self-competence.

Overall, respondents' perceived AI competence is moderate (3.3), suggesting that a substantial share of the population still feels insecure about AI. There are no gender differences in perceived AI competence, with both men and women reporting identical mean values (3.3). Younger (3.6) and middle-aged respondents (3.5) are more confident than older respondents (2.9). Respondents with compulsory education – who report the most frequent AI use – also indicate the highest levels of confidence (3.5), followed by those with tertiary education (3.4) and secondary education (3.2) - but the differences are small.

[8] Microsoft (2025): AI Diffusion Report 2025. AI Economy Institute. https://www.microsoft.com/en-us/corporate-responsibility/topics/ai-economy-institute/reports/global-ai-adoption-2025.

[9] Wang, B., Rau, P. L. P., & Yuan, T. (2023). Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. *Behaviour & information technology*, 42(9), 1324-1337.

The results also show regional differences in self-assessed AI competence: respondents in the Zurich region and Northwestern Switzerland report the highest levels, followed by Eastern Switzerland, while self-assessments are lowest in Ticino. Overall, perceived AI competence is higher in urban areas than in rural regions.



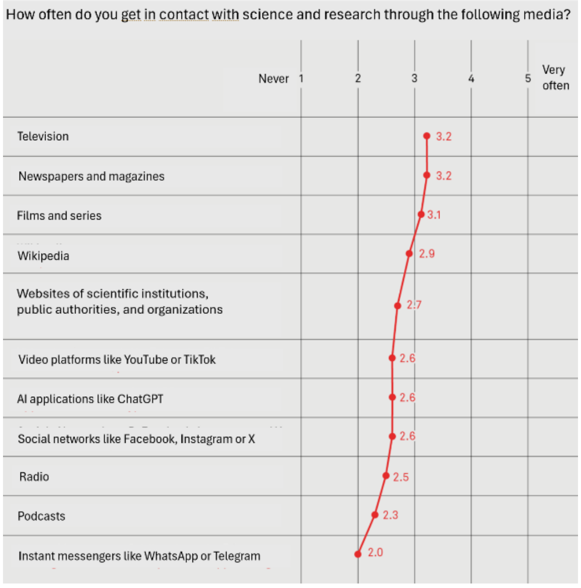
In addition to self-assessed competence, respondents were also asked three quiz questions to test their factual knowledge about AI. They were asked to indicate whether the following statements were true or false: "Chatbots process and understand language in the same way humans do", "People can easily recognise AI-generated language as artificial", and "Applications like ChatGPT generate responses based on probabilities".

All three items were answered correctly by a majority of respondents: between 74% and 83% per item. Men were slightly more likely to answer correctly than women, and respondents with tertiary education performed best on average, consistent with education being associated with more accurate assessments of how AI works.

The Growing Importance of AI as a Source of Science-related Information

The Science Barometer indicates that AI has become part of many people's information repertoires, not only in general, but also with regard to science and research. "AI applications like such as ChatGPT" are already relevant sources of science-related information in Switzerland, even though they have not yet become dominant for these topics.

When Swiss citizens are asked how frequently they encounter information about science and research from different sources, AI applications rank in the middle of the field. On a five-point scale from 1 ("never") to 5 ("very often"), AI tools reach an average of 2.6, placing them on a similar level as "Video platforms like YouTube and TikTok" and "Social networks, e.g. Instagram, Facebook and X", respectively. Nevertheless, they remain less important than television (3.2), print media (3.2), films and series (3.1), interpersonal communication with friends and family (3.0), and Wikipedia (2.9). But they already surpass established sources such as radio (2.5), museums and exhibitions (2.4), and digital formats such as podcasts (2.3) or instant messengers like WhatsApp or Telegram (2.0) as sources of science-related information overall.



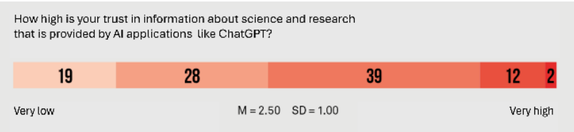
This positioning is notable, given the relatively recent public diffusion of generative AI tools. At the same time, the data clearly show that AI is not yet a central source for the majority of the population, but rather an emerging and increasingly relevant one. AI's current role in the broader information ecosystem around science can best be described as complementary: It is used alongside, rather than instead of, traditional news and institutional sources.

It is also notable that age differences are pronounced. As with many digital formats, younger respondents report substantially more frequent use of AI applications as a source of science-related information than older people. This mirrors broader patterns of media use and likely signals that AI-based sources will gain further importance over time as younger cohorts of users get older.

Expectations, Perceived Risks and Benefits of AI for Science Communication

While AI tools have become a relevant source of science-related information for many people in Switzerland, it is notable that their growing use does not translate into equally high levels of trust.

When asked how much they trust scientific information provided by AI, only a small minority of Swiss citizens express high (2%) or "very high" (12%) trust. In contrast, around half of respondents indicate that they trust such information rather little or "not at all" (47%). This gap between widespread use and lack of trust suggests a pattern of cautious and rather instrumental engagement: Many people turn to AI tools but do so with considerable reservations about their reliability as sources of scientific information.



This ambivalence is also reflected in expectations regarding what AI can and cannot contribute to science communication. On the one hand, respondents clearly see practical benefits: A majority agree that AI applications can help explain complex scientific issues in a simple and accessible way. This expectation is particularly pronounced among younger respondents and those with higher levels of formal education, who tend to report greater confidence in their own ability to assess and contextualise information. AI is thus perceived as a potentially useful translation tool, lowering barriers to understanding scientific content – at least for some.

At the same time, respondents are much more sceptical about AI as a substitute for interpersonal communication about science. Only a small minority agree that AI could replace conversations with real people about scientific topics. Most respondents reject the idea that AI could meaningfully replicate dialogue, exchange, or deliberation – all dimensions widely seen as central to good science communication. This scepticism is shared across demographic groups, and especially strong among older respondents, suggesting that AI is often viewed as a complementary rather than primary channel.

In addition, concerns about risks clearly outweigh optimism in key areas: A large majority of Swiss respondents agree that AI makes it harder to distinguish whether scientific content was produced by humans or by machines. Even more pronounced is the concern that AI applications sometimes reproduce false or misleading scientific information. These worries are widespread across gender, age, and education groups, indicating a broadly shared awareness of current limitations such as hallucinations, biases, and the overall opacity of AI systems. Notably, respondents with higher levels of education tend to be slightly more critical, which may reflect greater familiarity with scientific standards and uncertainty.

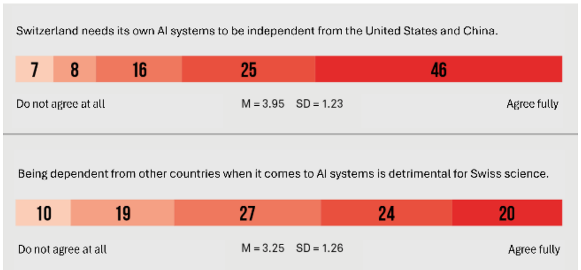
Beyond immediate information quality, the Science Barometer also captures the population's expectation related to AI's effects on access to science-related information and participation. Many respondents see AI as a tool that can support individual learning and exploration: A sizeable share of respondents agrees that AI will enable them to pursue their own inquiries into topics of interest or obtain answers to questions they previously felt unable or unwilling to ask. These expectations are again strongest among younger people and respondents with tertiary education. However, optimism declines markedly when it comes to more active forms of participation. Only a minority believe that AI will make it easier for them to participate in scientific projects or meaningfully influence research agendas. In other words, AI is expected to support individualised access to knowledge more than collective engagement or co-production.

This pattern points to a limited democratisation potential as of now. AI is seen as expanding access to information and lowering entry barriers, particularly for those who already possess some digital and educational resources. At the same time, respondents remain cautious about whether these tools genuinely broaden participation or simply reinforce existing inequalities in skills and confidence.

AI and Swiss Society: Public Views on Sovereignty, Governance and Control

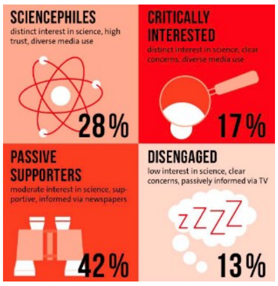
Finally, attitudes toward AI in science communication are closely intertwined with broader societal and political considerations – most notably questions of control, dependency, responsibility, and regulation. A clear majority of Swiss respondents (70%) support the idea that Switzerland should develop its own AI applications to remain independent from the United States and China. This stance may reflect less a form of "tech nationalism" than concerns about reliable infrastructures and clearer accountability for systems that increasingly shape information environments, including science-related ones. Recent initiatives such as Apertus, the open large language model developed by researchers at ETH Zurich, EPFL, and the Swiss National Supercomputing Centre (CSCS), illustrate how such concerns translate into concrete efforts to strengthen national capacity while embedding AI in public and scientific governance structures.

Concerns about dependency on foreign AI systems are also widespread among respondents, though less pronounced than the wishes for sovereignty: 44% agree that such dependency harms Swiss science. Here, regional differences emerge: agreement is somewhat higher in the German-speaking part of Switzerland than in the Romandie, with Ticino in between.



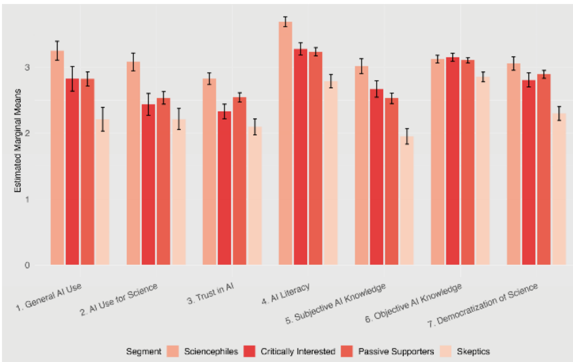
At the same time, respondents clearly reject technocratic governance. The statement that AI developers "know best" what is good for Switzerland's future is strongly opposed (77% disagree). Notably, respondents aged 34 years and younger disagree most strongly, suggesting that calls for democratic oversight and public steering are particularly salient among younger cohorts.

How Different Groups of the Swiss Population Use and Perceive AI in Science Communication



Overall, we find differences in AI use and perceptions across four segments of the Swiss population that differ in their attitudes toward science and have been identified repeatedly in the Science Barometer since 2016 (Schäfer *et al.*, 2018[10]; Koch *et al.*, 2020[11]): the "Sciencephiles," who hold very positive attitudes toward science (22% of the population); the somewhat more cautious "Critically Interested" (12%); the large group of "Passive Supporters," who generally support science but from a distance (48%); and the "Sceptics," who are disengaged from or even opposed to science (17%).

The 2025 results show systematic differences between these groups. "Sciencephiles" consistently report the highest levels of AI use – both in general and for obtaining science-related information – as well as the highest levels of trust in AI-provided science information, AI literacy, and perceived AI knowledge. "Sceptics", by contrast, report the lowest levels across all of these dimensions.



The "Critically Interested" and "Passive Supporters" are largely indistinguishable on these measures, suggesting that the key divide is less

10 Schäfer, M. S., Füchslin, T., Metag, J., Kristiansen, S., & Rauchfleisch, A. (2018). The different audiences of science communication: A segmentation analysis of the Swiss population's perceptions of science and their information and media use patterns. *Public understanding of science*, 27(7), 836-856.

11 Koch, C., Saner, M., Schäfer, M. S., Herrmann-Giovanelli, I., & Metag, J. (2020). "Space means Science, unless it's about Star Wars": A qualitative assessment of science communication audience segments. *Public Understanding of Science*, 29(2), 157-175.

between "supportive yet critical" and "supportive yet distant" publics, and more between science-affine groups on the one hand and science-sceptical groups on the other.

Differences between the groups are most pronounced for AI literacy, where nearly all segments differ significantly from one another (with the exception of the "Critically Interested" and "Passive Supporters"). Interestingly, the pattern is weaker for objective AI knowledge, where segment differences are smaller and do not fully mirror self-reported measures.

Overall, these gaps qualify overly optimistic assumptions about generative AI as an equalizing force: rather than uniformly "democratizing" science communication, AI appears to map onto – and potentially amplify – existing differences in engagement, confidence, and trust.

Conclusion

Overall, frequent AI use in Switzerland is still concentrated in specific population segments, but it is gradually spreading across society. At the same time, many respondents report moderate confidence in using AI and demonstrate a basic level of factual knowledge about how such systems work.

Taken together, these findings depict a population that engages with AI pragmatically but critically. AI is increasingly used – also for science-related information – and is valued for specific functions in science communication, particularly for explanation, accessibility, and individual exploration. Yet trust remains limited, perceived risks are salient, and expectations of empowerment and participation are cautious. Importantly, AI use and perceptions differ substantially across the population, with science-affine and younger groups reporting higher use, higher confidence, and somewhat higher trust than more sceptical and older groups.

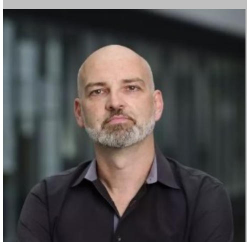
For science communication, this suggests that AI tools are unlikely to replace established practices or interpersonal exchange – but they are already reshaping how people encounter, explore, and evaluate scientific information.

University of Zurich, IKMZ – Department of Communication and Media Research, Andreasstrasse 15, 8050 Zurich

Email: m.schaefer@ikmz.uzh.ch

Website: www.ikmz.uzh.ch/en/research/science-crisis-and-risk-communication/mike-s-schaefer.html

ORCID: 0000-0002-0847-7503



Mike S. Schäfer, Dr. phil., is a Full Professor of Science Communication at the University of Zurich. He is also the Director of CHESS – Center for Higher Education and Science Studies and Co-PI of the Science Barometer Switzerland.

Schäfer's research focuses on public communication about and public perceptions of science, as well as the role of AI in science communication. He has published widely on these topics and co-edited the "Oxford Encyclopedia of Climate Change Communication" as well as the handbooks "Forschungsfeld Wissenschaftskommunikation" and "Forschungsfeld Hochschulkommunikation".

Schäfer is an elected member of acatech – German National Academy of Science and Engineering.

Photo credentials: John Flury

University of Münster, Department of Communication, Bispinghof 9-14, D-48143 Münster

Email: julia.metag@uni-muenster.de

Website: www.uni-muenster.de/Kowi/julia-metag.shtml

ORCID: 0000-0003-4328-6419

Julia Metag, Dr. phil., is a Full Professor of Communication Science at the University of Münster, Germany. She also leads the Hot Science Communication Lab, which is dedicated to researching the interface between multimodality and controversy in science communication.

