European Education Area Strategic Framework

Working Group on Digital Education: Learning, Teaching and Assessment

PLA, 30 May – 1 June 2023

Input paper: Generative AI for School Teachers

Generative AI for School Teachers

Introduction

In late November 2022, the world was introduced to ChatGPT, an Artificial Intelligence (AI)-based chatbot system based on GPT-3.5 that uses natural language processing (NLP) to generate conversations[1]. ChatGPT was the latest in a series of such tools but, unlike previous tools, it has captured the public interest and imagination. It had more than a million users within a week of its launch because of its ability to generate human-like text and its perceived implications and potential use in education, the workplace and everyday life. ChatGPT can answer questions and assist you with tasks like composing emails, essays, and code[2].

GPT-3 (Generative Pre-trained Transformer-3) is a Large Language Model (LLM) that was trained through pattern recognition on a vast quantity of data (499 billion data points – 800GB of data), a model some one hundred times larger than any previous models[3], and ChatGPT enables human-like conversations with this model. Put very simply, a Large Language Model is trained to predict the next word in a sentence, in somewhat the same way that a text message autocomplete feature works, and then to continue this process to generate larger quantities of text.

Users need to be aware that Large Language Models such as ChatGPT are weak on logical and mathematical tasks and have a tendency to 'hallucinate' facts and make reasoning errors (though the use of language, assigning such human like attributes to ChatGPT, rather hides what is really happening in terms of the processing). However, Large Language Models can be combined with standard data search in order to improve accuracy (and this has been done by Microsoft in incorporating GPT-4 with Bing[4], and by Google in developing Bard[5]) and can also be combined with quite different kinds of AI systems such as Wolfram Alpha[6] (which is based on computational approaches, rather than on pattern recognition) in order to better handle mathematical tasks.

ChatGPT is just one example of Large Language Models, which in turn are examples of 'Generative AI', which is a category of artificial intelligence (AI) algorithms that generate new outputs based on the data they have been trained on[7].

What is Generative AI?

Generative AI systems create new content in the form of images, text, audio, video, and more, for example:

- Images: Generative AI can create new images based on existing ones, such as creating a new portrait based on a person's face or a new landscape based on existing scenery
- Text: Generative AI can be used to write news articles, poetry, and even scripts. It can also be used to translate text from one language to another
- Audio: Generative AI can generate new music tracks, sound effects, and even voice acting[8]

---

[1] https://www.forbes.com/sites/bernardmarr/2022/12/21/chatgpt-everything-you-really-need-to-know-in-simple-terms/?sh=16674aadcbca

[2] https://www.zdnet.com/article/what-is-chatgpt-and-why-does-it-matter-heres-everything-you-need-to-know/

[3] https://hai.stanford.edu/news/how-large-language-models-will-transform-science-society-and-ai

[4] https://www.bbc.com/news/business-64562672

[5] https://www.bbc.com/news/technology-64546299

[6] https://writings.stephenwolfram.com/2023/03/chatgpt-gets-its-wolfram-superpowers/

[7] https://www.weforum.org/agenda/2023/02/generative-ai-explain-algorithms-work/

[8] https://www.weforum.org/agenda/2023/02/generative-ai-explain-algorithms-work/

These tools have captured the public imagination and people, in all walks of life, are exploring their possibilities. One of the reasons for the popularity of ChatGPT is that you can type natural-language prompts, such as "create an essay" and the system can create conversational, if somewhat stilted, responses[9].

Great claims have been made for the potential of AI in general and of Generative AI in particular – Bill Gates summarises these in his blog, *The Age of AI has begun*[10].

Sometimes these systems are able to perform tasks they were not explicitly trained on, for example ChatGPT is able to translate between some languages even though it was not trained on translation data. However, this may only be the tip of the iceberg, and some believe that these Large Language Models have much greater capabilities and these will emerge over time as the models grow in size and sophistication. The surprising abilities of ChatGPT have led to conflicting arguments about the implications of its abilities to generate very human sounding output, which have even (mis)led some to think of it as alive[11].

