Generative AI as cultural transmission technology: Challenges and opportunities for the education of future citizens

Pierre-Yves Oudeyer
Inria Centre of the University of Bordeaux
http://www.pyoudeyer.com/
March 2025

Reference: Oudeyer, P-Y. (2025) Generative AI as cultural transmission technology: Challenges and opportunities for the education of future citizens, Consensus Conference "New Knowledge and New Skills for Young People" of Cnesco, Oct. 2024.[1] *This text was written for non specialist audiences in the domain of education.*

Although it has appeared very recently, generative AI has radically transformed our digital landscape, and society at large, becoming perhaps the most rapidly adopted technology in human history. This unprecedented technological wave is sweeping through our educational institutions, with over 80% of secondary school students already using generative AI, either for school or outside.

We stand at a critical juncture where these powerful systems are reshaping cultural, political, economic, and educational domains before we fully comprehend their implications. The scientific community is struggling to keep pace; we are navigating what amounts to a terra incognita, where our knowledge is dwarfed by the magnitude of open questions. What makes this particularly concerning is that young people—the citizens of tomorrow—are adopting these technologies at extraordinary rates, embedding them into their learning processes and worldviews.

At their core, these systems function as unprecedented cultural transmission technologies, compressing and encoding societal values, knowledge, and biases from vast corpora of human-created content. They then amplify and propagate these cultural patterns at scale, shaping how people access information, form beliefs, and engage with ideas. This poses profound challenges for human culture and democratic societies in particular—where the intentional or unintentional encoding of certain worldviews within these systems could

[1] This text was initially written and published in French (here) for the Cnesco consensus conference entitled "New Knowledge and New Skills for Young People". Translation and adaptation in english was made with the help of generative AI.

influence public discourse, reinforce certain power structures, or even determine which voices and perspectives are amplified or suppressed.

This presents an urgent educational imperative: How do we prepare students not just to use these systems, but to understand them critically? What frameworks must we develop to address the multifaceted challenges they present—from the proliferation of sophisticated misinformation and the homogenisation of thought to environmental costs and workforce disruption? Equally important, how might we harness their potential benefits in supporting personalised learning, enhancing creativity, democratising access to knowledge, and improving work? Could these technologies, if thoughtfully guided, actually strengthen democracy, e.g. by making diverse cultural knowledge more accessible and bridging informational divides? The answers to these questions will fundamentally shape how our society evolves alongside these increasingly powerful technologies.

1. What is Generative AI (GenAI)?

While many secondary school students think that artificial intelligence (AI) refers to systems that appeared 5 or 10 years ago, it is a much older term and scientific field, and generative AI (GenAI) designates only one form of AI among others. The expression "artificial intelligence" was invented in 1955, shortly after the invention of computers, to designate a scientific field that studies the mechanisms of cognition and learning in living beings (particularly humans) using computer models, that is, by simulating certain aspects of these mechanisms. Today, many people use the word "AI" as a linguistic shortcut to designate the systems or machines manufactured by researchers and engineers in this field, and we will sometimes also use this shortcut in this document.

There are several forms of AI systems, and some have been present for a long time in our daily lives (Russell and Norvig, 2016). **Behavioural AI** models the role of interaction between certain sensorimotor reflexes, the body and the environment: this is the case, for example, of Gray Walter's Elmer and Elsie robots, created in 1949 (Walter, 1950; Cordeschi, 2002), whose light-attraction reflexes made it possible to reproduce complex behaviours very similar to those of many insects. A limitation of this approach is the difficulty in simulating abstract cognitive processes. **Symbolic AI** models a domain, such as chess or draughts, with a set of symbols and logical rules, and uses heuristics hand-programmed by engineers to calculate and deduce optimal action plans to solve a problem (Samuel, 1959). A limitation of this approach is its inability to adapt to new situations not anticipated by the engineers. **Statistical AI** refers to a set of techniques allowing a machine to learn new know-how by identifying regularities in data, through statistical calculations (Lecun et al., 2015). These techniques are also called "machine learning". For example, with "**supervised" learning**, and if we have access to a database associating photos with labels indicating what is in the photos (e.g., "cat", "dog", "plane", etc.), it is possible to develop a software system that will be able to predict the label associated with a new image. With "**reinforcement" learning**, a robot, for example, can learn a new movement strategy to grasp an object: at the start it tries random movements and measures their score for grasping the object, and it will gradually refine the parameters of the movements that have the best score (for example by trying small random variations of the best movements). These different AI techniques are already present for a long time in many contexts, for example on the internet in algorithms recommending a restaurant or a film, in driving assistance software for cars or planes, in logistics software, in voice or facial recognition software, or to program the behaviour of non-player characters in video games. However, until 4 or 5 years ago, all these approaches were still very limited in their capabilities, particularly in mastering language and tasks that are primarily expressed and solved in natural language.

The field of **generative AI (GenAI)** has brought a major evolution that has pushed back many of these limits (Brown et al., 2020), and explains the societal impact we will discuss below. Generative AI is a particular form of statistical AI, which became known to the general public with the release of ChatGPT 3.5 in 2022, followed by other software such as Midjourney, DALL-E, Mistral, Claude or Gemini more recently. These are software capable of producing text, images or sound from "prompts", that is, texts or images that are often used to give the software a context and an instruction or a question related to this context. For example, we can ask a GenAI questions such as "What are the monuments to visit as a priority in Paris?" or "What is the big bang?", and specify to whom the response should be addressed, e.g., "Write a response suitable for primary school children" or "Write a response for an adult with a scientific background". It is also possible to give a GenAI a pdf file, for example a history course, and ask for instance "Can you summarise this course?" or "Ask me questions to check that I have learned this course well". Finally, the result can also be an image, for example if we ask "Make a drawing showing a penguin skiing on an artificial snow slope in Paris".

Until recently, most of these tasks were performed very poorly by AI software. But their capabilities have considerably and very rapidly increased in just a few years, in a way that was unpredictable even for most scientists working in this field.