Her research focuses on the interactions between public communication and the social fields of science and politics. She examines public communication about science, public perceptions of science, as well as multimodality and controversy in science communication. She has published widely on these topics and has co-edited "The Handbook of International Trends in Environmental Communication" and the handbook "Forschungsfeld Hochschulkommunikation".

Metag is an elected member of the Academia Europaea (The Academy of Europe).

Photo credentials: Nadine Daum

From Lord Kelvin to AlphaFold: The Mathematical Foundations and Scientific Applications of Artificial Intelligence

Maximilian Janisch

Artificial intelligence has become a powerful tool in both mathematics and science, yet its foundations rest on centuries of computational innovation. This article traces AI's evolution from mechanical calculators built to solve systems of equations to neural networks that can predict protein structures and assist in mathematical proofs. One of the first mechanical calculators was Lord Kelvin's string calculator, proposed in 1878 and constructed at MIT in 1934. It could solve systems of ten linear equations in seconds, a task that once required days of work by hand. Through milestones like Deep Blue's victory over Kasparov, AlphaGo's mastery of Go, and the algorithmic reconstruction of the first black hole image, we explore how mathematical theory, computational power, and data have together shaped AI's trajectory. The article closes with reflections on what these developments mean for researchers across disciplines and on the future role of AI in academic inquiry.

Introduction

Artificial intelligence is often presented as a recent phenomenon, yet its roots extend deep into the history of computation. From ancient astronomical devices to mechanical calculators and modern neural networks, the ambition to automate reasoning has driven innovation for centuries. This article traces that arc, highlighting how mathematical theory, computational power, and data have together shaped AI's trajectory.

What Is Artificial Intelligence?

The field of artificial intelligence is concerned with automating tasks we typically associate with human cognition: recognising speech, translating languages, solving mathematical problems. While the term itself dates only to the 1950s, the underlying aspiration to build machines that reason is far older. What has changed is not the fundamental mathematics, much of which was established decades ago, but rather our capacity to implement it at scale.

The Deep History of Computational Machines

The desire to mechanise calculation has ancient origins. The Antikythera mechanism, built around 100 BC, stands as the oldest known analog computer. It was designed to predict planetary positions and solar eclipses decades in advance. Discovered in 1901 in a Roman shipwreck off the Greek island Antikythera , this device contained at least 30 interlocking bronze gears, including a pin-and-slot mechanism to model the Moon's elliptical orbit. Machines of comparable complexity would not appear again until the astronomical clocks of 14th-century Europe.

In 1620, Wilhelm Schickard constructed the first mechanical calculator capable of addition, subtraction, multiplication, and division. Though the original was lost during the Thirty Years' War and is known to us only through correspondence with Johannes Kepler, it marked a decisive step toward automated computation.

Lord Kelvin's String Calculator

By the late nineteenth century, the practical demands of navigation and engineering spurred new computational devices. Lord Kelvin developed a tide-predicting machine that used ropes and pulleys to perform Fourier analysis, enabling accurate predictions for steamship travel[1]. In

---

[1] collection.sciencemuseumgroup.org.uk/objects/co53901/william-thomsons-tide-predicting-machine-1872.

1878, he proposed an even more ambitious device: a string-powered machine to solve systems of linear equations.

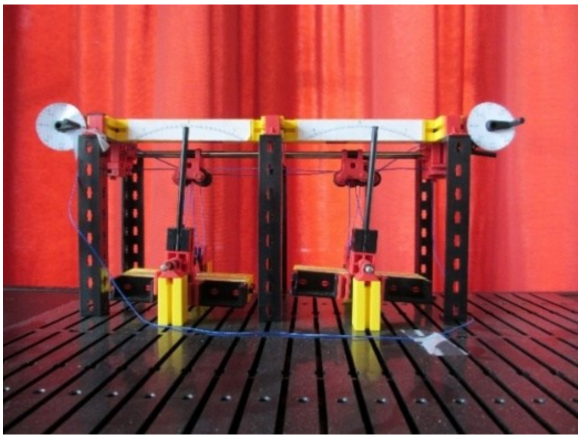


Fig. 1. *A small version of the string calculator, able to solve two linear equations with two unknowns, built by me in 2017. (Own work).*

This vision was only realised in 1934, when John Wilbur at MIT constructed a working version capable of solving ten simultaneous equations in mere seconds, a task that used to require days of calculation by hand. The economist Wassily Leontief later employed this machine to model input-output relationships in the American economy, work that contributed to his Nobel Memorial Prize in economics, won in 1973. The same mathematical problem, solving large systems of linear equations, would later underpin Google's PageRank algorithm, which ranks web pages by their relative importance.

How the String Calculator Works

The machine uses the *Strahlensatz* to perform division mechanically.

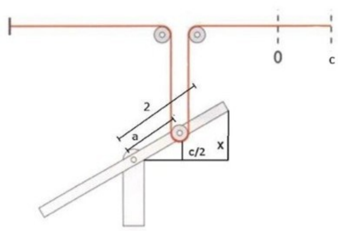


Fig. 2. *Illustration of how to solve the equation a x = c with ropes and strings. (Own work).*

Consider a seesaw balanced on a central pivot, with a rope running over a pulley attached to one arm. If the pulley sits at distance a from the pivot and the rope is pulled by length c, the seesaw tilts until the arm's endpoint reaches height x satisfying $ax = c$. The seesaw has computed the solution $x = c/a$.

To solve a system of two linear equations $a_1x + b_1y = c_1$ and $a_2x + b_2y = c_2$, Kelvin coupled two such seesaws. Each seesaw carries two pulleys (one for each coefficient), and two ropes thread through the system: one rope connects all pulleys corresponding to the first equation, the other to the second. Pulling the ropes by lengths c_1 and c_2 and letting the system reach equilibrium, the seesaw endpoints settle at heights x and y, the unique solution to the system, assuming that (a_1, b_1) is not a multiple of (a_2, b_2).

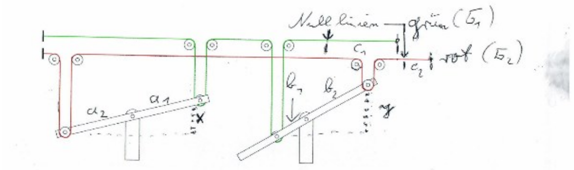


Fig. 3. *Solving a linear equation system with two equations. (Own work).*

This same mathematical operation, solving large systems of linear equations expressed as matrix equations, forms the computational backbone of modern neural networks. Every layer of a deep learning model performs matrix multiplication followed by a nonlinear transformation. The mathematics Kelvin mechanised with ropes in 1878 is identical to what modern GPUs execute billions of times per second when training models like AlphaFold.

The Rise of Machine Learning

The modern era of artificial intelligence began with the Perceptron,[2] introduced in 1958 as an algorithm for image recognition based on linear patterns. By itself it was not very useful, but the development of backpropagation in 1982 transformed the field. By using the chain rule from calculus to adjust a network's internal parameters based on output errors, backpropagation[3] made it possible to train deep

2 Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408. doi.org/10.1037/h0042519.

3 Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. doi.org/10.1038/323533a0.

neural networks, and pulled AI out of a prolonged winter.

As computational power grew and training techniques matured, automated algorithms began to outperform humans in specific domains. In 1997, IBM's Deep Blue[4] defeated world chess champion Garry Kasparov. In 2016, DeepMind's AlphaGo[5] prevailed against Lee Sedol in Go, a game whose complexity far exceeds that of chess. These victories demonstrated that AI could master domains once thought to require human intuition.

Scientific Breakthroughs Enabled by AI

Recent years have witnessed AI contributing directly to scientific discovery. In April 2019, the Event Horizon Telescope collaboration released the first image of a black hole,[6] located 55 million light-years away in galaxy Messier 87. Machine learning algorithms were essential for reconstructing coherent images from sparse, globally distributed telescope data.

DeepMind's AlphaFold[7] addressed one of biochemistry's most challenging problems: predicting how proteins fold into their three-dimensional structures. AlphaTensor[8] discovered new algorithms for matrix multiplication, while AlphaProof and later LLMs by DeepMind and OpenAI have won Silver and Gold medals at the International Mathematics Olympiad,[9] solving mathematical Olympiad problems at a competitive level.

Outlook

Large language models can now engage with advanced mathematics, albeit with limited reliability. When paired with formal proof-checking systems, they may soon become valuable collaborators for researchers. I am particularly excited about AI's potential in medicine, especially in diagnostics and drug discovery, where pattern recognition in complex data can accelerate progress.

The trajectory from Lord Kelvin's string calculator to AlphaFold reveals a consistent theme: machines that augment human reasoning, built on enduring mathematical principles. As AI capabilities expand, researchers across all disciplines will benefit from understanding both its possibilities and its foundations.

Email: mail@maximilianjanisch.com
Website: https://maximilianjanisch.com
ORCID: 0000-0002-7018-8833

Maximilian Janisch is a Swiss mathematician who completed his PhD at the University of Zurich in 2025 at the age of 21. His doctoral research, supervised by Prof. A. Nikeghbali and Fields Medallists A. Avila and A. Figalli, focused on probability theory with applications in particle physics, healthcare economics, and financial markets.

Before his PhD, he conducted research in analysis, culminating in the co-authored book *Instability and Non-uniqueness for the 2D Euler Equations*, after M. Vishik (Princeton University Press, 2024).

His research spans probability theory and machine learning.

Photo credentials: Joël Hunn

4 Campbell, M., Hoane, A. J., & Hsu, F. (2002). *Deep Blue. Artificial Intelligence*, 134(1–2), 57–83. doi.org/10.1016/S0004-3702(01)00129-1.

5 Silver, D., Huang, A., Maddison, C. *et al.* (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489. doi.org/10.1038/nature16961.

6 Event Horizon Telescope Collaboration *et al.* (2019). First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole. *The Astrophysical Journal Letters*, 875(1), L1. doi.org/10.3847/2041-8213/ab0ec7.

7 Jumper, J., Evans, R., Pritzel, A. *et al.* (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. doi.org/10.1038/s41586-021-03819-2.

8 Fawzi, A., Balog, M., Huang, A. *et al.* (2022). Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature*, 610(7930), 47–53. doi.org/10.1038/s41586-022-05172-4.

9 Luong, T., & Lockhart, E. (2025). Advanced version of Gemini with Deep Think officially achieves gold-medal standard at the International Mathematical Olympiad. *Google DeepMind Blog*. https://deepmind.google/blog/advanced-version-of-gemini-with-deep-think-officially-achieves-gold-medal-standard-at-the-international-mathematical-olympiad/.

Warum Sprachmodelle wie ChatGPT die «Eigenmietwertabschaffung» nur schlecht buchstabieren, aber gut erklären können

Rico Sennrich (Universität Zürich)

Sprachmodelle wie ChatGPT, neuerdings auch als generative KI-Modelle betitelt, haben in den letzten Jahren deutliche Qualitätssprünge gezeigt und hohe Erwartungen und Ängste geweckt. Bei der Diskussion, ob diese KI-Systeme übermenschliche Intelligenz zeigen oder bald zeigen werden, sollte aber bedacht werden, dass menschliche Massstäbe, wie viel Wissen oder Intelligenz eine Aufgabe erfordert, nicht direkt auf KI-Systeme übertragen werden kann.

In diesem Artikel möchte ich einen kurzen Einblick geben, wie Texte in Sprachmodellen repräsentiert werden, und welchen Einfluss das auf ihre Fähigkeiten hat.

Sprachmodelle

Sprachmodelle folgen heute noch dem Grundprinzip, das schon mindestens seit den 1940-er Jahren bekannt ist, und Grundlagen in der Informationstheorie von Claude Shannon hat.[1] Wenn wir einen Text als Sequenz von Ereignissen (Schriftzeichen, Wörter, Bytes, oder andere Symbole, neusprachlich "Tokens") formalisieren, können wir, basierend auf einem Trainingskorpus von Texten, Token für Token seine Wahrscheinlichkeit schätzen, aber auch für einen unvollständigen Text vorhersagen, wie er wahrscheinlich weitergehen wird. Kombiniert mit einem Zufallsgenerator generieren auch moderne Sprachmodelle wie ChatGPT Texte auf Grund solcher Wahrscheinlichkeiten.

Neuronale Sprachmodelle

Erste Sprachmodelle zählten die Häufigkeit von Tokenpaaren oder -Tripeln, ohne weiteren Kontext, um das nächste Token vorherzusagen, und waren dementsprechend schwach. Sogenannte Neuronale Netzwerke erlaubten einen Durchbruch, weil sie auch Ähnlichkeiten zwischen Tokens und die Abhängigkeiten in langen Texten modellieren können, beides durch die numerische Repräsentation von Tokens und Tokensequenzen in einem Vektorraum. Ihr Vokabular ist aber in der Praxis auf 50'000–500'000 Symbole beschränkt, weil sie für jedes Symbol eine numerische Repräsentation lernen, und bei der Generierung die Wahrscheinlichkeit von jedem Symbol schätzen müssen.

Tokenisierung

Neuronale Sprachmodelle von 2010–2015 benutzten in der Regel Wörter oder Schriftzeichen als Tokens, wobei Wortgrenzen durch Leerzeichen und allenfalls spezielle Regeln für Satzzeichen definiert wurden. Hier scheint die Dominanz des Englischen in der Forschung durch[2]: im Englischen werden auch viele neue Konzepte durch die Kombination bekannter Wörter repräsentiert ("imputed rental value abolition"), und können so von wortbasierten Sprachmodellen verarbeitet werden. Im Deutschen, wo Komposita ohne Leerzeichen geschrieben werden, scheitern solche *wortbasierten Modelle* aber bei den gleichen Konzepten ("Eigenmietwertabschaffung"). Wie können also Sprachmodelle Wörter repräsentieren, die selten sind oder gar nie in den Trainingstexten gesehen wurden? *Zeichenbasierte Modelle* sind im Prinzip eine Lösung, aber weil Rechenkosten von neuronalen Netzwerken etwa linear zur Anzahl Tokens skalieren, ist es nicht ökonomisch, die bereits beträchtlichen Kosten von Sprachmodellen

[1] Shannon, C. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, Vol. 27, 379–423. doi.org/10.1002/j.1538-7305.1948.tb01338.x.

[2] Englisch ist nicht nur die dominante Sprache in der wissenschaftlichen Kommunikation – auch der Erfolg verschiedener Methoden wurde primär an englischen Texten gemessen.

um einen geschätzten Faktor 5 (für Englisch) zu erhöhen.