The linguist Emily Bender famously described ChatGPT as a stochastic parrot[12], and Chomsky similarly argues: "The human mind is not, like ChatGPT and its ilk, a lumbering statistical engine for pattern matching, gorging on hundreds of terabytes of data and extrapolating the most likely conversational response or most probable answer to a scientific question. On the contrary, the human mind is a surprisingly efficient and even elegant system that operates with small amounts of information; it seeks not to infer brute correlations among data points but to create explanations."[13]

An example of experimental work supporting these perspectives is the work of Trotte et al.[14] who provide empirical evidence from comparing the responses to the False Belief task to human participants to and GPT-3and that that "while statistical learning from language exposure may in part explain how humans develop Theory of Mind, other mechanisms are also responsible."

Some authors (but not many) have gone the other way, suggesting that the success of ChatGPT may actually tell us something about human language, indicating that it is not as complex as we think it is and that humans may actually produce language in ways not too dissimilar from large language models, and that that we may need to move to a position in which " the distinction between true understanding and mere pattern recognition gives way to something new".[15]

Another perspective positions humans and AI systems as working in dialogue, arguing that "we are intelligent primarily in networks and systems rather than as lone individuals" and that "the best model for interacting with AI may well be one of human/machine pairing – where our own intelligence is augmented, not replaced by AI.[16]

Some examples to try

Generative AI tools are becoming more and more popular and some such as ChatGPT, are becoming so popular that it is impossible to access the free service at certain times of the day. However, there is nothing

[9] https://www.cnet.com/tech/computing/why-were-all-obsessed-with-the-mind-blowing-chatgpt-ai-chatbot/

[10] https://www.gatesnotes.com/The-Age-of-AI-Has-Begun

[11] https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/

[12] https://nymag.com/intelligencer/article/ai-artificial-intelligence-chatbots-emily-m-bender.html

[13] https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

[14] https://arxiv.org/abs/2209.01515

[15] https://www.noemamag.com/gpts-very-inhuman-mind/

[16] https://theconversation.com/has-gpt-4-really-passed-the-startling-threshold-of-human-level-artificial-intelligence-well-it-depends-202856

like trying out these tools to see what they can create in terms of text, images, music and much more. A good place to start is to use ChatGPT[17], Perplexity[18] or Bing Chat or Google's Bard (as of time of writing you must join a waiting list for both of these, and Bard is only available in UK and US) and start to engage in a conversation. Pick a topic you are interested in and ask the chatbot a direct question such as "*Tell me about*" or "*Create a lesson plan or a quiz about*". You might also ask the chatbot to proofread a paragraph of text and ask it to show you what changes it made. The better your prompts the better the responses will be, and you should strive to provide as much context as possible and use specific and detailed language. For example, you might include information about:

- your desired focus, format, style, intended audience and text length.
- a list of points you want addressed
- what perspective you want the text written from, if applicable
- and specific requirements, such as no jargon[19].

Moving away from text to images you can use prompts to create images generated from text descriptions, such as "the face of a shiba inu[20] blended into the side of a loaf of bread on a kitchen bench, digital art"[21]. Some tools you can try are: DALL.E 2[22], Stable Diffusion[23], Google's Imagen[24] and Bing Image Creator[25] that can generate images from text. You might also like to try out Google's Quick, Draw![26], which asks you to draw a doodle and it then attempts to recognise it.

How can Generative AI be used in education?

Generative AI tools have the potential to transform how we work, how we teach, how we learn and how we assess. In the world of work these tools are now being embedded into existing applications such as email, word processing, spreadsheets, presentation tools etc. to increase productivity and support us in generating and analysing information. We are already using AI tools as part of our word processing software to enhance our expression and flow, and there is an expectation that these features will become more sophisticated and powerful over time[27].

In education these tools have the potential to help teachers manage their administrative work, to generate lesson plans, to provide more personalised responses to student inquiries or provide students with instant feedback and support.

These tools can eliminate '*writers block*' for both the teacher and the student. It can assist teachers and students in the following ways:

- Drafting and brainstorming for lesson plans and other activities
- Help with the design of quiz questions or other exercises

---

[17] https://openai.com/blog/chatgpt

[18] https://www.perplexity.ai/

[19] https://theconversation.com/how-to-perfect-your-prompt-writing-for-chatgpt-midjourney-and-other-ai-generators-198776

[20] **Shibas** are currently one of the most popular companion dogs in Japan.