How do these GenAIs work and what has enabled these major developments? GenAIs are software which, given a text/image/sound as input, perform calculations to produce a text/image/sound as output. These calculations are carried out by billions of small elementary modules each performing relatively simple calculations and interacting with each other. These calculations are determined by internal parameters (numbers) for each small module. Initially, these parameters are random. From there, an algorithm will evolve them during "training". This training consists of two phases. First, in a supervised learning phase, the GenAI software is given texts/images with gaps, created from billions of texts found for example on the internet (and hiding bits of text or image to make the gaps). The GenAIs must then predict what text/image to place in the gaps. Each time they do this type of exercise, the parameters are reinforced if they allowed them to guess correctly what to put in the gap, and otherwise the parameters are modified very slightly to increase the probability of giving the right answer the next time. In a second phase of reinforcement learning (Ouyang et al., 2022), millions of tasks are given to GenAIs (for example summarising a text or answering a question), and the productions of the GenAIs are scored by human scorers. Based on these scores, the parameters are updated to increase the probability of getting a better score the next time. In short, GenAIs are software trained to produce the most probable words and images, and which will get the best scores from human scorers, given prompts in the form of texts and images.

The training of GenAIs thus consists of learning to do relatively elementary exercises. If we train GenAIs this way that have few parameters (a few million) and with few texts with gaps (a few million), then no notable capability appears. If on the contrary we train GenAIs that contain billions of parameters with billions of texts/images with gaps, then new and diverse capabilities appear quite suddenly. From a scientific point of view, the appearance of these capabilities in this context is still largely a mystery. To measure these capabilities, scientists use benchmarks, which are tests of knowledge, reasoning and know-how in many domains (Srivasta et al., 2023; Chang et al., 2024). Some of these tests are even directly tests made initially for university exams or competitions for human students. Thus, recent experiments on software such as GPT-4, Claude Sonnet 3.5, Llama 3 70B or Mistral Large have shown that they were capable of obtaining excellent marks for university exams in law, mathematics, or computer science, which would allow them to obtain degrees in these subjects. While some of these exams or competitions include knowledge tests and mainly require rote learning, others include exercises that have never been seen by these software during their training, and require advanced reasoning for humans: the ability of language models to perform these tasks thus shows a relatively powerful form of generalisation. However, as we will see below, this can also lead GenAIs to assert information or reasoning that is completely false with confidence.

GenAIs are cultural technologies. These processes for manufacturing GenAI software thus rely on two pillars: 1) learning the regularities appearing in the text/image/sound databases that enable training them; 2) the scoring of human evaluators. Thus, we can see GenAIs as systems that compress and encode the cultural regularities that are present in the text/image corpora (and human feedback) on which they are trained (Hershcovich et al., 2022; Bender et al., 2021; Johnson et al., 2022): they reproduce them when they are used (in the context of direct discussions with a human, or by other machines that will then simulate human populations, for example bots on social networks), which leads to their amplification. The large GenAI models (language, visual, multimodal) are thus fundamentally tools of cultural transmission: they are in fact models of culture.

More precisely, they encode cultural regularities along many dimensions, ranging from values, usual knowledge, socio-cultural norms, repertoires of concepts defining what is salient or not, interesting or not, true or not (Hershcovich et al., 2022). These regularities include all forms of **bias, that is, stereotypes**, which can be problematic for society, particularly for minority groups (Bender et al., 2021; Johnson et al., 2022).

Hallucinations and biases. While their behaviours and capabilities are often impressive (for example correctly answering a university-level mathematics problem), the same GenAI software can also **make major errors on elementary questions** (for example answering incorrectly to a primary school maths problem). This is referred to as **hallucinations or confabulations**. There are several reasons for this. First, GenAI systems are trained on very diverse texts from very varied sources: many of them contain errors, but also biases or even propaganda information deliberately intended to influence the thinking of their readers: these errors, biases and propaganda are thus encoded and reproduced by language models. Another major reason is that, as explained above, they are not trained to answer "correctly", but to give the most probable responses given the text or image databases used during their training (or to predict the responses that would probably get good marks from human scorers). Thus, GenAIs have no intrinsic notion of true or false (and have no metacognition, Guilleray et al., 2024), and are also made to try to guess the answers to new questions, which were not exactly present as such in their training.

This is what allows them to generalise, for example when it comes to summarising a completely new text, and in this case it is useful. It is also what leads them to fabricate facts that are not facts at all. Beyond factual errors, it is these same reasons that lead GenAI software to reproduce biases frequently present in the texts used to train them (e.g., gender, race, religion, professions, etc.), with the risk of amplifying their harmful consequences for diverse populations, particularly for minorities.

Many approaches are being explored today to better understand and limit these biases, for example by asking scorers to give bad marks when they observe productions that reflect them. This is called "alignment", meaning that we wish to "align" the behaviours of GenAIs with the values and cultural preferences of the groups of humans who use them (Ji et al., 2023). This poses numerous challenges. A recent example that helps understand the complexity is that of a GenAI system that was asked to generate an illustration of the Nazi army during the Second World War: this illustration showed a troop including women and people whose skin colours were diverse. In this case, the GenAI software followed mechanisms that guided it not to generate "bias" in representations (here related to the army): in doing so, it produced a result that is a factual error that distorts the representation of a real event. It would therefore be necessary that in this case the software could distinguish between a request that concerns a historical fact (a notion the objectivity of which is sometimes debated among historians and philosophers), and a request that is not related to a historical fact, which is still far from being the case today. This example also illustrates that beyond very complicated technical problems (Ji et al., 2023), the question of alignment also poses strong political issues: what are the cultural regularities that we wish to encode in this or that GenAI software, and what are the biases that we wish to preserve or on the contrary avoid?

Deepfakes: misinformation and privacy. The errors (hallucinations) of GenAIs are often unintentional. However, it is also possible to use these software to deliberately create texts or images that stage real people or places, but in situations that have not existed. In the case of images or videos, the quality of GenAI software today allows generating scenes for which it is almost impossible to guess whether it is a real image or an invented image: these are called deepfakes. The ability to produce **deepfakes** is unfortunately used today on a large scale for malicious purposes, whether by individuals or large organisations. Some states use them for massive disinformation campaigns. For example, in 2022, during Russia's invasion of Ukraine, a doctored video of Volodymyr Zelensky showed him asking the Ukrainian population to surrender. This type of usage has become common during elections, including in Western "democratic" states, and is amplified by the use of "bots", software that simulate real people on social networks to propagate this false information. This illustrates the major civic issues associated with them. Other malicious uses include commercial scams or false advertising. Finally, deepfakes are also being used increasingly by individuals, including teenagers, to stage, sometimes ridiculously, public figures or some of their acquaintances. Another use consists of staging oneself or using these software to show modify one's appearance and show an image of oneself that does not correspond to reality. These uses can thus have serious consequences on the lives of the people concerned.