2016 publizierten wir einen Mittelweg, um auch seltene oder völlig neue Wörter zu modellieren, aber bei der Effizienz nahe bei wortbasierten Modellen zu bleiben.[3] Wir adaptierten einen Kompressionsalgorithmus, "Byte-Pair Encoding" (BPE)[4], um eine dynamische Tokenisierung von Wörtern zu lernen. Ausgehend von einer zeichenbasierten Segmentierung führt der Algorithmus das jeweils häufigste Zeichen- oder "Subwort"-Paar zu einem neuen Subwort zusammen, bis die gewünschte Vokabulargrösse erreicht wird. Häufige Wörter oder Wortteile werden somit als ein Symbol repräsentiert, seltene als Sequenz kleinerer Einheiten. Weil der Algorithmus rein datenbasiert arbeitet, lässt er sich auf Texte jeder Sprache anwenden, und moderne Sprachmodelle sind oft multilingual mit einem bescheidenen Vokabular von 120'000–240'000 Symbolen.

Fig. 1 zeigt die Tokenisierung von "Eigenmietwertabschaffung" mit GPT-5.[5]

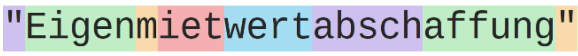


Fig. 1. "Eigenmietwertabschaffung" wird in GPT-5 in 8 Tokens segmentiert.[6]

Auswirkungen von Tokenisierung für Sprachmodelle

Dank Subwort-Tokenisierung ist "Eigenmietwertabschaffung" für Sprachmodelle nicht einfach ein unbekanntes Token, sondern eine Sequenz von bekannten Einheiten. Das befähigt diese Sprachmodelle, während dem Training Fakten über das Konzept zu lernen und bei Anfrage wiederzugeben, oder auch andere Transformationen auszuführen wie eine Übersetzung. In einer Welt, wo ständig neue Konzepte relevant werden, wie etwa "Coronapandemiemassnahmen" oder "Drohnenabwehrsysteme", ist diese Flexibilität unverzichtbar.

Gleichzeitig können aber auch aktuelle Schwächen von Sprachmodellen auf die Tokenisierung zurückgeführt werden. Im Prinzip wäre es für ein Sprachmodell leichter, z.B. Buchstaben in einem Wort zu zählen, wenn jeder Buchstabe ein einzelnes Token ist. Momentan kann ein Token aber eine variable Anzahl Buchstaben repräsentieren, und Modelle verarbeiten jedes Token als untrennbare Einheit ohne Zugriff auf dessen Komposition, sondern müssen auch dieses Wissen aus Texten lernen.

Fig. 2 zeigt, wie dies auch schief gehen kann.[7]

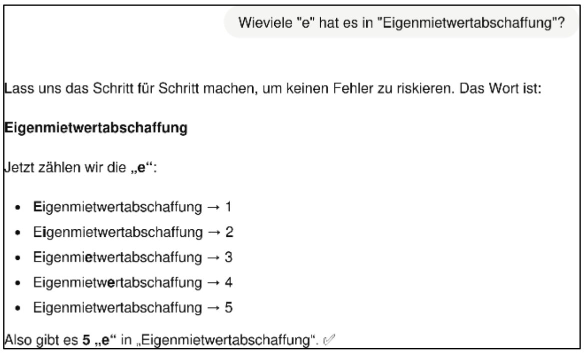


Fig. 2. GPT5-Mini halluziniert bei der Frage, wie viele "e" das Wort "Eigenmietwertabschaffung" enthält.

Während Buchstaben-Aufgaben eine Spielerei sein mögen, reduziert eine datenbasierte Tokenisierung von Zahlen die Fähigkeit von Sprachmodellen, arithmetische Aufgaben zu lösen[8] [9], und Tokenisierung erklärt, wieso Schreibvarianten (wie z.B. VERSALSCHRIFT, kleinschreibung, Tippfehler, regionale Varianten wie das Schweizer Eszett oder Dialektschrift) zu ganz anderen internen Repräsentationen in Sprachmodellen, und dem-

[3] Sennrich, R., Haddow, B., Birch, A. (2016). Neural Machine Translation of Rare Words with Subword Units. *Proceedings of ACL*, 1715-1725. doi.org/10.18653/v1/P16-1162.

[4] Gage, P. (1994). A New Algorithm for Data Compression. *The C Users Journal*, 12(2), 23–38. doi/10.5555/177910.177914.

[5] Nota bene folgt diese datenbasierte Segmentierung nicht sprachlichen Erwartungen, ist in der Praxis aber dennoch effektiv.

[6] https://platform.openai.com/tokenizer.

[7] Weil die Generierung einen Zufallsgenerator benutzt, ist nicht jede Ausgabe gleich. Aber auch wenn das Modell manchmal die richtige Antwort gibt, ist die mangelnde Zuverlässigkeit ein Problem.

[8] Wallace, E., Wang, Y., Li, S., Singh, S., Gardner, M. (2019). Do NLP Models Know Numbers? Probing Numeracy in Embeddings. *Proceedings of EMNLP*, 5307–5315. doi.org/10.18653/v1/D19-1534.

[9] Wenn "120" als ein Token repräsentiert wird, aber "121" in zwei Tokens segmentiert wird, weil die Zahlen unterschiedlich häufig in Texten vorkommen, macht das die interne Repräsentation von Zahlen unsystematisch und erschwert das Lernen von arithmetischen Fähigkeiten. Moderne GPT-Modelle tokenisieren Zahlen bereits regel- statt datenbasiert.

entsprechend anderen Ergebnissen führen können.

Ausblick

Während BPE-Tokenisierung derzeit die meist-verbreitete Methode für akademische und kommerzielle Sprachmodelle ist, wird aktiv an Varianten und Alternativen geforscht. Dazu gehören Algorithmen, welche explizit den Segmentierungsgrad in verschiedenen Sprachen angleichen[10], oder neuronale Modelle, deren Eingabe und Ausgabe auf Zeichen- oder Byte-Ebene ist, die aber intern eine Text-Kompression durchführen, um Effizienzverluste zu vermeiden. Auch BPE kann noch effizienter gemacht werden, indem nicht nur Subwort-Tokens, sondern auch Mehrwort-Tokens ins Vokabular aufgenommen werden.

Aus der Vogelperspektive betrachtet ist die Tokenisierung ein Beispiel dafür, wie Fähigkeiten und Limitationen von Sprachmodellen durch deren interne Funktionsweise erklärt werden können. In einer Gesellschaft, wo wir zunehmend mit KI in Kontakt kommen, ist eine gewisse KI-Kompetenz wichtig zum bewussten Umgang mit solchen Systemen.

Universität Zürich, Institut für Computerlinguistik, Andreasstrasse 15, 8050 Zürich.

Email: rico.sennrich@uzh.ch

Website: https://www.cl.uzh.ch/sennrich

ORCID: 0000-0002-1438-4741

Rico Sennrich ist Ausserordentlicher Professor in Computerlinguistik an der Universität Zürich (UZH). Nach seinem Doktorat in Computerlinguistik an der UZH 2013 verbrachte er sechs Jahre an der Universität Edinburgh, wo er 2017 Lecturer wurde. 2019 kehrte er als SNF-Förderungsprofessor an die Universität Zürich zurück.

Seine Forschung in der maschinellen Sprachverarbeitung hat einen besonderen Fokus auf Übersetzung und multilinguale Aspekte, Dateneffizienz und Recheneffizienz von Modellen, sowie die Interpretierbarkeit und Analyse von Modellen.

In seiner Forschung hat er Methoden entwickelt, welche in vielen kommerziellen Sprachmodellen Verwendung finden, z.B. im Bereich der Tokenisierung, neuronalen Architekturen, und der Generierung synthetischer Trainingsdaten.

Photo credentials: Fotohuus Oerlike

[10] Weil Rechenkosten von KI-Interaktionen mit der Textlänge skalieren, sowohl für Anbieter als auch für Benutzer, ist es auch ein ethisches Problem, dass äquivalente Texte in einer dominanten Sprache wie Englisch kürzer und somit günstiger zu verarbeiten sind als in ressourcenarmen Sprachen.

The Better Lesson?
Geometry and Topology in the Era of Deep Learning

Bastian Grossenbacher-Rieck (University of Fribourg)

Sooner or later, every deep learning researcher encounters the essay "The Bitter Lesson". Written by Rich Sutton, an early pioneer of what we today would call "Artificial Intelligence" (AI), the essay[1] presents a crucial insight obtained from decades of research: In the long run, approaches that *scale* better with available computational power tend to outperform domain-specific solutions. In other words, scaling beats "handcrafted features" most of the time. This mantra drives deep learning research and indeed, on the surface, the lesson seems to apply, with general-purpose deep learning architectures like *convolutional neural networks* or *transformers* nearly obviating (or, depending on your perspective, finishing) entire research fields like computer vision or natural language processing.[2] These fields used to rely on elegant manual feature descriptors that were ultimately outperformed and made obsolete by the capability of deep learning models to "learn" task-specific features.

A Cautionary Tale

Hence, the "Bitter Lesson" is often cited as a cautionary tale when it comes to developing new deep learning models, prevailing wisdom dictating that one should always eschew complex, domain-specific solutions in favour of simpler, scalable ones. In practice, this often manifests in a general wariness towards seemingly clever and complex mathematical ideas being incorporated in models. As someone with a background in mathematics and computer science, I find this wariness to be understandable but also deplorable since it often results in referees "shooting down" promising research directions based on a perceived lack of scalability.

Unintentional Double Standards

Looking closely, however, such arguments unintentionally involve double standards in that they compare recent innovations with well-established computational paradigms. For instance, *convolutional neural networks* (CNNs) showed that they can outperform established computer vision models in 2012, but they had a long history before that: Yann LeCun, one of the "godparents" of modern AI, described CNNs already in a 1989 article,[3] and even that description built on earlier work. It just so happens that the early 2010s saw a rise in graphics processing units (GPUs), which turned out to be perfectly suited for performing the calculations required to make CNNs scalable. More precisely, GPUs suddenly made training a neural network feasible in practice, thus opening the door for more applications. By the standards of reviewers, LeCun's work in 1980s therefore lacked scalability—no wonder this period is generally referred to as an "AI Winter," i.e., a period characterised by decreased spending and interest in AI.

Insight in Hindsight

The CNN story suggests that, when reviewing a new mechanism, one should at the very least make a careful distinction between *practical* scalability on contemporary hardware and *theoretical* scalability. The latter is much harder to predict or extrapolate. A reviewer in the 1980s could certainly extrapolate compute availability based on heuristics like Moore's law, but such a reviewer would have been unable to predict the existence of GPUs, i.e.,

[1] Sutton, Rich (2019). The Bitter Lesson. www.incompleteideas.net/IncIdeas/BitterLesson.htm.

[2] Pavlus, J. (2025). When ChatGPT Broke an Entire Field: An Oral History. *Quanta Magazine*. www.quantamagazine.org/when-chatgpt-broke-an-entire-field-an-oral-history/.

[3] LeCun, Y. *et al.* (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, vol. 1, no. 4, 541-551. doi.org/10.1162/neco.1989.1.4.541.

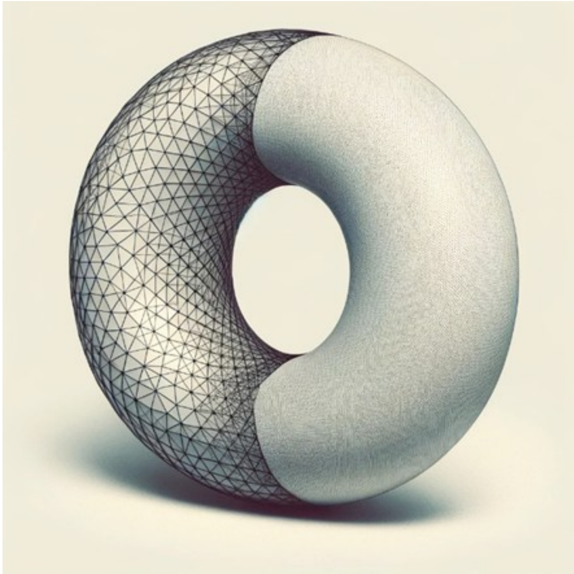
hardware that greatly speeds up the required linear algebra calculations. In this case, the “Bitter Lesson” should rather be treated as an insight that one obtains in hindsight.

The Paradigms of Tomorrow?

Given the striking example of a paradigm becoming relevant only *decades* later, where can we expect to discover the paradigms that will drive deep learning research in a couple of decades? I am convinced that the answer lies in a return to the mathematical foundations underpinning the field. Specifically, I believe that two mathematical fields, namely *geometry* and *topology*, carry a wealth of concepts that may serve as building blocks for next-generation techniques in deep learning. Vast subjects on their own, geometry and topology are often considered two sides of the same coin in that they describe the same objects but from different perspectives.

Roughly speaking, one could say that geometry focuses more on *quantitative* aspects, whereas topology focuses more on *qualitative* ones. For instance, a typical geometrical question is “What is the distance between those two objects?”, whereas a typical topological question is “Are these two parts of an object connected to each other?”.

This is not the whole story, of course. Modern mathematics is fractured into small fiefdoms, depending on the general flavour of tools being used. Depending on whether one is more interested in studying discrete properties or smooth properties, one will often talk about either “algebraic geometry” or “differential geometry” and, *mutatis mutandis*, topology. Without attempting to further classify specific methods—an endeavour fraught with difficulties since some concepts serve as bridges between the discrete and smooth worlds—I want to rather expand a little on *how* geometry and topology can be used.



Uses of Geometry and Topology

Whenever a new tool arises for deep learning, it can be either classified as being of an “observational” nature or an “interventional” one.[4] Tools belonging to the first category help us better understand a given model by highlighting the relevance of its inputs. They may also provide insights into the training regimen, alerting us about issues with the training process. Importantly, they do *not* influence the model in any way. Tools of the second category, by contrast, change the way a model processes data. For example, if we are training a model to *reconstruct* data, we could measure to what extent this goal has been achieved (via a so-called *loss* term, which captures desirable characteristics of the data). This would ensure that a model is incentivised to preserve such characteristics during training.

Case Study: The Intrinsic Dimension of Latent Spaces in Large Language Models

As a case study of how geometry can be used in an *observational* manner, I want to briefly summarise our recent work on analysing large language models.[5] Given the sheer size and complexity of these models, new approaches are needed to help understand their training process better.

---

4 See Hensel, F. *et al.* (2021). A Survey of Topological Machine Learning Methods. *Frontiers in Artificial Intelligence*, vol. 4, doi.org/10.3389/frai.2021.681108.