[21] https://theconversation.com/ai-art-is-everywhere-right-now-even-experts-dont-know-what-it-will-mean-189800

[22] https://openai.com/product/dall-e-2

[23] https://stablediffusionweb.com/#demo

[24] https://imagen.research.google/

[25] https://www.bing.com/create

[26] https://quickdraw.withgoogle.com/

[27] https://www.nytimes.com/2023/02/01/technology/personaltech/chatgpt-ai-bots-editing.html

- Providing grammatical or structural feedback on portions of writing
- Customising materials for different preferences (simplifying language, adjusting to different reading levels, creating tailored activities for different interests)[28]

These tools can very quickly "mindlessly generate text"[29], which has led to concerns over plagiarism. This has led to some jurisdictions and institutions seeking to ban these tools, though other are seeing this as an opportunity to rethink how they teach and assess student learning[30]. With the upcoming incorporation of tools based on these technologies by Microsoft and Google into their search engines, and office applications the prospect of effective bans on their use appear likely to be futile. Perhaps a more effective way forward is to see these new tools as partners and helpers rather than replacements for either teachers or students work. For example, these tools can quickly synthesise large datasets and identify patterns and trends that could inform teaching strategies and curriculum development.

The good news is that there are many groups of teachers getting together online to develop approaches to using these new technologies – join one!

The Risks associated with Generative AI

The Technical Report[31] which accompanied the release of GPT-4 (the latest version of GPT) in March 2023, provides some insights into potential risks. The first indicator is that unlike in reports for earlier versions of GPT they provide no data on the training set or training methods for this version – citing both competition and safety concerns. Some specific indicators in the report are:

- GPT-4 was able to do things the designers did not expect and did not predict – an example given relates to 'Hindsight neglect' (p.4)
- When given unsafe inputs, the model may generate undesirable content, such as giving advice on committing crimes (p. 12)
- Indications of some proficiency in generating text that favours autocratic regimes when prompted to do so in multiple languages (p.51)
- The emergence of concerning novel capabilities "including the ability to create and act on long-term plans, to accrue power and resources ('power-seeking'), and to exhibit behaviour that is increasingly 'agentic'." (p. 54) One of the studies referenced in the Technical Report highlights "the increasingly evident fact that ML systems are not fully under human control"[32]

Such risks have led to an open letter entitled "Pause Giant AI Experiments" signed by a large number of AI researchers and backers (including Elon Musk) calling for a pause in the development of these models. The call talks of "AI labs locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one – not even their creators – can understand, predict, or reliably control" and argues that there are "profound risks to society and humanity."[33]

A more wide-ranging discussion of potential risks of Large Language Models was set out in paper by Bender and her co-authors in 2021[34], and they responded to this open letter writing: "It is indeed time to act but the focus of our concern should not be imaginary "powerful digital minds." Instead, we should focus on the very

[28] https://nationalcentreforai.jiscinvolve.org/wp/2023/03/14/getting-started-with-chatgpt/

[29] https://nymag.com/intelligencer/article/ai-artificial-intelligence-chatbots-emily-m-bender.html

[30] https://www.nytimes.com/2023/01/16/technology/chatgpt-artificial-intelligence-universities.html

[31] https://cdn.openai.com/papers/gpt-4.pdf

[32] https://arxiv.org/abs/2302.10329

[33] https://futureoflife.org/open-letter/pause-giant-ai-experiments/

[34] https://dl.acm.org/doi/10.1145/3442188.3445922

real and very present exploitative practices of the companies claiming to build them, who are rapidly centralizing power and increasing social inequities."[35]

On the other side of this argument, see Donald Clark's response entitled 'Moral panic and AI regulation'[36] opposing bans on AI development, arguing that we already have sufficient regulation in place.

What are the dilemmas Generative AI presents for teachers?

Generative AI will have an impact on all the media inputs around you, on education and on your students' everyday lives. The boundaries between human created and machine created media will be further blurred and this will necessarily impact on forms of teaching and assessment and provide further challenges for the development of digital competencies.