The new and very broad capabilities of GenAI technologies, but also their nature that makes them powerful tools of cultural transmission, which includes a whole set of misinformation mechanisms, thus pose today to society in general, and for young and future citizens in particular, great challenges and great opportunities that we will discuss below.

2. The Global Societal and Educational Issues of Generative AI

2.1 The Massive and Growing Use of Generative AI.

A few months after the release of the GPT-3 model from the company OpenAI, which marked a technological and scientific turning point, the ChatGPT 3.5 model was launched in November 2022: it reached 1 million users in 5 days, 100 million in two months, and today it has 180 million. Many other software have appeared, also allowing the production of images (e.g., MidJourney with 20 million users, DALL-E or Stable Diffusion) or videos (e.g., Sora or Runway). While the first GenAI software were "private", that is, their internal parameters were secret, a growing number of institutional and individual actors have begun to develop and share "open-weights" models, meaning they shared the code of these models on the internet for free use. For example, the Hugging Face platform today hosts more than 400,000 "open-weights" GenAI models, that is, software whose entire code and parameters are accessible and freely downloadable.

In France, according to an Ifop/Talan survey (2024), nearly 70% of 18-24 year olds personally use GenAI software, compared to only 47% of 25-34 year olds and 22% of those aged 35 and over. Users estimate they gain 38% in productivity and efficiency thanks to generative AIs. In particular, 46% of 18-24 year olds estimate this productivity increase at more than 40%. Finally, 44% of users (and 61% of 25-34 year olds) use the results of generative AIs as they are without modifying them and 35% declare that they would have difficulty doing without generative AIs. A pilot survey in Nouvelle Aquitaine among high school students, conducted by the Flowers team at Inria in June 2024, indicates that more than 90% of tenth grade students have already used generative AI software to help them with their homework. Moreover, a large proportion of them, after having experimented with ChatGPT, use it directly afterwards to search for information, without going through traditional search engines. Overall, in France the increase in the number of users has been 60% in one year, and it is mainly a massive use by younger generations.

In the professional world, the uses of generative AI by employees are very diverse, ranging from help with writing emails or reports, generating responses to customer requests, automatic summarisation of reports or meetings, translation, help with brainstorming or problem solving, semantic analysis of written productions, help with classifying CVs or funding applications, help with producing graphic illustrations, or help with programming for computer developers.

Among individuals in general, a large proportion of users use generative AI mainly for recreational purposes. However, among younger users, who are also the biggest users, the use of this software for homework help is growing.

Overall, content produced with generative AIs is also very widely shared online, both because individuals share the content they have produced with GenAIs, but also because we observe a massive and growing use of organisations deploying GenAI software on a large scale on social networks, simulating real users, and thus producing content read by both humans and other GenAI software. According to Europol, 90% of online content could have been generated by generative AI by the end of 2026 (Europol, 2022).

2.2 Generative AI at School: Today's Challenges

Several converging studies indicate that in Europe and North America, more than 80% of 14-18 year olds have already used ChatGPT to do their homework. Among them, an American study indicates that 38% did so without telling their teacher (CommonSense Media and Impact Research, 2023). These uses may consist of asking ChatGPT to solve their maths exercises, propose compositions in French or history, translate a text into a foreign language or prepare a presentation in physical sciences. At university, usage has also become massive: for example, the 2024 De Vinci Higher Education, RM conseil and Talan study, conducted among 1600 fourth-year university students in management, engineering and computer science, shows that 92% of them use GenAI regularly, and 30% of them pay a subscription of 20 euros per month to have access to the best versions. Moreover, 65% of them say that the presence of generative AI will be a major criterion in choosing the companies in which they will want to work.

This poses several major challenges for students and teachers. First for students, when they use generative AI for homework, for example, the risk is that these tools are used in a way that short-circuits the cognitive effort necessary for effective learning (Kasneci et al., 2023; Abdelghani et al., 2023). More precisely, the image of "super-intelligence" conveyed in many media, combined with the confident tone of GenAI software (although they are incapable of metacognition, that is, incapable of evaluating their own uncertainties), can lead many students to overestimate both the skills of GenAIs and their own skills, limiting the development and expression of their curiosity, critical thinking and metacognition which are nevertheless essential for effective and motivating learning (Abdeghani et al., 2023; Oudeyer et al., 2016). These effects are amplified by the absence of a pedagogical posture in the behaviour of GenAI: indeed, they have been trained to predict the most probable words and images, as well as to respond as directly and efficiently as possible to users' questions. Consequently, when a student asks them a question or gives them an exercise, they will have a very strong tendency to give the answer right away, instead of giving pedagogical clues to help and allow the learner to make the effort to find the answer by themselves (Macina et al, 2023; Jurenka et al., 2024).

These challenges are related both to the nature of generative AI software, but also to the cognitive biases of student learners and their limited understanding of these systems (Kidd and Birhane, 2023). First, the human brain tends to use its estimation of the level of competence of other individuals to decide which information and beliefs to adopt or question when these individuals express them (Orticio et al., 2023). Moreover, humans also have cognitive biases favouring the attribution of agency and intention to objects (Heider and Simmel, 1944), and this applies particularly to GenAI. Finally, the human brain is also biased in such a way that once certain knowledge or beliefs have been acquired from sources it believed solid, it is then difficult to correct these beliefs (Thompson and Griffiths, 2021). These three biases combined thus lead humans to tend to attribute agency to GenAIs, to think that their knowledge is solid given their assured and affirmative tone, and thus potentially learn false information, which is particularly problematic given that GenAIs encode many stereotypes (e.g., racial, gender, religious, etc.) (Bender et al., 2021).