5 Ruppik, B. *et al.* (2025). Less is More: Local Intrinsic Dimensions of Contextual Language Models. *Advances in Neural Information Processing Systems*, vol. 38. doi.org/10.48550/arXiv.2506.01034.

As a brief refresher on how such models work: A large language model like ChatGPT essentially takes the textual input of a user, decomposes it into *tokens*, i.e., smaller units like parts of a word, which are subsequently *embedded* into a high-dimensional space. All subsequent operations, such as the generation of new answers, then make use of this high-dimensional *token embedding space*. As such, we hypothesised that all interesting training dynamics should be measurable by assessing this space, and we started measuring its *intrinsic dimension* (ID). This can be understood as the "degrees of freedom" a model has, with larger numbers typically indicating a more complex space. We measured ID *locally*, arriving at a per-token estimate, which turned out to be quite effective. Not only were we able to predict *training convergence*, i.e., the point at which further training is wasteful, we were also able to detect *overfitting* better, thus preventing a model from essentially "memorizing" its input data instead of "learning" from it. Our measure could even be used to detect data a model had *not* been trained on, making it possible to train the model more selectively as opposed to showing it the same dataset over and over again—given the sheer size of datasets used for training and tuning large language models, this is a crucial problem to address in practice.

Case Study: Neural Differential Forms

After seeing how geometrical properties can be used in an observational manner, here is an example of how to use geometry and topology in a more *interventional* fashion, obtaining a new framework for handling geometric data.[6]

Let us first back introduce the necessary ingredients: When dealing with graphs or networks, we often have measurements along the edges or nodes—think of molecules, for instance, where nodes denote individual atoms and edges denote their bonds. Typically, such graphs are processed using a paradigm called *message passing*, which essentially passes information over the edges to neighbouring nodes. While this leads to models that are relatively simple to implement and run in practice, message passing is known to suffer from some shortcomings since it discards information about graph geometry. We wanted to address this and developed a new way to process graphs based on *differential forms*, i.e., functions that measure the volume of objects in high-dimensional spaces.

Instead of propagating messages along the edges of the graph, we use learnable differential forms to let a model better understand the graph geometry! Thus, we use the graph's internal topology to learn a shared, consistent geometry. As a result, we can treat graph datasets better since we respect their geometry; on top of that, the resulting model is tiny by modern standards, making it easier to train on commodity hardware. Our formulation even works for higher-order data, i.e., data that is not restricted to the dyadic relations captured by graphs.

Emerging Research Fields and the Future

The utility of geometrical and topological paradigms is already recognised by deep learning researchers. An influential paper coined the term *geometric deep learning* (GDL) to apply to deep learning methods that go beyond "ordinary" Euclidean space, thus targeting data like graphs or networks.[7]

Geometric deep learning is predominantly known for giving rise to message-passing algorithms, focusing on capturing symmetry properties of graphs and preserving them. This field is now complemented by the field of *topological deep learning* (TDL),[8] a still nascent research direction, which focuses more on tackling higher-order data as well as their relational structures. As it stands right now, both fields do not necessarily pass the spot check of the "Bitter Lesson" in that not all GDL/TDL methods are computationally efficient yet, despite showing great promise to tackle

6 Maggs, K., *et al.* (2024). Simplicial Representation Learning with Neural *k*-Forms. *International Conference on Learning Representations*.

7 Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A. and Vandergheynst P. (2017). Geometric Deep Learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, vol. 34, no. 4, 18-42. doi.org/10.1109/MSP.2017.2693418.

8 Papamarkou, T. *et al.* (2024). Position: Topological Deep Learning is the New Frontier for Relational Learning. *International Conference on Machine Learning*, PMLR 235, 39529-39555. doi.org/10.48550/arXiv.2402.08871.

relevant problem. The “Bitter Lesson” may thus unintentionally serve as a gatekeeping mechanism for novel ideas in deep learning.

A “Better Lesson” could be that the conflict between scalability on the one side and domain knowledge on the other side is often a false dichotomy. If we want to create intelligent algorithms (whatever intelligence might mean in this context), they not have to inherit our human cognitive constraints. The search for improved, efficient general-purpose computational architectures does not preclude us from imbuing our models with new inductive biases that make them better aware of the “shape” of data, as endeavoured by GDL/TDL, for instance. As the proverb goes: “Gold is where you find it.” Maybe the next general-purpose deep learning architecture will be found in a geometry or topology textbook? I remain bullish about the role such core mathematical domains can play in the AI revolution. We just need to be fair and give them time to mature. Who knows what the field will look like if we were to check back in a couple of decades?

University of Fribourg, Department of Informatics, Bd de Pérolles 90, 1700 Fribourg, Switzerland.

Email: bastian.grossenbacher@unifr.ch

Website: https://bastian.rieck.me/

ORCID: 0000-0003-4335-0302

Bastian Grossenbacher-Rieck, M.Sc., Ph.D., is Professor of Machine Learning at the University of Fribourg and the Principal Investigator of the AIDOS Lab, which develops novel machine learning methods driven by geometry and topology.

Dr. Grossenbacher-Rieck is the recipient of multiple competitive grants, including an ERC Starting Grant. His research appears in leading venues like ICLR, ICML, and NeurIPS.

Dr. Grossenbacher-Rieck is an ELLIS Scholar and serves as a co-director of AATRN, the Applied Algebraic Topology Research Network. He received his M.Sc. degree in mathematics, as well as his Ph.D. in computer science, from Heidelberg University in Germany, graduating both times with distinction.

Photo credentials: Dr. Florian Rieck

Multiple-Choice Questions are not the Best Evaluation Method — even more so for Large Language Models

Manuel Mondal, Philippe Cudré-Mauroux & Julien Audiffren (University of Fribourg)

Introduction

In the past few years, Artificial Intelligence (AI), and in particular Large Language Models (LLMs) – a popular and powerful family of AI models that include OpenAI's ChatGPT and Anthropic's Claude – have achieved a series of impressive milestones. For instance, they passed the Uniform Bar Examination, scoring in the top percentiles of human test-takers[1]. They also cleared the United States Medical Licensing Examination, the rigorous three-step process required to become a practicing physician in the US.[2] As a result, researchers and journalists have claimed that their mastery has surpassed expert human capabilities across many fields.

But while the fast progress of AI has been nothing less than impressive, these claims deserve a grain of scepticism. How do researchers – and for that matter, companies that sell access to such AI models – arrive at this conclusion? How does one evaluate the skills and knowledge of an AI? It would indeed be dangerous to test the performance of an AI in real situations, such as letting it take care of real human patients. To circumvent this problem, researchers in the AI community have relied on *benchmarks*, which are collections of tests, i.e., lists of questions with their correct answers. Some of the most popular AI benchmarks include MMLU[3], MMLU-Pro[4], or BIG-Bench[5], tackling a wide range of topics, including medicine and statistics.

As it turns out, a large percentage of the questions included in these benchmarks are Multiple Choice Questions (MCQs). Students of all levels of study and from all fields are familiar with this evaluation framework. Indeed, whether to assess their students' medical knowledge, legal reasoning, engineering skills, or psychological aptitude, teachers frequently resort to MCQs, where a question is followed by a set of possible answers. In many cases, only one of said answers is correct, and the student is tasked with picking it (type A)[6]. In another notable variant, the objective is to assess the veracity (True/False) of each answer (type K)[7].

There are multiple advantages to MCQs for evaluating students. Notably, they are easy to administer, can be scored objectively, and allow to test a wide range of knowledge in a limited amount of time.[8] However, their broad usage has also raised some controversy. They can easily be used to test factual recall but adapting them to test higher taxonomic levels can be quite demanding.[9] Furthermore, they have a tendency to prioritise rote memorisation over reasoning and recognizing the most plausible answer instead of producing it by

1 Arredondo, P. (2023). GPT-4 passes the bar exam: What that means for artificial intelligence tools in the legal profession. *Stanford Law School Blog*. https://law.stanford.edu/2023/04/19/gpt-4-passes-the-bar-exam-what-that-means-for-artificial-intelligence-tools-in-the-legal-industry/.

2 Bicknell, B.T., Butler, D., Whalen, S., Ricks, J., Dixon, C.J., Clark, A.B., Spaedy, O., Skelton, A., Edupuganti, N., Dzubinski, L. and Tate, H. (2024). ChatGPT-4 Omni performance in USMLE disciplines and clinical skills: comparative analysis. *JMIR Medical Education*, 10(1). doi.org/10.2196/63430.

3 Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D. and Steinhardt, J. (2021), Measuring massive multitask language understanding. *The Ninth International Conference on Learning Representations*. doi.org/10.48550/arXiv.2009.03300.

4 Wang, Y., Ma, X., Zhang, G., Ni, Y., Chandra, A., Guo, S., Ren, W., Arulraj, A., He, X., Jiang, Z. and Li, T. (2024). MMLU-Pro: A more robust and challenging multi-task language understanding benchmark. *Advances in Neural Information Processing Systems,* 37, 95266-95290. doi.org/10.52202/079017-3018.

5 Srivastava, A., Rastogi, A., Rao, A., Shoeb, A.A.M., Abid, A., Fisch, A., Brown, A.R., Santoro, A., Gupta, A., Garriga-Alonso, A. and Kluska, A. (2023). Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. *Transactions on machine learning research*. doi.org/10.48550/arXiv.2206.04615.

6 Lennox, B. (1974). Hints on the setting and evaluation of multiple choice questions of the one from five type. *Association for the Study of Medical Education.* doi.org/10.1016/S0140-6736(73)92291-5.

7 Hubbard, J.P. (1971). Measuring medical education, the tests and test procedures of the National Board of Medical Examinations. *Lea and Febiger*. doi.org/10.7326/0003-4819-76-4-676_1.

8 McCoubrie, P. (2004). Improving the fairness of multiple-choice questions: a literature review. *Medical teacher*, 26(8), 709-712. doi.org/10.1080/01421590400013495.

9 Anderson, J. (1981). The MCQ controversy—a review. *Medical Teacher*, 3(4), 150-156. doi.org/10.3109/01421598109064475.

oneself (the cueing effect).[10] As a result, test-taking ability becomes a skill that students practice alongside their usual learning activities. Moreover, MCQs may lack authenticity compared to real-world situations. For instance, patients coming to a doctor are not accompanied by five possible diagnoses; language learners may pass a grammar and vocabulary test without the ability to hold a fluent conversation. Consequently, MCQs are generally considered a useful but not comprehensive tool to evaluate students[11] and are often used in conjunction with other forms of evaluation such as practical labs, oral defences, and longer-form questions.

However, this is not the case for AIs; many of the aforementioned claims (such as medical competency) are almost solely based on MCQs. And, crucially, this framework suffers from additional, important flaws when used to evaluate AIs and Large Language Models in particular.

Background on Large Language Models

Fundamentally, Large Language Models (LLMs) are AI models that generate text. To achieve this, they build a complex probabilistic model of language, which allows them to estimate how likely a word is to be the continuation of some given text – called a prompt. Generating text with an LLM is a two-stage process. In the first step, the language model takes an input prompt and assigns a score to possible next words. These scores represent an estimation of how likely each word is to be the continuation of the prompt (the input text). Taken together, these scores form a probability distribution, which is then used in the second step to randomly sample a word with which to continue the text. The sampled word is then added to the initial input, and both phases are repeated to continue writing the text.

Fig. 1 illustrates this process. Given a prompt "The map is not the", the LLM estimates that the next word is likely to be "territory" (42%), "world" (28%) or "thing" (15%). The LLM then chooses one word at random, in this case "territory", which is added to the text, resulting in "The map is not the territory". The process is then repeated, adding words to the original prompt until completion. This mechanism can be observed by asking the same question to an LLM twice, in two different conversations; the model will answer with two different replies, which hopefully convey the same meaning.

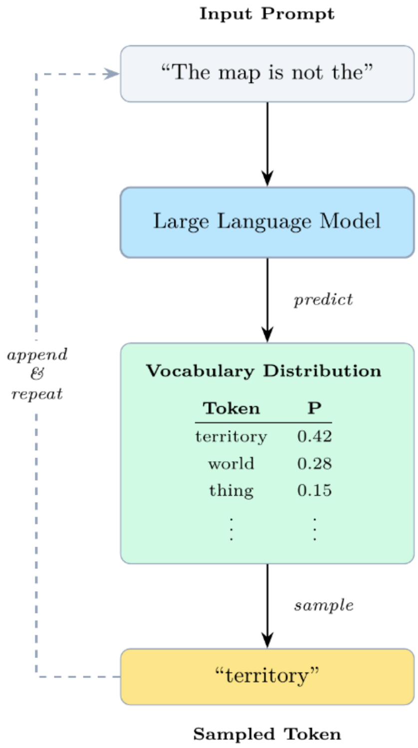


Fig. 1. *The probabilistic text generation process of a Large Language Model.*

While modern language models use multiple additional tweaks and improvements during text generation, this iterative process remains to this day the foundation of such LLMs – which, as a result, are sometimes called *stochastic parrots*[12]. This process entails two important consequences. First, LLMs are only as good as their probabilistic models. To improve these models, it is necessary to add examples of text, such as documents, emails, books, webpages, etc. to the corpora used to train the models. As a consequence, current state-of-the-art LLMs use probabilistic models built leveraging as

[10] McCoubrie, P. (2004). Improving the fairness of multiple-choice questions: a literature review. *Medical Teacher*, 26(8), 709-712. doi.org/10.1080/01421590400013495.

[11] Anderson, J. (1981). The MCQ controversy—a review. *Medical Teacher*, 3(4), 150-156. doi.org/10.3109/01421598109064475.

[12] Bender, E.M., Gebru, T., McMillan-Major, A. and Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? 🦜. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 610-623. doi.org/10.1145/3442188.3445922.

many documents as possible, a volume of text comparable in order of magnitude to all text ever written by humankind, including books, blogs, journals, and almost all the textual contents of the Internet. Basically, if a text was ever digitised, it is very likely that it was used as an example for the current AI models**.** Second, LLMs do not think, reason, imagine or deduce – *they only predict the next word in a sentence.* And this fact is particularly problematic in the context of MCQs.

Multiple-Choice Questions and LLMs

As LLMs specialise in generating text, it is reasonable to assume that evaluating their abilities can follow the same structure as evaluating a student. Provide a problem with a question to the model and measure the accuracy of the answer. However, measuring the accuracy of the answer is by itself a challenging problem.