Whilst it is easy to get Generative AI systems to produce some output, the task of producing good prompts is not trivial and there is much to learn about how this is best done, though there are a number of resources available to support teachers in doing this[37].

In using Generative AI in ways suggested in the previous section, the limitations of the tools need to be kept in mind, with the user (whether teacher or student) constantly reviewing what is generated to ensure it is appropriate for the particular context. Large Language Models ultimately derive their responses from the data they were trained on and so can reflect all the errors that may be found there, and they do not have the ability to, for example, consult the internet to update their information (though increasingly they will be combined with other systems to enable this to be done – as we see in the new Bing search and in Bard for Google).

Reports giving empirical evidence of the impact on using Large Language Models are starting to be published, for example a paper by Noy and Zhang[38] provides evidence of the impact of using ChatGPT on occupation specific writing tasks of college educated professionals, demonstrating both productivity and quality gains.

There is a growing trend to embed Generative AI into existing educational products. Examples include Khan Academy and Duolingo which are actively working on incorporating these technologies into their products. Duolingo is using them to provide additional grammatical explanations, and to generate role play conversations. Khan Academy is exploring how ChatGPT-4 can act as a virtual tutor and help learners when they get stuck or have a question[39]. Such technologies will be likely to spread to many other educational products and this will throw up challenges for teachers in integrating these into their teaching. The promises may however not be achieved easily, a recent small empirical study[40] examining the relative effectiveness of ChatGPT and human tutor generated algebra hints, found that the human created hints resulted in substantial and statistically significant better results than ChatGPT hints.

Hopes have been expressed that ChatGPT will be able to help with marking. In a number of small-scale studies[41] [42] Christodoulou has explored the value of ChatGPT-4 in marking children's essays with rather poor results – finding poor agreement between human judging and GPT marks; poor predictive validity for GPT

---

[35] https://www.dair-institute.org/blog/letter-statement-March2023

[36] http://donaldclarkplanb.blogspot.com/2023/03/moral-panic-and-ai-egulation-primer.html

[37] For example: https://drive.google.com/file/d/1q9exc7gm3DpRAygeV8hgZ-7sVnyrTq6b/view

[38] https://economics.mit.edu/sites/default/files/inline-files/Noy_Zhang_1.pdf

[39] https://www.youtube.com/watch?v=rnIgnS8Susg

[40] https://arxiv.org/abs/2302.06871

[41] https://substack.nomoremarking.com/p/can-gpt-3-mark-writing-the-data-is

[42] https://substack.nomoremarking.com/p/more-gpt-marking-data-is-it-better

marks; and clustering of marks given by GPT. Another study she carried out on using ChatGPT to provide written feedback[43] on essays found that ChatGPT provided feedback that was as good as that provided by teachers –though she was critical of the value of the feedback by both teachers and ChatGPT.

The success that ChatGPT-4 has had on taking public examinations has raised some questions about the value of these examinations and of the curriculum they are assessing. Some caution is needed here as the great success of ChatGPT on the Bar Exam, for example, has been questioned[44], raising the possibility that it was being tested on data it had been trained on.

The linking of Wolfram Alpha to ChatGPT leading to success in a UK Maths-A level exam has led to the claim that high school maths may need to be completely revamped[45] though Christodoulou argues that to some extent this is a misreading of what assessment is for[46].

The use of AI tools in computer programming[47] is raising questions about the nature of computer science education. Coupled with this is the observation that the use of systems like ChatGPT relies not on traditional programming but on scripting 'prompts' (and libraries of such prompts are now available online[48]). It just might be that at school level there should be a reconsideration of the role of coding and computational thinking in digital competences and increased emphasis on the rather different skill of writing prompts for Generative AI.

Addendum to the Paper based on the May 4th online Meeting on Generative AI

This highly informative event shared multiple perspectives and practices on the role generative AI is and can play in education. AI is no longer something we encounter in scientific publications or movies, it is here today and we can engage with across all levels of education. The meeting stressed that there is a need to engage with AI, and specifically generative AI, and not ignore it hoping that it fades away, because it is real and it is already impacting on education. Multiple speakers and contributors noted that schools have a key role to play in ensuring that young people are familiar with AI and on how it can be used for good in their lives. It was noted that we are at the beginning of a journey and that we have much to learn from each other, particularly from those involved in the DELTA WG by engaging in focused conversations around specific AI issues. AI has the potential to transform many existing aspects of our education and training systems, yet there was a call to **critically consider what we might transform and why,** and for greater collaboration among WG members on these issues. Having reflected on the discussions it appears that the following issues should be further considered by DELTA members, in their own organisations and collectively with colleagues, in the future.