The role of these biases is also illustrated by a recent experiment conducted with a group of business school students in France during a behavioural economics course (Hill, 2023). Each student had to solve two case studies (chosen randomly from 14): one for which he/she had to find the answer by himself/herself, and the other for which an answer was already provided, and he/she had to correct or improve it. In the second case, students were given answers to correct that were produced either with ChatGPT or by a student from the previous year. The students were informed of the procedure, but did not know whether a given answer came from another student or from ChatGPT. They were also told that ChatGPT's responses were often of poor quality. The results were very clear: students scored on average 28% higher in the first case (without a proposed answer) than in the second (answer to correct/improve). The qualitative analysis of their productions showed a great difficulty in moving away from the initial proposition, and in particular from those provided by ChatGPT, which nevertheless were of lower quality than if they had answered without this answer proposition.

The educational challenges related to the limits in understanding GenAI software are also illustrated by a recent study conducted in several colleges in Nouvelle Aquitaine (Abdelghani, 2024). In this study, science exercises were given to 72 students in 8th and 9th grade in four colleges: each exercise included an illustration and a short text presenting an observation of a natural phenomenon, and they had to research and write a short explanation. ChatGPT was the research tool that was made available to them, and the objective of the study was to understand how they formulate their questions and to what extent they manage to formulate a context allowing the software to give a relevant response. Indeed, the ability to formulate precise and informative investigation questions is essential in learning processes, and is associated with learners' metacognitive abilities (Abdelghani, 2024). First, the data collected shows that 73% of these students have already used ChatGPT. Moreover, they have great confidence in ChatGPT's responses: 82% of them think they are reliable. At the same time, only 33% of them say they do not know the limits of ChatGPT. Next, the experimentation shows that their ability to formulate relevant and well-contextualised questions (or choose the right questions among several proposed), is weak (in 49% of cases they choose a question that is not suitable, which is the equivalent of a random choice). Finally, 79% of participants ask only one question to ChatGPT and do not question the accuracy of its response, leading them to a low success rate on exercises (43%).

The challenges are also major for teachers. First, the growing and massive use of GenAIs for help with completing homework makes their evaluation by teachers very difficult. In particular, while AI software tools have appeared with the objective of automatically detecting texts or images generated by AI, many scientists agree that this objective is almost impossible to achieve (Oravec, 2023): new GenAI software appears permanently, in particular with functionalities to circumvent the automatic detection of their use, and with a high risk of detecting false positives, that is, attributing to a GenAI software texts genuinely written by humans.

Overall, the arrival of GenAI has led many teachers to modify their teaching practices. In a large-scale study conducted among 908 primary and secondary teachers in Estonia (Laak and Aru, 2024), it is observed that the arrival of GenAI has led 49% of teachers to modify their practices, by eliminating a large part of homework, and including activities promoting critical thinking. However, this study, as well as other converging studies (e.g., the Impact Research 2024 study conducted among 1003 teachers in the United States, and Klopfer et al., 2023), also show uses evaluated as positive by teachers (74% of them in the study by Laak and Aru, 2024), particularly to help them in the preparation of creative and motivating activities on a subject, or the development of course plans and associated exercises/quizzes, or finally to respond to emails sent by parents. On the other hand, the use of this software to help them evaluate student work is judged not very relevant and not very effective by teachers (Impact Research, 2024). Finally, many of them would like to receive more training on GenAI and how to take it into account and include it in teaching.

Generative AI and education: opportunities for tomorrow? Beyond these major challenges associated with the role of generative AI in school learning, these technological developments could also bring various educational opportunities.

As explained above, generative AI is only one approach to AI among many others, and there is a long tradition of work combining AI, cognitive sciences and educational sciences, having led to the development of educational software proposing learning activities based on strong cognitive and/or pedagogical principles and using AI for more accessibility and personalisation (Nkambou et al., 2010). For example, recent work in the Flowers team at Inria has led to the development of a personalisation algorithm for exercise sequences based on the principles of our understanding of the mechanisms of curiosity-driven learning in children (Oudeyer et al., 2016). Experiments in primary school with about 1000 children have shown that this approach allowed, compared to exercise sequences made by hand by an expert in mathematics didactics, to significantly improve learning efficiency and motivation, particularly for students differing from the "average" student (Clement et al., 2015, Clément et al., 2024). This approach was then transposed into the AdaptivMaths software, developed as part of the French P2IA programme, and supported by the Ministry of Education (https://www.adaptivmath.fr/).

In the same way, a growing part of the scientific community working on the development of pedagogically relevant uses of generative AI, has begun to study ways to use generative AI for the benefit of students, teachers, and more broadly the educational ecosystem. Let's take some examples that illustrate this diversity and these perspectives. In a study conducted among 8762 students from 146 countries, enrolled in an online introductory computer programming course, the authors studied the effect of making available a free version of GPT-4 parameterised to provide pedagogical help to students when they asked it questions about the course (Nie et al., 2023). The study first showed that the impact of the availability of GPT-4 was very different between countries with a high Human Development Index (HDI) and countries with a low human development: in high HDI countries, **engagement and (optional) participation of students** in the exam decreased significantly compared to a control group that did not have access to GPT-4 (the study does not allow understanding why), while it **increased significantly in low HDI countries**. Furthermore, the study shows that the **exam score of participants who adopted the use of GPT-4 was significantly higher compared to those who had not used it**. These results, some aspects of which are still difficult to interpret, nevertheless show the positive role that GenAI could have in populations from countries whose educational system is poorly developed.

Generative AI systems are also being studied to **help generate personalised content in educational software**: exercises, personalised hints and feedback, and even explanations. In an extension of the above study, conducted among the 8762 students from 146 countries, the authors compared, for programming exercises, the effectiveness of error messages generated by classical methods with error messages generated with GPT (Wang et al., 2024). They showed that error messages generated with GPT led students to learn to solve the exercises significantly faster.

Another example is the study presented in (Abdelghani et al., 2023), which studies the use of GPT-3 to **pre-generate exercises and training hints for formulating curious questions, and associated metacognitive skills**. Indeed, these skills play a key role in learning in general. Here, the software was tested with a group of 75 children aged 9-10 in primary schools in Nouvelle Aquitaine. Based on short texts on science subjects, the students had the objective of formulating complex questions related to the text but whose answer was not in the text. The study compared the effectiveness of syntactic and semantic hints generated by hand by human experts, with hints generated with GPT-3: it concluded that the hints generated with GPT-3 had a quality at least equal to, and an impact at least as positive as, those made by hand (and even a greater positive impact for a category of hints).