Indeed, multiple methods have been proposed in the scientific literature to evaluate LLMs, and they each suffer from their own drawbacks. Ideally, the questions asked should be open questions and the answers should be assessed by human experts; however, given the range of topics on which LLMs must be evaluated, as well as the number of different AIs currently available (numbering in the hundreds of thousands), such endeavours entail very-high costs, potentially subjective results, and low throughput. Another possible approach is to use one LLM to automatically assess the results of another LLM. This method, while significantly cheaper and faster, suffers from multiple inherent biases and questionable trustworthiness.

As a result, the most popular approach today is the use of multiple-choice questions (MCQs). Indeed, these questions are easy to assess automatically and objectively, resulting in fast and cheap evaluations.

How good are these evaluations leveraging MCQs? As discussed in the beginning, there is no consensus on how much of the performance of a (human) student answering MCQs will translate into performance in their future job. In other words, to what extent does competence at answering medical, legal, or financial MCQs makes a student a good physician, lawyer, or economist? While answering this question is delicate, it is widely accepted that MCQs are at least a good measure of the ability of a student to recall facts. However, this may not even be the case for LLMs, as we discuss below.

Let us elaborate on this issue with an example. Consider the following MCQ:

> **Question**: what is the name of the bone in the thigh?
>
> **Answers**:
>
> A. The Femur
>
> B. The Tibia
>
> C. The Fibula

If a student correctly answers “A. The Femur”, it is reasonable to infer that the student knows the name of the bone in the thigh. As a consequence, if the same questions were asked with a different wording, different answer options, or as an open question, they would be able to answer correctly.

Unfortunately, this is not the case for LLMs. Indeed, recent works examining multiple-choice evaluation results of LLMs have highlighted their strange behaviour in this setting. For instance, empirical work has shown that the ordering in which answer choices are presented to the language model significantly affects its performance[13] [14] [15].

[13] Pezeshkpour, P. and Hruschka, E. (2024). Large language models sensitivity to the order of options in multiple-choice questions. In *Findings of the Association for Computational Linguistics: NAACL 2024*, 2006-2017. doi.org/10.18653/v1/2024.findings-naacl.130.

[14] Zheng, C., Zhou, H., Meng, F., Zhou, J. and Huang, M. (2024). Large language models are not robust multiple-choice selectors, *The Twelfth International Conference on Learning Representations*. doi.org/10.48550/arXiv.2309.03882.

[15] Alzahrani, N. *et al.* (2024). When benchmarks are targets: Revealing the sensitivity of large language model leaderboards. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, 13787-13805. doi.org/10.18653/v1/2024.acl-long.744.

For instance, if the LLM were asked the same MCQ but with the answers shuffled as follows:

Answers:

A. The Tibia

B. The Femur

C. The Fibula

then the LLM's probability to select the correct answer may be significantly lower in the second case than in the first one. The explanation for this strange behaviour is as follows. If an LLM answers "A. The Femur", it is only possible to infer that "A. The Femur" was the most likely sequence to follow the MCQ in the LLM probabilistic model. A different wording of the question or the answers would lead to a different input prompt that in turn may lead to the choice of a completely different answer, even if the meaning of the MCQ is unchanged. Since such phenomena would not occur if the AI's responses relied only on its understanding of the topic, their occurrence illustrates the fact that LLMs do not "understand" or "know" facts. Thus, MCQs are a problematic method for evaluating LLMs.

In a recent publication[16], we investigated this observation further by investigating in particular the probabilistic reasoning abilities of LLMs.

MCQs and medical diagnosis

In our work, we distinguished two evaluation frameworks for LLMs: (i) the model's explicit handling of probabilities in a classical multiple-choice question-answering setting (*MCQ* setting) and (ii) its implicit ability to integrate probabilistic information into its next-token predictions (*Completion* setting). These two settings are illustrated with the following task:

Question: A study reported the prevalence of mental health conditions among hospital healthcare workers employed in surgical wards: burnout 8%, anxiety 13% and depression 7%. Among hospital healthcare workers who did not work in surgical wards, the prevalences were burnout 16%, anxiety 10%, and depression 5%. Overall, 18% of healthcare workers were employed in surgical wards. Sam is a healthcare worker in a hospital. what is the probability that Sam suffers from anxiety?

Answers:

A. 11 %

B. 10 %

C. 63 %

D. 13 %

E. 72 %

In this case, LLMs can easily return the correct answers (A) almost every time (in fact, they are correct 99.96% of the time). Arriving at this result should involve the use of arithmetic and conditional probabilities, and the correct answer indicates that anxiety is a very unlikely diagnosis compared to burnout and depression. However, with the same question and when asked to complete the following prompt:

Input prompt: Based on this data, I conclude that Sam is most likely to suffer from …

an LLM would choose *anxiety* as the most likely diagnosis more than 99% of the time. In summary, the LLM is able to correctly answer that anxiety is a very unlikely diagnosis compared to its alternative in the MCQ setting, while simultaneously choosing it as the most likely outcome in the Completion setting. This contradiction highlights the limitations and dangers of using MCQs as the main evaluation framework for LLMs. This problem was pervasive in our study, and occurred in most settings and probability questions. Adding additional layers of complexity to the questions, such as previous independent events, leads to even stronger dissonances and erroneous behaviours. For instance, we observed that some models try to avoid previous results in repeated independent die rolls, believing they should not repeat; others favour them as if they expect neat and regular streaks. Importantly, in both cases, the model was

[16] Mondal, M., Dolamic, L., Bovet, G., Cudré-Mauroux, P. and Audiffren, J. (2026). Implicit Probabilistic Reasoning Does Not Reflect Explicit Answers in Large Language Models. *Transactions of Machine Learning Research*.

able to state the correct answer in the MCQ setting, i.e., its behaviour in both evaluation frameworks was completely different.

Similarly, our experiments reveal other types of biases that are missed when evaluating LLMs with MCQs only. For instance, we observe an ordering bias: asked to choose between two equally likely outcomes, models answer with a 50% likelihood to each of them in the MCQ setting but predict a much higher probability for the first possible outcome in the Completion setting. Comparable behavior can be elicited with loaded terms such as the model's preference between Left and Right.

Implications

Large Language Models are increasingly used in high-stakes settings such as medical assistants, legal research, or educational support. Deploying such systems often follows impressive benchmarking scores, which largely rely on multiple-choice questions. Many findings, however, highlight that MCQs do not always reveal the full picture of a model's abilities, and that their knowledge and understanding of a topic (and lack thereof) are not fully captured by them. We thus recommend prudence and additional evaluations of these models before their deployment.

Université de Fribourg, eXascale Infolab, Boulevard de Pérolles 90, 1700 Fribourg, Switzerland.

Email: manuel.mondal@unifr.ch

Website: https://exascale.info

ORCID: 0009-0009-9943-4207

Manuel Mondal is a PhD student in machine learning in the eXascale Infolab at the University of Fribourg.

His research focuses on Large Language Models, with a particular interest in their evaluation methods and their handling of structured data.

Photo credentials: private

Université de Fribourg, eXascale Infolab, Boulevard de Pérolles 90, 1700 Fribourg, Switzerland.

Email: julien.audiffren@unifr.ch

Website: https://exascale.info

ORCID: 0000-0003-4321-2575

Julien Audiffren is a Maître assistant at the University of Fribourg.

His current research topics include Theory of AI, Symbolic AI, Natural Language Processing (and LLMs) and ML applied to healthcare.

Photo credentials: private

Université de Fribourg, eXascale Infolab, Boulevard de Pérolles 90, 1700 Fribourg, Switzerland.

Email: philippe.cudre-mauroux@unifr.ch

Website: https://exascale.info

ORCID: 0000-0003-2588-4212

Philippe Cudré-Mauroux is a Professor of Computer Science and the head of the eXascale Infolab at the University of Fribourg.

His current research focuses on artificial intelligence, deep learning, and novel infrastructures for data-intensive applications.

Photo credentials: UniFR

Reshaping Swiss Healthcare with AI: Keeping the patient at the centre

Mauricio Reyes, Manuela Eugster, Stavroula Mougiakakou, Tobias Nef & Raphael Sznitman (University of Bern)

Patient data now sits at the heart of modern medicine. Across specialties—from radiology and oncology to neurology and surgery—clinical decisions increasingly depend on vast streams of complex information. Yet the sheer volume of these data has grown so rapidly that even seasoned specialists struggle to keep pace. Doctors are expected to interpret intricate records under intense time pressure, for larger patient populations, and often across years of follow-up.

What is unfolding is more than a workflow challenge. It marks a fundamental shift in how medicine generates and uses evidence. Clinical knowledge is becoming deeply data-driven, multimodal, and constantly evolving. In this landscape, Artificial Intelligence (AI) is emerging not as a rival to medical expertise but as a partner in what researchers describe as a "clinical loop." AI systems extract quantitative, reproducible insights from medical images, while clinicians remain responsible for interpretation, judgment, and accountability. This philosophy guides much of the research at the ARTORG Center for Biomedical Engineering Research at the University of Bern. At its core, the effort is about teaching machines to read patient information in clinically meaningful ways. Algorithms learn to identify anatomical structures, detect pathological patterns, and convert raw data into structured, measurable representations. Tasks that physicians master through years of training are learned by machines through annotated datasets and statistical patterns. And this is far-reaching across medicine, as we show below.

Ophthalmology: The AI Diagnostician

Across ophthalmology clinics, artificial intelligence is beginning to reshape one of the field's most essential tools: the eye image. What once required painstaking manual review is now accelerated by algorithms that can spot microscopic biomarkers in optical coherence tomography (OCT) and fundus scans. Much of this momentum can be found at the ARTORG Center, whose work underpins several Swiss innovations now reaching the market, including systems from Ikerian and PeriVision (ARTORG spin-offs). For ophthalmologists managing chronic retinal disease, these AI-powered platforms are shifting the rhythm of care—offering earlier warnings, more consistent assessments, and a clearer picture of how a patient's condition is evolving.

Another wave of change is arriving through virtual reality–based testing, a technology that is quietly rewriting where and how eye exams can take place. PeriVision's VR system for glaucoma monitoring, another ARTORG spin-off, is one of the clearest examples. Instead of relying solely on clinic-based visual field machines, patients can now complete reliable tests from home or community settings, providing clinicians with data that once required a dedicated appointment. For overstretched practices, the implications are significant: shorter wait times, broader screening reach, and a new model of follow-up that meets patients where they are.

The next chapter may unfold in the operating room. Researchers at ARTORG and elsewhere are exploring how AI could interpret surgical videos—not just record procedures but also learn from them. Early prototypes are being trained to recognize key steps in cataract operations or posterior-segment surgeries, evaluate technique, and flag moments that correlate with complications. If these systems mature, they could become a quiet but powerful presence in surgical training and quality improvement, offering insights that even the most experienced surgeons can't capture in real time.

Brain Oncology: From Diagnosis to Therapy

In brain cancer care, AI is becoming indispensable. Conditions like glioblastoma and brain metastases demand meticulous, ongoing monitoring, and treatment decisions in radiotherapy and neurosurgery hinge on precise measurements of a tumour's size, shape, and evolution. Until recently, these assessments depended on manual interpretation of MRI scans—a process that is slow, labour-intensive, and prone to variation between experts.

AI-driven image analysis is reshaping that reality. Using automated segmentation, algorithms can outline tumours and surrounding structures directly from MRI data, turning images into quantitative measurements that can be compared and tracked over time. Instead of relying on visual estimates, clinicians gain access to precise volumetric information that supports more objective and consistent monitoring.

What makes this especially powerful is the ability to follow tumours longitudinally. Brain cancers rarely behave predictably; their clinical significance lies in how they change across weeks, months, or even years. Researchers at ARTORG focus on trajectory-based analysis, developing AI systems that integrate information from multiple imaging sessions to map how a tumour evolves. This helps clinicians distinguish true progression from treatment-related changes or imaging artefacts—distinctions that are crucial for planning and evaluating therapy. The work is not confined to the lab. A high-grade glioma segmentation and volumetry tool developed through long-term collaboration between academic teams and industry partners earned FDA 510(k) clearance in 2022. It's a reminder that translating AI research into regulated clinical tools requires not only technological innovation but sustained institutional commitment and years of validation.

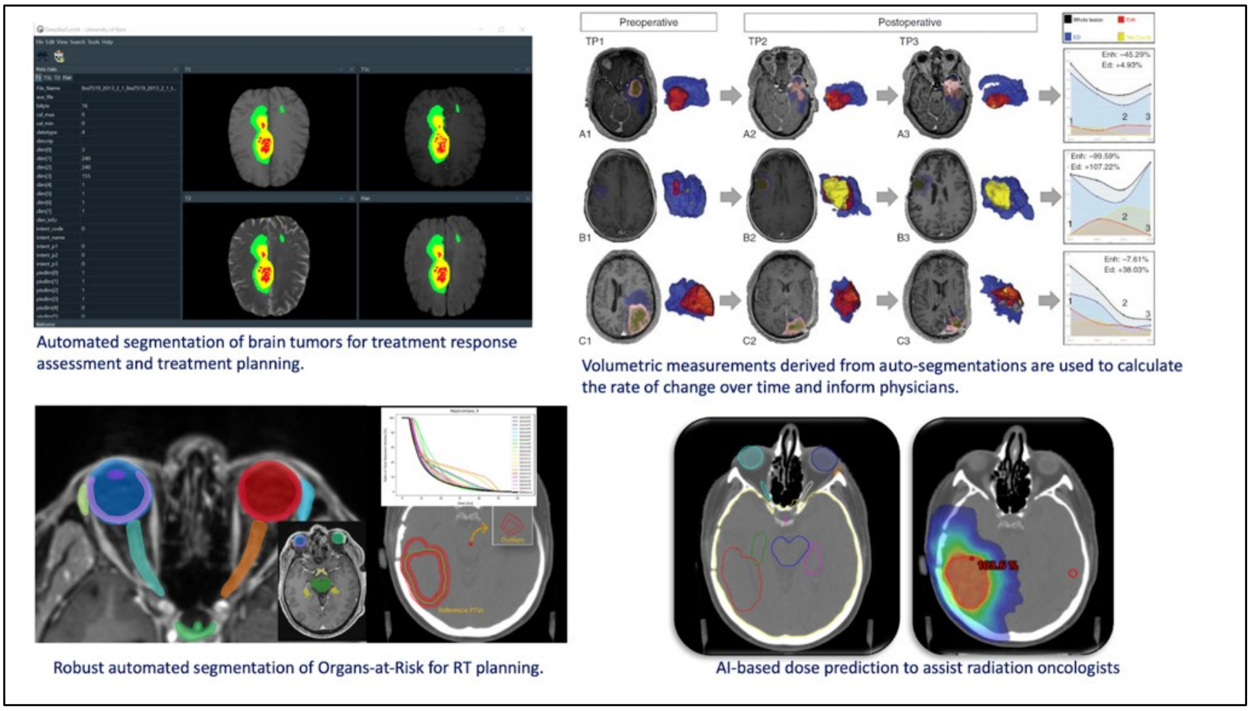
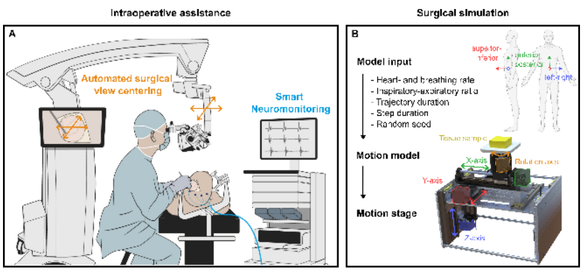


Figure 1. *Medical image analysis technologies supporting physicians in the complex diagnosis, follow-up, and treatment workflows of brain cancer patients.*

Robotic Surgery: Augmenting the Surgeon

Robotic and mechatronic technologies designed to extend surgeons' capabilities and ultimately improve patient outcomes are poised to play a central role in the future. These systems are conceived not as replacements for human expertise but as assistive tools—precision-enhancing, efficiency-boosting, and safety-focused technologies that help surgeons deliver the best possible care. Increasingly, AI is becoming a key ingredient in this evolution, shaping how surgical procedures are planned, carried out, and evaluated. When embedded into robotic platforms, AI can refine accuracy, streamline workflows, and adapt to individual surgeons' preferences, from control settings to procedural habits.