---

[43] https://substack.nomoremarking.com/p/can-chatgpt-provide-feedback-our

[44] GPT-4 and professional benchmarks: the wrong answer to the wrong question (substack.com)

[45] https://www.conradwolfram.com/writings/game-over-for-maths-a-level

[46] https://substack.nomoremarking.com/p/if-we-are-setting-assessments-that-a-robot-can-complete-what-does-that-say-about-our-assessments-cbc1871f502

[47] https://www.theregister.com/2022/10/20/ai_programming_tools_mean_rethinking/

[48] https://promptbase.com/

- AI is here to stay and we need to ensure we educate our teachers and our students about what it is and how it can be used ethically and legally in society, and specifically how it might be used to enhance administrative, teaching, learning and assessment practices. The provision of appropriate education and training experiences will be key for educators to ensure they are confident and competent in using and developing AI literacy in schools. It was suggested that schools should provide a safe space where teachers can experience AI, using approaches such as Project Based Learning, to develop their competences and those of their students in this area. While there was broad agreement on this issue, it raises questions around **what kinds of supports**, member states and other organisations, are providing to schools so that educators are equipped with the **confidence and competence to use AI effectively** (i.e. this includes pedagogical, ethical and legal approaches) in their work. Furthermore, it raises questions around how best to capture and share this information between DELTA members.
- The meeting reiterated that ChatGPT has captured the public imagination in many countries since its launch in November 2022. The meeting showcased multiple ways in which it can be used to support teaching, learning and assessment practices, specifically in teaching English as a second language. The examples showcased how creative teachers can use generative AI tools for good with their students. Others mentioned that it could potentially transform how we teach other subjects in schools, such as language teaching and mathematics. There appears to be a need to explore and identify practices/areas where AI can work well in education and specifically **where it can support** school leaders, teachers and students carry out their activities. In addition there also seems to be a need to outline practices/areas where AI applications should **not** be used in schools. The capturing and sharing of such practices could again be facilitated through European collaboration.
- There has been a significant focus in the media, and elsewhere, on how generative AI is disrupting many traditional forms of assessment across all levels of education and training. It was noted during the meeting that the Schools WG is in the process of developing an online course on the role of AI in Formative Assessment for this September. The use of generative AI tools can support learning when used appropriately, but they can prevent learning when used inappropriately. There appears to be a need to further explore this issue and to better understand how member states are dealing with these challenges and in particular if summative assessment practices are being altered because of the arrival of generative AI. Ultimately there is a need to consider how the advent of generative tools might impact on **how schools perceive and design assessment tasks in the future**, in light of recent developments.
- Much of the work around AI is currently being funded by US multinational companies, such as Open AI, Microsoft and Google. They are predominantly developing Large Language Models (LLM) for the English language and there is a fear that European languages may suffer in the future. During the recent meeting there were calls for the creation of European Large Language Models (LLM) to counteract such developments. The creation of these LLM models, based on English, could have implications for EU languages, specifically **how we use technology to teach and to promote the use of all our languages in the future**. There is potentially a need to better understand what these possibilities are and to fully understand what the creation of such LLMs for Europe might entail for member states and specifically education and training organisations.

- Ministries of Education are facing many questions and challenges around the uses of AI, specifically generative AI in schools. The French DELTA WG member outlined how they are dealing with this ongoing complex challenge, by adapting a flexible model that revolves around bestowing trust in teachers, scientists and legal experts to ensure that AI is used ethically and legally in schools. Other member states, such as Luxembourg, have also developed approaches to AI focussing on *learning with, about and despite AI*. It was noted that there is a need for greater collaboration and sharing of approaches among member states around their approaches to AI. It was suggested that there is a need to capture how ministries are dealing with AI at present, in general and in education, so that member states can learn from colleagues and devise collaborative responses, where possible.