Another use of generative AI is being experimented in the GPTeach project (Markel et al., 2023): **simulating the behaviour of diverse students to train apprentice teachers to use pedagogical strategies**. More specifically, the system studied simulates students with varied personalities, needs and learning objectives, and scripts teaching situations that can include a group of simulated students. Although this study involves a small number of participants (24), the results show a very positive appreciation from these apprentice teachers: these training situations allow them to try and repeat pedagogical interventions without the pressure they might have to find themselves immediately in front of students in a real classroom.

2.3 Generative AI: a Wide Variety of Uses, from Sciences to Arts

As mentioned above, generative AI is now increasingly used in the professional world. In companies, a major use is writing assistance. A randomised study, conducted among 453 professionals occupying varied functions (e.g., consultants, human resources, data analysts, managers), for example showed that the use of ChatGPT for common tasks (emails, summary reports, press releases) made it possible to achieve a time saving of 40%, while increasing quality by 18% on average (Noy and Zhang, 2023). With other uses such as decision support, or access to a company's knowledge bases, generative AI is thus causing major transformations in the world of work, involving numerous challenges for employees and organisations themselves (GPAI, 2023).

Beyond business, uses are also numerous in various sectors. **In the field of sciences, generative AI is beginning to open extraordinary perspectives to help physicists, chemists, biologists, or mathematicians make new discoveries**. For example, many laboratories today are studying the use of generative AI systems to efficiently generate new chemical or physical structures (Park et al., 2024), opening the possibility of discovering for example new proteins relevant for applications in health or agriculture (Zambaldi et al., 2024), or new materials (Merchant et al., 2023). In mathematics, recent projects have shown how language models could allow exploring and finding new solutions to open problems (Romera-Paredes et al., 2024), how generative AI could also be used to help mathematicians formalise and develop theorem proofs (Wu et al., 2022), and current research is interested in how these systems could generate interesting new conjectures (Bengio and Malkin, 2023).

Generative AI can also serve other sciences such as archaeology. As part of the "Vesuvius Challenge" (https://scrollprize.org/), a team has indeed succeeded in deciphering the remains of texts on papyrus scrolls severely damaged during the eruption of Vesuvius. In ethology, the

Earth Species project explores the use of generative AI to help decode animal communication signals (https://www.earthspecies.org/), which is not without difficulties that still remain largely to be solved (Yovel and Rechavi, 2023).

Other domains in which the uses of generative AI are developing include assistance for artistic creation, particularly the creation of images, music, voices and films with systems such as MidJourney, DALL-E, HeyGen Suno AI or Sora - which raises many ethical and legal questions - or assistance in the development of video games (Bruce et al., 2024).

2.4 Cultural and Democratic Issues

Thus, the opportunities and societal challenges of generative AI, particularly seen as a set of **tools for cultural transmission and amplification**, are today very great, and are beginning to be well described and characterised in the literature. In this perspective, the development of **training corpora** (and guidance by human scoring) can be seen as a form of "education" of GenAI models, in the sense that these corpora will define the cultural orientations that will be encoded in them. There is therefore an emerging issue that is very similar to that of the education of humans for organisations, e.g., States: **what are the knowledge and values, and more generally the culture, that we wish to see encoded/learned by GenAI models? For what purposes and for what uses and users?** This is primarily a political and cultural issue, rather than a technological one. In the short term, a first step consists of understanding how the data used to train large generative AI software could be made accessible, at least to trusted third-party institutions, in order to be able to verify a certain number of cultural and legal dimensions associated with this data: this is one of the issues in the application of the recent European Digital AI Act (https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai).

While today, in most States, there is a long tradition (theoretical and practical) of methods for developing school curricula (used including by totalitarian States that wish to control the thinking of their population), there is today in the AI ecosystem no global and organised approach to this issue. It is even almost absent from most current debates, which focus on technological issues, whereas the biggest questions about LLMs are political and cultural questions.

In practice, today large private organisations (e.g. OpenAI, Anthropic, Google, Deepseek, Mistral, etc) constitute very large and heterogeneous corpora, from very varied sources that they capture on the internet, and filtered with a mixture of massive automation and human work (micro-workers). For human scoring, it is done on a very large scale by resorting to **micro-work.** In a number of cases, some of these organizations have been reported to recruit microworkers in countries where workers rights are very limited, with very low salaries and in human conditions that are highly criticisable (Gray et al., 2019). The constitution of these corpora, these modes of feedback, and the filters to "correct"/"eliminate" "unacceptable" generations are based on values and cultural models that are not sufficiently detailed publicly by these organisations, and which are designed most often internally, without clear articulation with public bodies as well as with user populations. In practice, they reflect a mixture between a vision of values and culture carried within the ecosystems of these organizations (Soleiman and Dennison, 2021; Rozado, 2022), and a superposition of values and cultures present on the internet (Kovac et al., 2023), without there being a good understanding of these values and cultures and the way they are encoded.

Several ambitious actions and projects have been developed recently to establish **more solid, and especially societally more positive, inclusive and acceptable methodologies for the constitution of corpora and modes of feedback**. In particular, we can note the work of the BigScience consortium which has developed an approach for thinking about the **governance** of the constitution of these corpora (Jernite et al., 2022), and has designed a corpus (ROOTS, Laurençon et al., 2022) with this method and shared it with the scientific community. However, this corpus has only been superficially designed and analysed from a cultural point of view: only language and a few relatively simple biases have been taken into account. A notable exception is the work presented by Johnson et al. (2022), which presents an advanced study of the cultural dimensions of the corpora used to train GPT-3 (and the values encoded by the model).

Overall, almost everything remains to be done, both from a scientific point of view and from a political and cultural point of view, to understand how to constitute corpora and approaches to human feedback that aim to align models on sets of well-understood cultural regularities. This even raises fundamental questions in the humanities to characterise these cultural regularities, which are posed in a new way with these technologies of cultural transmission and amplification.