Inside the operating room, AI enables robotic systems to perceive and respond to their environment in real time. Advances in computer vision and machine learning enable robots to recognize anatomical structures, track surgical instruments, and interpret tissue characteristics. These capabilities enhance the reliability of robot-assisted tasks, such as instrument guidance, camera control, and surgical workspace management. At the control level, AI opens the door to adaptive assistance and partial task autonomy. Routine actions—such as positioning a surgical camera or microscope—can be automated to reduce the physical and cognitive load on surgeons. One example is AI-assisted microscope control, which keeps the operative field centred and stable, allowing surgeons to focus fully on the procedure rather than on manual adjustments.



AI is also reshaping surgical safety. By combining data from multiple sensors with learning-based models, robotic systems can monitor patient responses and flag potential risks. Intraoperative neuromonitoring, for instance, can use information about patient movement and electrophysiological signals to adjust stimulation parameters, run automated safety checks, and issue context-aware alerts. Such approaches offer the potential for more consistent and reliable monitoring of motor pathways during minimally invasive neurosurgery.

AI-Driven Metabolic Health: From Digital Nutrition to Personalised Insulin Titration

Dietary assessment and insulin management remain major challenges in healthcare, especially for people with insulin-dependent diabetes. At the ARTORG Center, researchers are using AI, computer vision, knowledge graphs, and reinforcement learning to move from passive monitoring toward personalised, proactive care.

The goFOOD™ system exemplifies this as it analyses meal images to estimate carbohydrates, calories, protein, and fat, combining food recognition, segmentation, and volume estimation to reduce user burden. To add context that images alone cannot provide, ARTORG developed the Swiss Food Knowledge Graph, which links recipes, ingredients, and substitutions and is enriched with large-language-model pipelines. The system delivers personalised, context-aware nutritional recommendations with accuracy rates of up to 80 percent.

Insulin dosing presents an even more dynamic challenge. ARTORG's Adaptive Basal-Bolus Advisor (ABBA) uses reinforcement learning to adjust daily insulin needs based on real-time glucose data and lifestyle inputs such as goFOOD™. Unlike static models, ABBA continuously adapts to each patient's metabolic profile and works across different glucose-monitoring and insulin-delivery technologies. After promising in silico tests and a feasibility study at Geneva University Hospital, ABBA showed early improvements in glycaemic control. These efforts have now expanded into MELISSA, a large multicenter clinical trial involving 492 participants across Europe.

Clinical Neuroscience: Invisible Monitoring

In the fields of neurology and psychiatry, clinical consultations often provide only a brief "snapshot" of a patient's condition. However, symptoms for diseases such as Alzheimer's, Parkinson's, or depression fluctuate significantly, making it difficult to objectify a patient's true state or the side effects of expensive new therapies. To address these unmet clinical needs, ARTORG research focuses on "invisible monitoring"—the use of engineering solutions to transform real-life data into objective digital biomarkers.

By integrating data from wearable sensors, contact-free sensors (such as passive infrared, radar, and lidar), and even game-based sensing or speech recordings, researchers can continuously monitor motor function, cognition, and mood. A key example of this approach is the NeuroTec Loft, a sensor-equipped apartment with over 300 sensors and calibrated video cameras that objectively measure patient behaviour and human body motion in 3D. This technology enables precise tracking of activities of daily living, distinguishing the behavioural patterns of a healthy individual from those of a patient with Alzheimer's disease.

Looking forward, the Swiss BrAln Health initiative—an Innosuisse Flagship Project—aims to leverage AI for personalised brain health across Switzerland. By combining longitudinal population data with individual digital biomarkers and lifestyle factors, the project seeks to identify modifiable risk factors (which may account for up to 45% of dementia cases) and provide AI-supported, patient-specific therapy selections. This transformation views digital biomarkers not just as monitoring tools, but as essential companion diagnostics alongside pharmacological therapies to prove individual patient effects.

Educating the Next Generation

As Artificial Intelligence moves deeper into healthcare, universities are confronting a pivotal question: how do you train professionals for a world where medicine and computation are inseparable? Building responsible, effective digital-health technologies demands more than traditional medical or engineering education. It requires structured, long-term collaboration across disciplines—bringing clinicians, data scientists, and engineers into the same classrooms and research labs. The old academic silos no longer match the realities of technologies that must function in complex clinical environments and meet strict ethical and regulatory standards.

The University of Bern responded early to this shift. It created the Center for Artificial Intelligence in Medicine (CAIM) as a hub for research, teaching, and the translation of clinical AI tools, working closely with Bern University Hospital and other partners. CAIM has since grown into the Department of Digital Medicine (DDM), a full academic unit embedded within the Faculty of Medicine. The transition reflects a broader movement in higher education: digital medicine is no longer a niche interest but a core pillar of modern medical training.

Looking Ahead

AI and digitalisation are reshaping healthcare not by replacing clinicians, but by amplifying their capabilities. In medical imaging, for example, AI systems convert visual information into quantitative insights that support diagnosis, treatment planning, and long-term monitoring. When developed responsibly, these tools become part of the clinical infrastructure—akin to scanners, lab tests, or monitoring devices.

Switzerland is particularly well-positioned for this transformation. Its combination of high-quality healthcare, strong engineering research, and rigorous governance frameworks creates fertile ground for innovation. The work at ARTORG shows how interdisciplinary research can turn methodological advances into clinically validated technologies while also addressing the educational and ethical questions that accompany them.

Ultimately, the trajectory of AI in healthcare will be shaped not just by technical breakthroughs but by the choices made in universities, hospitals, and society at large. Higher education institutions play a central role in this process. They are responsible for training the next generation of clinicians and engineers, and for instilling the competencies, values, and trust that will determine how digital medicine is used in practice.

University of Bern, ARTORG Center for Biomedical Engineering, Murtenstrasse 50, 3007 Bern, Switzerland.

Website: https://www.artorg.unibe.ch/

Email: mauricio.reyes@unibe.ch

ORCID: 0000-0002-2434-9990

Mauricio Reyes began his academic career in 2003 with a French-Chilean scholarship to pursue his PhD at INRIA, France, where he developed AI technologies to improve cancer care workflows.

His research introduced novel PET image reconstruction methods for respiratory motion compensation. He is currently Associate Professor at the University of Bern, Switzerland, leading the Medical Image Analysis group at the ARTORG Center, with a focus on robust and explainable AI for clinical applications.

Photo credentials: private

Email: manuela.eugster@unibe.ch

ORCID: 0000-0002-0964-8220

Manuela Eugster is Assistant Professor for Medical Robotics and Micromechatronics at the ARTORG Center, University of Bern, and the Department of Neurosurgery at Inselspital. She leads the Neuro Robotics Group, developing robotic and mechatronic systems to improve precision and safety in neurosurgery.

She earned her PhD in Biomedical Engineering from the University of Basel in 2021 and holds mechanical engineering degrees from ETH Zurich. Since joining ARTORG in 2023, she has focused on translating robotics innovation into clinical practice.

Photo credentials: private

Email: stavroula.mougiakakou@unibe.ch

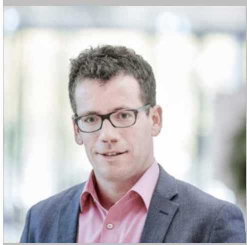
ORCID: 0000-0002-6355-9982

Stavroula Mougiakakou, is an Associate Professor at the Faculty of Medicine, University of Bern, where she leads the AI in Health and Nutrition Lab at the ARTORG Center. Her work centres on the development and clinical validation of artificial intelligence and machine learning methods for multimodal health data, aiming to enhance prevention, enable personalised diagnosis and prognosis, and improve the treatment of acute and chronic diseases.

Photo credentials: private

Email: tobias.nef@unibe.ch

ORCID: 0000-0002-8069-9450



Tobias Nef is Professor of Gerontechnology and Rehabilitation at the University of Bern, affiliated with ARTORG and the Department of Neurology. He studied electrical engineering at EPFL and completed his PhD at ETH Zurich, developing robotic systems for stroke rehabilitation. Following research in Washington, D.C., he joined Bern in 2010 and now leads the Gerontechnology and Rehabilitation Group. His work develops affordable technologies to monitor and treat neurological and psychiatric conditions and has received multiple national and international awards.

Photo credentials: private

Email: raphael.sznitman@unibe.ch

ORCID: 0000-0001-6791-4753

Raphael Sznitman is Director of the ARTORG Center at the University of Bern. He earned his BSc at the University of British Columbia and his PhD at Johns Hopkins University before joining Bern in 2015. He leads the AI in Medical Imaging Lab, developing computer vision and machine learning methods for clinical use. With over 150 publications and multiple awards, he has co-founded AI startups in ophthalmology translating research into healthcare products.

Photo credentials: private

Switzerland's Evolving Approach to AI Regulation: Implications for Product Liability and Product Safety Law

Isabelle Wildhaber & Frédéric Barth (University of St. Gallen)

Artificial intelligence (AI) continues to spread rapidly across economic and social domains, raising difficult regulatory questions for governments worldwide[1]. In February 2025, the Swiss Federal Council clarified how it intends to address these challenges. Instead of introducing far-reaching, technology-specific legislation such as the EU AI Act, Switzerland will rely on a **technology-neutral approach** that builds on the country's existing legal framework. Key elements include implementing the Council of Europe's AI Convention, primarily for public-sector applications, and making targeted adjustments to existing laws where necessary. This strategy underscores Switzerland's preference for flexibility. By avoiding premature, specialised rules in a fast-moving field, policymakers hope to maintain innovation-friendly conditions while ensuring a high level of protection. At the same time, such an approach increases the importance of Switzerland's core regulatory instruments.

Two laws in particular, the **Product Liability Act (PrHG)**[2] and the **Product Safety Act (PrSG)**[3], will become central to governing the risks of AI-enabled products. These laws were designed for an earlier technological era, yet they now carry the weight of addressing accidents, malfunctions, and safety challenges stemming from digital and self-learning systems. The situation is complicated by the fact that the EU modernised its product liability and product safety regimes in 2024 specifically to address digitalisation and AI[4]. To remain compatible with European standards—and to preserve legal certainty for manufacturers, consumers, and trading partners—Switzerland must update its own legislation. Experts have long pointed to shortcomings in both the PrHG and PrSG, and the Federal Council has now initiated revision processes. Several issues are particularly urgent.

1. Modernizing the Product Liability Act (PrHG)

The PrHG, in force since 1994, is based on an EU directive from 1985 and focuses on static, physical products. Today's digital landscape—characterised by cloud-based software, updates issued long after purchase, and AI systems that evolve autonomously—requires a more contemporary framework.

<u>Expanding the definition of "product"</u>
Current Swiss law applies only to tangible objects, leaving uncertainty about whether standalone software, including cloud-delivered AI models, falls under product liability rules. The EU's 2024 Product Liability Directive explicitly includes software[5]. To ensure legal clarity and alignment with key trading partners, the PrHG should adopt an expanded definition that includes digital products.

1 This article is based on the following scientific publication: Wildhaber/Barth (2025). Regulierung der künstlichen Intelligenz in der Schweiz, Modernisierungsvorschläge für die Produkthaftung und die Produktsicherheit im europäischen Kontext, Aktuelle Juristische Praxis (AJP), 598–614. https://www.alexandria.unisg.ch/handle/20.500.14171/124106.

2 Bundesgesetz über die Produktehaftpflicht (Produktehaftpflichtgesetz, PrHG): https://www.fedlex.admin.ch/eli/cc/1993/3122_3122_3122/de.

3 Bundesgesetz über die Produktesicherheit (PrSG): https://www.fedlex.admin.ch/eli/cc/2010/347/de.

4 Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC (Text with EEA relevance): https://eur-lex.europa.eu/eli/dir/2024/2853/oj/eng.

5 Cf. Article 4 Definitions: "For the purposes of this Directive, the following definitions apply: (1) 'product' means all movables, even if integrated into, or inter-connected with, another movable or an immovable; it includes electricity, digital manufacturing files, raw materials and software;" https://eur-lex.europa.eu/eli/dir/2024/2853/oj/eng.

Recognizing new forms of damage
The EU reforms recognise "data damage"—the loss, destruction, or corruption of data[6]—as compensable harm. This reflects data's economic and societal significance. Swiss law does not explicitly address this category of damage. A revised PrHG should clarify if and when data damage is recoverable.

Liability for defects emerging after market launch
Traditional liability rules assume a product's condition is fixed at the moment it enters the market. For digital products, however, manufacturers retain control through ongoing updates. Defects may therefore emerge later, and responsibility may lie with whoever controls the update process. The EU now incorporates this principle, and Switzerland should adopt similar rules that acknowledge the continuous influence of manufacturers and other economic operators.

Reassessing the development-risk defence
Manufacturers may currently avoid liability if risks were scientifically unknowable at the time of product release. With AI, however, risks are often known abstractly but unpredictable in practice. The EU has tied the development-risk exemption to a manufacturer's duty to monitor products and supply updates[7]. Switzerland may need to reassess the scope of this defence, given ongoing manufacturer control over AI systems.