It is also essential to keep in mind that all the concepts and techniques that aim to align large language models with particular human values are approaches and techniques that allow controlling the cultural values that will be transmitted. **In this sense, these approaches can be as useful for transmitting values and knowledge aligned for example with those of democracies, but also by organisations, small or large, state or private, that wish to influence, or even control, what targeted populations think.**

2.5 AI and Environmental Impact

The massive use of generative AI systems also poses **major environmental challenges** (Trystram et al., 2021). Indeed, each generative AI software requires very large computing resources both at the time it is trained, but also each time it is used to give a response to a "prompt". The computing centres that enable this and host supercomputers can be gigantic in size (several hectares): they occupy land in place of green spaces or housing, and consume considerable amounts of electricity. In many countries where these computing centres are hosted, electricity is produced by means of fossil energy, which thus causes the **release of large quantities of CO2 and numerous pollutants**. It is estimated that in the last 6 years, global electricity consumption related to AI has been multiplied by one million, and some studies estimate that it could reach 10% of global electricity consumption in 2030. To cool the processors, these **computing centres also consume large quantities of water** (Li et al., 2023). To cool the processors, some computing centres also consume large quantities of water (Li et al., 2023) - though new energetic design approaches aim to mitigate this aspect.

Alongside these negative impacts, projects are also studying different ways in which **generative AI could help us manage climate change** (Cowles et al., 2023). Some projects aim to simulate and better predict climate evolution and extreme events, while others aim to use GenAI to optimise energy management in buildings or transport, or finally to analyse satellite data to help farmers develop more environmentally friendly cultivation strategies.

Furthermore, many scientific projects are working today to understand how it would be possible to **develop much less energy-consuming generative AI software (what is called frugal AI)**. This involves, for example, developing, instead of a large generalist model, a set of small AI models (Touvron et al., 2023), which can be specialised for the particular needs of a group of users (Fu et al., 2023). Generally, a number of researchers and organisations are working to increase transparency in work in this field, by leading designers to measure and share the energy cost of the production and use of generative AI models.

3. Education in Generative AI Literacy: Paths and Tools

Given these societal issues, it thus appears fundamental today to develop education in generative AI, that is, to **enable children and adolescents (and beyond) to progressively acquire literacy in generative artificial intelligence**. This is not about technical education aimed at preparing them for a profession in this field: it is on the contrary about enabling them to have a sufficiently developed culture on the mechanisms, uses and societal issues, allowing them on the one hand to use these tools wisely, and on the other hand to **express in an informed**

manner their opinion as citizens to contribute to the major collective orientations related to generative AI (and therefore related to major cultural and democratic issues). In this perspective, it also seems fundamental to develop acculturation and **literacy in generative AI among teachers, and among parents in general**, both to conceptually equip adults in their role as citizens, but also so that they can contribute to building the ecosystem in which younger people will be able to develop this literacy. In short, the development of literacy in generative AI is an issue for everyone.

Several pedagogical approaches and tools are beginning to be developed, but efforts are still at the beginning: work on evaluating and adapting these pedagogical tools will be necessary over time. The Flowers team, at the Inria Centre of the University of Bordeaux, has recently developed several resources aimed at developing literacy in generative AI for the widest audience, and in particular for secondary school students and their teachers. First, the pedagogical series "**ChatGPT explained in 5 minutes**" consists of a set of short videos (between 5 and 10 minutes) aimed at introducing the mechanisms, uses and societal issues of generative AI (Torres-Leguet et al., 2024). For example, the first video introduces the notion of language model (what it does, how they are made), and introduces a summary of all the limitations of these systems (e.g., fixed knowledge, hallucinations, private economic actors). The second video focuses on the limitations, and explains in more detail the notions of bias and alignment, static and non-sourceable knowledge, the amount of data necessary for training models, hallucinations, or the problem of the lack of grounding of generative AIs in the physical world. Another video presents the strengths of language models, for example the great mastery of languages, the use of external tools to make calculations, or the way they can help brainstorm for creative projects. Finally, several videos explain the different ways to construct prompts and lead generative AIs to simulate and explain reasoning, allowing learners to ask questions by providing a context that will allow them to obtain relevant answers and better interpret these answers. Another video explains the applications in different domains (daily life, work, health, education, sciences, cultural diversity, ...). Finally, the last video discusses how various cognitive biases in human brains can lead to misuses of these technologies, or be exploited purposefully by malevolent organizations in the context of generative AI. These videos are all under Creative Commons CC-BY licence, which authorises free of charge all uses of these videos, including commercial uses or their integration into broader pedagogical tools, and are freely downloadable on the website: https://developmentalsystems.org/chatgpt_5_minutes/en/

In parallel, the book "C'est pas (moi), c'est l'IA" (It's not (me), it's AI) (Roy and Oudeyer, 2024, English translation in the works), in an illustrated format accessible from 12 years old, explains all these subjects, delving deeper into the dimensions related to education, deepfakes and misinformation, the environment, and open scientific challenges for the future. This book is associated with a website that lists a set of varied and accessible resources for better understanding this field, and in particular allowing teenagers, but also teachers, to experiment with different generative AI systems: https://github.com/flowersteam/clia

Beyond discovering the mechanisms and societal issues of generative AI, an essential challenge is to develop tools and approaches allowing teenagers to **train their critical thinking**, their ability to ask curious questions and to question information they receive, to learn actively and by themselves. These abilities are closely linked to the mechanisms of **curiosity** (Oudeyer et al., 2016) as well as those of **metacognition** (Proust, 2019), which are essential in general for learning, and particularly important when these are carried out in interaction with generative AI systems (Abdelghani et al., 2023). Although these transverse skills are still very little taken into account in school teaching, several works have begun to show how certain **exercises and pedagogical gestures can be used to train the curiosity and metacognition of children in the school context** (Abdelghani et al., 2023b; Guilleray et al., 2024). This path thus opens complementary and promising perspectives not only for the development of literacy in generative AI, but for the education of future citizens in general.

4. Acknowledgements

The writing of this article has greatly benefited from discussions and projects carried out with Didier Roy (in particular the book "C'est pas (moi), c'est l'IA" that we co-wrote), Rania Abdelghani, Chloé Desvaux and Hélène Sauzéon (in particular for studies of the use of generative AI in primary school and secondary school), Alexandre Torres-Leguet, Clément Romac and Thomas Carta (for the series of pedagogical videos "ChatGPT Explained in 5 minutes"), and all the members of the Flowers team at the Inria Centre of the University of Bordeaux. However, the opinions and potential errors present in this article are entirely my own.

5. References

Abdelghani, R. (2024) Guiding the minds of tomorrow: Conversational Agents to Train Curiosity and Metacognition in Young Learners, Thesis from the University of Bordeaux.

Abdelghani, R., Sauzéon, H., & Oudeyer, P. Y. (2023). Generative AI in the Classroom: Can Students Remain Active Learners?. In Workshop on Generative AI in Education at the Conference on Neural Information Processing Systems.

Abdelghani, R., Law, E., Desvaux, C., Oudeyer, P. Y., & Sauzéon, H. (2023b). Interactive environments for training children's curiosity through the practice of metacognitive skills: a pilot study. In Proceedings of the 22nd Annual ACM Interaction Design and Children Conference (pp. 495-501).

Abdelghani, R., Wang, Y. H., Yuan, X., Wang, T., Lucas, P., Sauzéon, H., & Oudeyer, P. Y. (2024). GPT3-driven pedagogical agents to train children's curious question-asking skills. International Journal of Artificial Intelligence in Education, 34(2), 483-518.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?🦜. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610-623

Bengio, Y., & Malkin, N. (2024). Machine learning and information theory concepts towards an AI Mathematician. Bulletin of the American Mathematical Society, 61(3), 457-469.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. In Proceedings of the 34th International Conference on Neural Information Processing Systems (pp. 1877-1901).

Bruce, J., Dennis, M. D., Edwards, A., Parker-Holder, J., Shi, Y., Hughes, E., ... & Rocktäschel, T. (2024, February). Genie: Generative interactive environments. In Forty-first International Conference on Machine Learning.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., ... & Xie, X. (2024). A survey on evaluation of large language models. ACM Transactions on Intelligent Systems and Technology, 15(3), 1-45

Cordeschi, R. (2002). The discovery of the artificial: Behavior, mind and machines before and beyond cybernetics (Vol. 28). Springer Science & Business Media

Common Sense Media and Impact Research (2023) https://www.commonsensemedia.org/sites/default/files/featured-content/files/common-sense-ai-polling-memo-may-10-2023-final.pdf

Clement, B., Roy, D., Oudeyer, P. Y., & Lopes, M. (2015). Multi-Armed Bandits for Intelligent Tutoring Systems. Journal of Educational Data Mining, 7(2).

Clément, B., Sauzéon, H., Roy, D., & Oudeyer, P. Y. (2024). Improved Performances and Motivation in Intelligent Tutoring Systems: Combining Machine Learning and Learner Choice. arXiv preprint arXiv:2402.01669.

Cowls, J., Tsamados, A., Taddeo, M., & Floridi, L. (2023). The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations. Ai & Society, 1-25.

De Vinci Higher Education, RM conseil and Talan (2024) Study 2024 "The impact of generative AI on students"

Europol (2022), Facing reality? Law enforcement and the challenge of deepfakes, an observatory report from the Europol Innovation Lab, Publications Office of the European Union, Luxembourg. https://www.europol.europa.eu/cms/sites/default/files/documents/Europol_Innovation_Lab_Facing_Reality_Law_Enforcement_And_The_Challenge_Of_Deepfakes.pdf

GPAI (2023) Future of Work Working Group Report, https://gpai.ai/projects/future-of-work/Future%20of%20Work%20Working%20Group%20Report%20v2%20(November%202023).pdf

Guilleray, F., Proust, J., Fernandez, J. (2024) Glossaire de la métacognition, Conseil Scientifique de l'Education Nationale/Réseau Canopé.

Hershcovich, D., Frank, S., Lent, H., de Lhoneux, M., Abdou, M., Brandl, S., ... & Søgaard, A. (2022). Challenges and strategies in cross-cultural NLP. arXiv preprint arXiv:2203.10020.

Ifop/Talan (2024) Baromètre 2024 Ifop pour Talan « Les Français et les IA génératives », https://www.talan.com/actualites/detail-actualites/news/barometre-2024-ifop-pour-talan-les-francais-et-les-ia-generatives/

Impact Research (2024) AI Chatbots in Schools, https://www.waltonfamilyfoundation.org/ai-in-the-classroom

Ji, J., Qiu, T., Chen, B., Zhang, B., Lou, H., Wang, K., ... & Gao, W. (2023). Ai alignment: A comprehensive survey. arXiv preprint arXiv:2310.19852.

Jurenka, I., Kunesch, M., McKee, K. R., Gillick, D., Zhu, S., Wiltberger, S., ... & Ibrahim, L. (2024). Towards responsible development of generative AI for education: An evaluation-driven approach. arXiv preprint arXiv:2407.12687.

Johnson, R. L., Pistilli, G., Menédez-González, N., Duran, L. D. D., Panai, E., Kalpokiene, J., & Bertulfo, D. J. (2022). The Ghost in the Machine has an American accent: value conflict in GPT-3. arXiv preprint arXiv:2203.07785

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and individual differences, 103, 102274.

Kidd, C., & Birhane, A. (2023). How AI can distort human beliefs. Science, 380(6651), 1222-1223.

Kovač, G., Sawayama, M., Portelas, R., Colas, C., Dominey, P. F., & Oudeyer, P. Y. (2023). Large language models as superpositions of cultural perspectives. arXiv preprint arXiv:2307.07870

Laurençon, H., Saulnier, L., Wang, T., Akiki, C., Villanova del Moral, A., Le Scao, T., ... & Jernite, Y. (2022). The bigscience roots corpus: A 1.6 tb composite multilingual dataset. Advances in Neural Information Processing Systems, 35, 31809-31826.

Markel, J. M., Opferman, S. G., Landay, J. A., & Piech, C. (2023). Gpteach: Interactive ta training with gpt-based students. In Proceedings of the tenth acm conference on learning@ scale (pp. 226-236).