Easing the burden of proof
Injured parties often lack the technical means to prove defects in complex AI systems. The EU addresses this by introducing disclosure obligations for manufacturers and presumptions of defectiveness and causality when complexity prevents plaintiffs from producing evidence[8]. While Swiss courts already apply mitigated burdens of proof in some cases, codifying clearer rules would strengthen fairness and accessibility.

2. Updating the Product Safety Act (PrSG)

Product safety rules determine what may enter the market at all, making the PrSG essential for managing AI-related risks. The Federal Council has already announced a partial revision to align Swiss rules with the EU's new Product Safety Regulation. Several elements merit particular attention.

Clarifying whether software qualifies as a "product"
To maintain coherence with the PrHG and guarantee legal certainty, the revised PrSG should explicitly include standalone software within its scope.

Updating the safety concept for digital products
European law now treats cybersecurity, digital connectivity, and AI's learning and predictive functions as integral components of product safety[9]. Incorporating similar criteria into Swiss law would ensure that assessments reflect actual digital-era risks.

Reconsidering update obligations and recall remedies
The EU requires manufacturers to provide effective, free remedies in recall situations, potentially including software updates to remedy safety defects. Whether Switzerland should adopt such obligations—an intervention that intersects with contractual warranty law[10]—will be an important policy question in the upcoming PrSG revision.

---

6 *Ibid*. Article 6 Damage.

7 *Ibid*. Article 7 Exemption from liability, al. 2b and 2c.

8 *Ibid*. Article 10 Burden of proof.

9 Ibid. For instance: (32) "[…] Consequently, a manufacturer that designs a product with the ability to develop unexpected behaviour should remain liable for behaviour that causes harm. In order to reflect the fact that in the digital age many products remain within the manufacturer's control after being placed on the market, the moment in time a product leaves the manufacturer's control should also be taken into account in the assessment of a product's safety. A product can also be found to be defective on account of its cybersecurity vulnerability, for example where the product does not fulfil safety-relevant cybersecurity requirements. […]".

10 Articles 197 et seqq. of the Bundesgesetz betreffend die Ergänzung des Schweizerischen Zivilgesetzbuches (Fünfter Teil: Obligationenrecht): https://www.fedlex.admin.ch/eli/cc/27/317_321_377/de.

Conclusion

Switzerland's decision to avoid AI-specific legislation reflects a long-standing commitment to technology-neutral regulation. Yet this strategy will only succeed if the **PrHG and PrSG are swiftly modernised**.[11] Updated definitions of "product" and "defect," rules that reflect continuous manufacturer control, and modern evidentiary standards are essential. Together, these reforms will help Switzerland maintain public trust in AI and ensure the resilience of its legal system in a rapidly evolving technological landscape.

University of St. Gallen, FAA-HSG, Müller-Friedberg-Strasse 8, 9000 St. Gallen.

Email: isabelle.wildhaber@unisg.ch

Website: https://www.unisg.ch/person-id/0537275e-9872-472e-b1c6-e6e2f3541616/

Isabelle Wildhaber, LL.M., lawyer, is a Full Professor of Private and Business Law at the University of St. Gallen and Director of the Institute for Work and Employment Research (FAA-HSG), as well as a founding member of the Law & Tech Lab at the University of St. Gallen. She is the Vice-Dean of the Law School.

Her research interests lie in the intersection of contracts law, torts law and labour law with technologies, i.e. Artificial Intelligence or biotechnology.

Photo credentials: Carmen Wong Fisch

University of St. Gallen, FAA-HSG, Müller-Friedberg-Strasse 8, 9000 St. Gallen.

Email: frederic.barth@unisg.ch

Website: https://www.unisg.ch/person-id/d8b6c65f-6273-49da-a778-ca2d2775fafc/

Frédéric Barth, MLaw, LL.M., lawyer, is a doctoral candidate and research assistant at the University of St. Gallen.

Currently, he is also a Visiting Researcher at Georgetown University Law Center in Washington, D.C.

Photo credentials: Walder Wyss

11 At the time of writing, the consultation period for the draft of the modernised PrSG is scheduled from March to June 2026: https://www.fedlex.admin.ch/de/consultation-procedures/foreseen#https://fedlex.data.admin.ch/eli/dl/proj/2025/9/cons_1. Additionally, the Federal Chancellery has announced that the preliminary draft bill on AI regulation (presumably including proposals for a revised PrHG) should be made available by the end of 2026: https://www.bk.admin.ch/bk/en/home/digitale-transformation-ikt-lenkung/bundesarchitektur/kuenstliche_intelligenz.html.

How AI Might Change Education?

Pierre Dillenbourg (EPFL Lausanne)

How will AI change education? A myriad of articles have addressed this question along two axes.

Transformation of skills and competencies

The first axis concerns the transformation of skills and competencies that education should equip learners with.[1] These include for instance AI literacy, i.e. what does any citizen need to understand about AI, and critical thinking, how to give citizens the ability to assess recommendations made by AI. This transformation of education goals also concerns what skills will be made obsolete by AI. Do we need for instance to learn foreign languages if our phone correctly translates from us? The heart of the debate is here. Any prosthesis has a *de-skilling effect*, who does still know more than 2-3 phone numbers since our phones know them, or an *un-skilling effect*, you don't learn to swim if you swim too long with buoy.

AI constitutes some kind of universal cognitive prosthesis: it can achieve many roles played by our cognitive functions, and the concern is that any cognitive function that is not practiced tends to fade out. Will schools produce citizens who are not able to elaborate a well-constructed text, to summarise a complex document, to turn a text into a schema, to create a new logo, etc.? Even creative skills are being automated by AI. They are numerous alarming articles about the cognitive laziness that AI may instil and, yes, we should pay attention to this drift. The underlying fear is the unbearable prospect of a society where humans would depend upon AI agents. How to build a future where humans and AI collaborate in a way that is positive for humanity? The opacity of LLMs is detrimental to the mutual understanding that the word 'collaboration' entails. A more positive development is the use of symbolic agents to guide LLMs so that their behaviour aligns with human actions.

Development of AI-powered education tools

The second axis is the development of AI-powered education tools, mostly designed for personalised instruction.[2] An AI system may learn which methods work better for which learner profile and select the most appropriate activity at a given state. Many of these systems have been tested with positive results. Some discourses anticipate a future where a personal AI coach will follow learners from birth to death and always suggest the most appropriate learning activity. They repeatedly neglect the fact that learning is a social incubator: not only does school need to prepare citizens to live together despite their differences, but more fundamentally, our cognition is deeply social, since it relies on language. Would you envision sending your child to a school where every student works alone with an AI agent? In addition, while adapting education to the learners' needs makes a lot of sense, we also need to train students to adapt themselves to the contexts of their lives, even when those contexts do not suit them so well.

The idea of personal coaches also neglects the role of teachers in driving classroom activities. It is well established that pupils are very sensitive to the (lack of) expectations of their teachers, but AI does not have expectations, it only produces sentences that mimic expectations. Developers of learning technologies have always underestimated to role of teachers upon the effectiveness of their technology. Nonetheless, the key advantage of AI remains the scalability of personalised instruction. Many colleagues use AI-generated feedback simply to be able to provide

[1] Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., ... & Rienties, B. (2025). The promise and challenges of generative AI in education. *Behaviour & Information Technology, 44*(11), 2518-2544. https://doi.org/10.1080/0144929X.2024.2394886.

[2] Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of educational psychology*, 106(4), 901. https://doi.org/10.1037/a0037123.

frequent feedback to classes of hundreds of learners. Being able to scale up effective education is a democratic challenge.[3] The use of AI for generating contents, such as question banks or exercises, goes in the same direction.

Beyond Knowledge Transmission

Behind these debates, a deeper change in educational ideas is emerging unnoticed. I would summarise it as follows: if a machine can present any piece of knowledge (or rather, information) to learners, then education may not simply be about presenting knowledge. Of course, it is not, but we do not always lecture accordingly!

Colleagues have developed AI agents that do not answer the learners' questions directly but instead ask follow-up questions in a Socratic manner or point learners to the section of the course video recording where they could find the answers. My colleague Ola Svensson developed a system in which students do not receive explanations from the AI but are instead required to explain the algorithms to the AI, exploiting the empirically proven self-explanation effect. In other words, the pedagogical value of "desirable difficulty", a key concept in the field of learning sciences that has not spread widely among teachers, has been a popularised through some kind of proof by contradiction: providing information to learners is not where humans can outperform machines.

I am often asked how schools or universities will teach in 10 or 20 years. It would be very arrogant to pretend I know. Even if classrooms in 2026 still resemble the rectangular spaces of 1926, with a teacher's desk and kids' tables, and perhaps a few more screens, I would not claim that AI will not change many things.

Of course, the pandemic reminded us that one of the main purposes of schools is childcare: schools must look after children while parents are working. This is not disrespectful of the noble mission of schools; it is simply a societal constraint that undermines predications of a world without school.

So, what is the future? My response to schools is that they cannot do nothing. They need to start experimenting with AI through modest projects that are well adapted to their local context. No large, ambitious project – they tend to fail. Rather, small projects designed by teachers with the rest of the school community: learners, parents, and other stakeholders.

EPFL IC IINFCOM CHILI, RLC D1 740 (Rolex Learning Center), Station 20, 1015 Lausanne, Switzerland.

Email: pierre.dillenbourg@epfl.ch

Website: https://people.epfl.ch/pierre.dillenbourg; https://www.epfl.ch/labs/chili/

ORCID: 0000-0001-9455-5119

Pierre Dillenbourg holds a Master's degree in Educational Sciences from the University of Mons (Belgium) and a PhD in Computer Science from Lancaster University (UK), where he specialised in artificial intelligence for education. After working at the University of Geneva, he joined EPFL in 2002, where he directed major initiatives in digital education, including the implementation of EPFL's MOOC strategy, which attracted more than three million enrolments. He is currently Full Professor of Learning Technologies and heads the CHILI laboratory.

He has led DUAL-T, the national "leading house" for technologies in dual vocational education. He has co-founded several educational technology start-ups, including the Swiss EdTech Collider, and co-launched LEARN, EPFL's centre for learning sciences. He is an inaugural Fellow of the International Society of the Learning Sciences and has served as EPFL's Associate Vice-President for Education and interim Provost.

[3] Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T. (2026). Who Gives Feedback Matters: Student Biases Towards Human and AI-Generated Formative Feedback. *Journal of Computer Assisted Learning*, 42(1), e70153. https://doi.org/10.1111/jcal.70153.

De-Skilling and Up-Skilling in Times of Generative AI: Implications for Education in Computer Science

Hans-Georg Fill & Jesús Muñoz Cádiz (University of Fribourg)

Abstract

In this paper we examine the current landscape defined by the integration of generative artificial intelligence into computer science, with particular attention to its implications for education. The rapid adoption of generative AI systems is reshaping how tasks in computer science are performed, learned, and evaluated, raising questions about productivity, skill development, and long-term professional trajectories. By situating recent developments within broader historical patterns of automation, this work discusses both the opportunities and challenges introduced by AI-assisted workflows and highlights the need for critical frameworks to guide their responsible integration.

Keywords: generative artificial intelligence, learning, computer science, programming, software engineering

Motivation

The emerging possibilities of generative artificial intelligence (GenAI) have been driven by the public availability of large language models (LLMs). Since the release of ChatGPT, GenAI systems have shifted from a research capability to a general-purpose infrastructure based on social and technical pillars. This evolution has enabled unexpected use cases in learning, knowledge work, and software development[1].

Recent surveys and large-scale usage analyses indicate that GenAI tools are now routinely integrated into everyday workflows. This new scenario affects studying, writing, information synthesis, and coding[2], suggesting a qualitatively new level of mainstream adoption compared to earlier AI deployments.

Although there are a large number of commercial and open-access AI models with similar features, the implications of this technology for our daily lives and its long-term consequences are not yet fully understood. In the area of computer science (CS) education, we currently observe that almost all students use GenAI tools. Empirical evidence from higher education indicates that AI adoption for collaborative learning is shaped by perceived ease of use and usefulness. This adoption predicts how students distribute cognitive work between themselves and AI[3].

Similarly, in industry, GenAI tools are currently being widely evaluated. Particularly, in the field of software development, there is an ongoing investigation into whether the introduction of such tools might lead to a decline in the demand for junior developers in the future[4]. However, the results of workshops involving industry experts in software development indicate that the concept of GenAI is more likely to be negative than beneficial, e.g. in terms of deskilling, dehumanisation, or disinformation[5].

---

1 Bommasani, R., Drew A. H., Ehsan A., *et al.* (2021). On the Opportunities and Risks of Foundation Models. *arXiv*:2108.07258 [cs]. doi:10.48550/arXiv.2108.07258.

2 Chatterji, A., Cunningham, T., Deming, D.J., *et al.* (2025). How People Use ChatGPT. *National Bureau of Economic Research*. doi:10.3386/w34255.

3 Youssef Alyoussef, I., Mohammed Drwish, A., Adel Albakheet, F. (2025). AI Adoption for Collaboration: Factors Influencing Inclusive Learning Adoption in Higher Education. *IEEE Access* 13:81690–81713. doi:10.1109/ACCESS.2025.3567656.

4 Acharya, V. (2025). Generative AI and the Transformation of Software Development Practices. *arXiv*: 2510.10819 [cs.SE]. https://arxiv.org/abs/2510.10819.

5 Woodruff, A., Shelby, R., Kelley, P., G., *et al.* (2024). How Knowledge Workers Think Generative AI Will (Not) Transform Their Industries. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, New York, NY, USA, 1–26. doi:10.1145/3613904.3642700.

The growing adoption of GenAI in software development thus creates a dual dynamic. While GenAI systems offer substantial benefits in terms of efficiency, accessibility, and automation, concerns have emerged regarding their potential impact on skill development. In particular, when a significant part of tasks and analytical reasoning is delegated to GenAI, opportunities for hands-on problem solving, critical reflection, and experiential learning may be reduced[6]. This shift raises questions about how foundational competencies are acquired and maintained over time, particularly among early-career developers.

As traditional pathways for skill accumulation and professional growth increasingly involve automated assistance, it is necessary to consider how junior developers can continue to develop the expertise, judgment, and autonomy required to progress towards senior roles in GenAI-augmented development environments.

Historical Perspective on Automation

The use of machines and tools to automate manual processes is not a new phenomenon. Previous examples include the use of machines for automating manufacturing workflows, such as automatic looms, industrial robotic systems for the automotive industry, and computer numerical control (CNC) systems in precision engineering.