Nkambou, R., Mizoguchi, R., & Bourdeau, J. (Eds.). (2010). Advances in intelligent tutoring systems (Vol. 308). Springer Science & Business Media.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187-192.

Oravec, J. A. (2023). Artificial intelligence implications for academic cheating: Expanding the dimensions of responsible human-AI collaboration with ChatGPT. Journal of Interactive Learning Research, 34(2), 213-237.

Oudeyer, P. Y., Gottlieb, J., & Lopes, M. (2016). Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. Progress in brain research, 229, 257-284.

Park, H., Li, Z., & Walsh, A. (2024). Has generative artificial intelligence solved inverse materials design?. Matter, 7(7), 2355-2367.

Romera-Paredes, B., Barekatain, M., Novikov, A., Balog, M., Kumar, M. P., Dupont, E., ... & Fawzi, A. (2024). Mathematical discoveries from program search with large language models. Nature, 625(7995), 468-475.

Fu, Y., Peng, H., Ou, L., Sabharwal, A., & Khot, T. (2023, July). Specializing smaller language models towards multi-step reasoning. In International Conference on Machine Learning (pp. 10421-10430). PMLR.

Gray, M. L. and Suri, S. (2019) Ghost work: how to stop Silicon Valley from building a new global underclass. Eamon Dolan Books.

Hill, B. (2023). Taking the help or going alone: ChatGPT and class assignments. HEC Paris Research Paper Forthcoming.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. The American journal of psychology, 57(2), 243-259.

Jernite, Y., Nguyen, H., Biderman, S., Rogers, A., Masoud, M., Danchev, V., ... & Mitchell, M. (2022). Data governance in the age of large-scale data-driven language technology. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 2206-2222)

Klopfer, E., Reich, J., Abelson, H., & Breazeal, C. (2024). Generative AI and K-12 Education: An MIT Perspective.

Laak, K. J., & Aru, J. (2024, July). Generative AI in K-12: Opportunities for Learning and Utility for Teachers. In International Conference on Artificial Intelligence in Education (pp. 502-509). Cham: Springer Nature Switzerland

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). Making ai less "thirsty": Uncovering and addressing the secret water footprint of ai models. arXiv preprint arXiv:2304.03271.

Macina, J., Daheim, N., Chowdhury, S., Sinha, T., Kapur, M., Gurevych, I., & Sachan, M. (2023, December). MathDial: A Dialogue Tutoring Dataset with Rich Pedagogical Properties Grounded in Math Reasoning Problems. In Findings of the Association for Computational Linguistics: EMNLP 2023 (pp. 5602-5621).

Merchant, A., Batzner, S., Schoenholz, S. S., Aykol, M., Cheon, G., & Cubuk, E. D. (2023). Scaling deep learning for materials discovery. Nature, 624(7990), 80-85.

Nie, A., Chandak, Y., Suzara, M., Malik, A., Woodrow, J., Peng, M., ... & Piech, C. (2024). The GPT Surprise: Offering Large Language Model Chat in a Massive Coding Class Reduced Engagement but Increased Adopters' Exam Performances (No. qy8zd). Center for Open Science.

Oudeyer, P. Y., Gottlieb, J., & Lopes, M. (2016). Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. Progress in brain research, 229, 257-284.

Orticio, E., Meyer, M., & Kidd, C. (2023). Children flexibly adapt their evidentiary standards to their informational environment. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 45, No. 45).

Roy, D., Oudeyer, P-Y. (2024) C'est pas (moi), c'est l'IA, Nathan.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. Advances in neural information processing systems, 35, 27730-27744.

Proust, J. (2019). La métacognition: les enjeux pédagogiques de la recherche, in: S. Dehaene (ed.), Les sciences au service de l'école. Odile Jacob.

Rozado, D. (2022) The political orientation of the ChatGPT AI system Applying Political Typology Quizzes to a state-of-the-art AI Language model, https://davidrozado.substack.com/p/the-political-orientation-of-the

Russell, S. J., & Norvig, P. (2016). Artificial intelligence: a modern approach. Pearson.

Salesforce Generative AI statistics (2024) https://www.salesforce.com/news/stories/generative-ai-statistics/

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. IBM Journal of research and development, 3(3), 210-229.

Solaiman, I., & Dennison, C. (2021). Process for adapting language models to society (palms) with values-targeted datasets. Advances in Neural Information Processing Systems, 34, 5861-5873.

Srivastava, A., Rastogi, A., Rao, A., Shoeb, A. A. M., Abid, A., Fisch, A., ... & Wang, G. X. (2023). Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models. Transactions on Machine Learning Research.

Thompson, B., & Griffiths, T. L. (2021). Human biases limit cumulative innovation. Proceedings of the Royal Society B, 288(1946), 20202752.

Torres-Leguet A., Romac, C., Carta, T., Oudeyer, P-Y. (2024) ChatGPT en 5mn: une série pédagogique pour le grand public. https://developmentalsystems.org/chatgpt_5_minutes/fr/

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... & Lample, G. (2023). Llama: Open and efficient foundation language models.arXiv:2302.13971.

Trystram, D., Couillet, R., Ménissier, T. (2021) Apprentissage profond et consommation énergétique : la partie immergée de l'IA-ceberg, The Conversation, https://theconversation.com/apprentissage-profond-et-consommation-energetique-la-partie-immergee-de-lia-ceberg-172341

Walter, W. G. (1950). An electro-mechanical «animal». dialectica, 206-213.

Wang, S., Mitchell, J., & Piech, C. (2024, March). A large scale RCT on effective error messages in CS1. In Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1 (pp. 1395-1401).

Wu, Y., Jiang, A. Q., Li, W., Rabe, M., Staats, C., Jamnik, M., & Szegedy, C. (2022). Autoformalization with large language models. Advances in Neural Information Processing Systems, 35, 32353-32368.

Yovel, Y., & Rechavi, O. (2023). AI and the Doctor Dolittle challenge. Current Biology, 33(15), R783-R787.

Zambaldi, V., La, D., Chu, A. E., Patani, H., Danson, A. E., Kwan, T. O., ... & Wang, J. (2024). De novo design of high-affinity protein binders with AlphaProteo. arXiv preprint arXiv:2409.08022.