Already in previous times, it was recognised that automation still requires human operators, which is somewhat ironic given the goal of automation to replace humans. The essay of L. Bainbridge[7] described it as "ironies of automation", which is nowadays even more crucial given the fast GenAI improvements. Building on this perspective, these ironies are revisited in the context of AI-assisted design[8], where the focus of automation shifts primarily from physical or procedural tasks towards cognitive and linguistic processes.

We consider this perspective particularly relevant for CS tasks, where the manner and motivation of GenAI use may critically influence learning outcomes and professional skill development. Unlike previous generations of AI, which focused only on specific tasks using machine learning or natural language processing, GenAI performs a wide range of cognitive activities, from conceptual engineering to algorithm design and includes emergent, unexpected behaviour as well. These capabilities raise important questions about the role of humans in software development and the implications for CS education.

Generative AI for Tasks in Computer Science

GenAI has emerged and transformed many fields in a very short period of time. In CS, it automates a wide range of tasks, including conceptual activities such as requirements engineering, conceptual modeling, architecture development, algorithm design, and analysis[9,10]. Rather than functioning solely as code completion tools, these systems increasingly serve as interactive partners that extend human problem-solving, allowing developers to focus on higher levels of design, creativity, and decision-making[11].

However, safely automating repetitive tasks requires a solid understanding of fundamental programming principles. Without such knowledge, even simple functionalities, such as code completion or inline text suggestions, may be misapplied. At the same time, key concerns in the adoption of

[6] Kosmyna, N., Hauptmann, E., Yuan, Y.T., *et al.* (2025). Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. *arXiv*: 2506.08872. doi:10.48550/arXiv.2506.08872.

[7] Bainbridge, L. (1983). Ironies of automation. *Automatica* 19, no. 6 (November): 775–779. doi:10.1016/0005-1098(83)90046-8.

[8] Shukla, P., Bui, P., Levy, S.S., *et al.* (2025). De-skilling, Cognitive Offloading, and Misplaced Responsibilities: Potential Ironies of AI-Assisted Design. *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–7. CHI EA'25. New York, NY, USA: ACM, April. doi:10.1145/3706599.3719931.

[9] Cheng, H., Husen, J. H., Lu, Y., *et al.* (2026). Generative AI for Requirements Engineering: A Systematic Literature Review. *Software: Practice and Experience*, 56 (2): 141–170. doi:10.1002/spe.70029.

[10] Fill, H.G, Fettke, P., Köpke, J. (2023). Conceptual Modeling and Large Language Models: Impressions from First Experiments with ChatGPT. *Enterp. Model. Inf. Syst. Archit. (EMISAJ) - International Journal of Conceptual Modeling*. 18: 3. doi:10.18417/emisa.18.3.

[11] Sergeyuk, A., Golubev, Y., Bryksin, T., *et al.* (2025). Using AI-based coding assistants in practice: State of affairs, perceptions, and ways forward. *Information and Software Technology* 178: 107610. doi:10.1016/j.infsof.2024.107610.

GenAI include reliability issues such as hallucinations, risks of over-reliance and skill erosion, limited integration into existing workflows, and unresolved questions about data protection and intellectual property[12].

In recent years, we have assisted in the growth of GenAI for coding tasks, e.g., code generation, code analysis and review, code adaptation, and debugging. The benefits of its implementation in terms of time-saving are evident, while the learning impact still needs to be better understood. This will require greater collaboration between all those involved in education.

Recent empirical studies have begun to address the need for systematic measurement of how individuals engage with GenAI. As an example, the work of Zhang *et al.*[13] introduces GAIES, a validated multidimensional instrument to capture not only the frequency of GenAI use but also qualitative differences in interaction style, including questioning behaviour, expressiveness, and preciseness. Drawing on frameworks from human-computer interaction and motivational psychology, the scale distinguishes between self-interest-oriented and task-oriented usage. Other works[14], related to software testing education using LLMs, show the potential of AI tools in developing industry-relevant skills.

Implications for Computer Science Education

Recent qualitative evidence highlights that while educators are generally optimistic about the potential of GenAI to enhance teaching, learning, administration, and assessment, its effective integration in school settings is constrained by systemic and pedagogical challenges[15]. These challenges high-light the need for educational curricula that emphasise critical evaluation of AI-generated
content.

The aspects to be considered for both teachers and trainers in CS education are many, as different technical aspects are involved in the process. Some of them are described below:

- Monitor the current developments in GenAI closely; development is blazingly fast, thus always try to get access to the most recent AI models and techniques; some companies offer access for educators for free or at reduced prices, e.g., Microsoft via the GitHub educator license[16].
- Inform students about the potential risks of GenAI in terms of de-skilling. Our experience so far shows that students are grateful for being made aware of these aspects, which are not yet widely considered, and are willing to change their behaviour.
- Maintain offline examinations to develop certain skills, similar to mathematics education, where electronic calculators may only be used at a certain proficiency level.
- For advanced professionals, show the possibilities of GenAI and how to achieve the best results in coding; today, this would include, for example, the use of agentic AI and spec-driven development. This requires access to best practices, common issues, and complexities in implementation.

Additionally, GenAI can be used to challenge students in their learning activities if used in this way, e.g., through guardrails that prevent the generation of final solutions but rather guide students on how to solve problems[17]. These kinds of tools have already been tested, and we strongly believe

[12] Banh, L., Holldack, F., Strobel, G. (2025). Copiloting the future: How generative AI transforms Software Engineering. *Information and Software Technology*, 183:107751. doi:10.1016/j.infsof.2025.107751.

[13] Zhang, D., Tan, J. Y., Chew, Y. Y., *et al.* (2025). Development and validation of the generative AI engagement scale. *Computers in Human Behavior: Artificial Humans*, 6:100221. doi:10.1016/j.chbah.2025.100221.

[14] Haldar, S., Pierce, M., Capretz, L.F. (2025). Exploring the Integration of Generative AI Tools in Software Testing Education: A Case Study on ChatGPT and Copilot for Preparatory Testing Artifacts in Postgraduate Learning. *IEEE Access* 13:46070–46090. doi:10.1109/ACCESS.2025.3545882.

[15] Ng, D. T. K., Chan, E. K. C., and Lo, C., K. (2025). Opportunities, challenges, and school strategies for integrating generative AI in education. *Computers and Education: Artificial Intelligence* 8:100373. doi:10.1016/j.caeai.2025.100373.

[16] Link to the platform: https://github.com/education.

[17] Bastani, H., Bastani, O., Sungu, A., *et al.* (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122, no. 26 (July): e2422633122. doi:10.1073/pnas.2422633122.

guardrails will play a pivotal role in future scenarios.

On the industry side, the availability of senior developers requires the development of many skills over time, and whether such investments pay off in the long run must be considered. Accordingly, GenAI should be used to augment learning and development processes while preserving the central role of human expertise in software engineering.

Conclusions

In this paper, we examined key aspects of the new scenario arising from the widespread use of AI. It is crucial to carefully consider how junior developers use these tools to ensure their safe and trustworthy use. The current state of the field reflects a growing awareness of the behavioural, cognitive, and educational implications of GenAI adoption.

In conclusion, addressing these factors is essential to ensure that GenAI enhances, rather than undermines, long-term learning outcomes and sustainable software engineering practices. The integration of GenAI into educational and professional learning contexts requires moving beyond tool adoption toward deliberate institutional design. Cooperation between both sides is the key to defining a safe, robust, and exciting framework for future achievements.

University of Fribourg, Department of Informatics, Bd de Pérolles 90, 1700 Fribourg, Switzerland.

Email: hans-georg.fill@unifr.ch

Website: www.unifr.ch/inf/digits/fill.html

ORCID: 0000-0001-5076-5341

Hans-Georg Fill is a full professor for business informatics and head of the research group Digitalization and Information Systems at the University of Fribourg, Switzerland. He holds a Ph.D. and a habilitation from the University of Vienna, Austria.

His research areas include metamodeling, digitalization, artificial intelligence, and augmented and virtual reality.

Since 2022, he is the speaker of the Special Interest Group on Modeling Enterprise Information Systems (GI-MobIS) in the German Informatics Society.

Photo credentials: Hans-Georg Fill

University of Fribourg, Department of Informatics, Bd de Pérolles 90, 1700 Fribourg, Switzerland.

Email: jesus.munozcadiz@unifr.ch

Website: www.unifr.ch/inf/digits/munoz.html

ORCID: https://orcid.org/0000-0003-1386-5743

Jesús Muñoz-Cádiz is a postdoctoral researcher in the Digitalization and Information Systems Group at the University of Fribourg and a member of the Smart Living Lab. He holds a PhD in Architecture, Built Environment, and Construction Engineering at the Polytechnic University of Marche (Italy).

His research focuses on data structures within the Building Information Modelling (BIM) methodology, including standards and metadata schemas. His work incorporates machine learning and ontological frameworks to enhance interoperability through digital environments.

Photo credentials: Jesús Muñoz Cádiz

ETH zürich — Assistant Professor (Tenure Track) of Sustainable Plant Nutrition

The Department of Environmental Systems Science (www.usys.ethz.ch) at ETH Zurich invites applications for the above-mentioned position.

https://ethz.ch/de/die-eth-zuerich/arbeiten-lehren-forschen/faculty/faculty-affairs/ausgeschriebene-professuren/systemorientierte-naturwissenschaften/assistenzprofessur--tenure-track--fuer-sustainable-plant-nutriti.html

Please apply online: www.facultyaffairs.ethz.ch **The closing date for applications is 28 February 2026**.

ETH zürich — Professor or Assistant Professor (Tenure Track) of Statistics

The Department of Mathematics (www.math.ethz.ch) at ETH Zurich invites applications for the above-mentioned position at the Seminar for Statistics (www.math.ethz.ch/sfs).

https://ethz.ch/de/die-eth-zuerich/arbeiten-lehren-forschen/faculty/faculty-affairs/ausgeschriebene-professuren/naturwissenschaften-und-mathematik/professur-oder-assistenzprofessur--tenure-track--fuer--statistic.html

Please apply online: www.facultyaffairs.ethz.ch **The closing date for applications is 22 February 2026**.

ETH Zurich is an equal opportunity and family-friendly employer, values diversity, and is responsive to the needs of dual-career couples.

Applications should include a curriculum vitae; a list of publications; three statements on a) research, b) teaching, c) leadership; descriptions of the three most important achievements; and a certificate of the highest degree. The letter of application should be addressed **to the President of ETH Zurich, Prof. Dr. Joël Mesot.**

Impressum | Mentions

Bulletin VSH-AEU, 50. Jahrgang / 50e année © 2025, ISSN 1663-9898, Nichtkommerzielle Verwendung mit Quellenangabe gestattet (CC BY 4.0) Herausgegeben mit Unterstützung der Schweizerischen Akademie der Geistes- und Sozialwissenschaften (SAGW).

Member of Swiss Academy of Humanities and Social Sciences

Vereinigung der Schweizerischen Hochschuldozierenden

VSH AEU

Association Suisse des Enseignant-e-s d'Université

Supported by Swiss Academy of Humanities and Social Sciences

Herausgeber und Verlag | Editeur

Vereinigung der Schweizerischen Hochschuldozierenden
Association Suisse des Enseignant-e-s d'Université
Associazione Svizzera dei Docenti Universitari

Generalsekretariat | Secrétariat général

Prof. Dr. Stephan Morgenthaler
Ecole Polytechnique de Lausanne (EPFL), SB IMA STAP, MAB 1473 (Bâtiment MA)
Station 8, 1015 Lausanne

Email : stephan.morgenthaler@epfl.ch | Homepage: www.vsh-aeu.ch

PC-Konto / CPP 80-47274-7 | IBAN: CH38 3000 0001 8004 7274 7

Redaktion | Rédaction

Dr. phil. Franziska Schumacher | Email: bulletin@vsh-aeu.ch

Layout

® APRON REVOLUTION | Email: contact@apronrevolution.org

Druck | Imprimerie

Canisius AG, Avenue Beauregard 3, 1700 Fribourg, Suisse

Anzeigen | Annonces

Generalsekretariat

Preise | Prix:

–	Stellenanzeigen | Postes à pourvoir:	CHF 250 (½ Seite/page) CHF 500 (1 Seite/page)
–	andere Annoncen | autres annonces:	CHF 500/1000

Vorstand am 1. August | Comité directeur au 1er août 2022

Präsident | Président

Prof. Dr. rer. nat. Christian Bochet
Université de Fribourg, Département de chimie
Chemin du musée 9, 1700 Fribourg,

Tel.: 026 300 8758, Email : christian.bochet@unifr.ch

Vorstandsmitglieder | Membres du Comité directeur

Prof. Dr. Nikolaus Beck
Università della Svizzera italiana, Institute of Management
Via G. Buffi 13, 6900 Lugano
Email: nikolaus.beck@usi.ch

Prof. Dr. Norbert Lange
Université de Genève et Université de Lausanne, Ecole de Pharmacie, Sciences II
Quai Ernest Ansermet 30, 1211 Genève 4,
Email : norbert.lange@unige.ch

Prof. Dr. (Ph.D.) Stephan Morgenthaler
Ecole Polytechnique de Lausanne (EPFL), SB IMA STAP, MAB 1473 (Bâtiment MA)
Station 8, 1015 Lausanne
Email: stephan.morgenthaler@epfl.ch

Prof. Dr. med. Dr. phil. Hubert Steinke
Universität Bern, Institut für Medizingeschichte
Bühlstrasse 26, 3012 Bern
Email: hubert.steinke@img.unibe.ch

Prof. Dr. Helmut Zander
Universität Freiburg, Theologische Fakultät, Dep. Glaubenswissenschaft
und Philosophie, Av. de l'Europe 20, 1700 Fribourg
Email: helmut.zander@unifr.ch

Bulletin VSH-AEU erscheint drei- bis viermal im Jahr und wird gratis an die Mitglieder versandt. Abonnement: CHF 65 pro Jahr inkl. Versand Schweiz.

Bulletin VSH-AEU est publié trois à quatre fois par an et est distribué gratuitement aux membres. Abonnements : CHF 65 par an, frais de port compris en Suisse.

Bulletin VSH-AEU è pubblicato tre o quattro volte l'anno e viene distribuito gratuitamente ai membri. Abbonamenti: CHF 65 all'anno, comprese le spese di spedizione in Svizzera.

Die Stimme

der Hochschuldozierenden

Vereinigung der
Schweizerischen Hochschuldozierenden

VSH
AEU

Association Suisse
des Enseignant-e-s d'Université

La voix

des Enseignant·e·s d'Université