

Exploring the Use of GPT-4 in Co-creating Personalized Case Scenarios for Higher Education.

Pablo Flores^{1*}, Guang Rong¹ and Benjamin Ultan Cowley^{1,2}

^{1*}Faculty of Educational Sciences, University of Helsinki,
Siltavuorenpenger 5, Helsinki, 00014, Finland.

^{2*}Cognitive Science, Faculty of Arts, University of Helsinki.

*Corresponding author(s). E-mail(s): pablo.flores@helsinki.fi;
Contributing authors: guang.rong@helsinki.fi ; ben.cowley@helsinki.fi;

Abstract

In the fast evolving landscape of Artificial Intelligence in Education (AIED) Large Language Models (LLMs) like GPT-4 have emerged as powerful and versatile tools capable of adapting for a wide set of natural language tasks. This study explores into an interaction design where higher education students utilized GPT-4 to craft personalized case scenarios tailored for their coursework. This process not only facilitates tailored learning experiences, but also engaged students into speculative reflections about the potential futures of AIED Add up on results and conclusions when done

Keywords: ChatGPT, Large Language Models, Higher Education, Human-AI Co-creation ,Personalized Learning, Speculative Scenarios

1 Introduction

Upon every technological breakthrough, the educational domain reconsiders its methods and strategies. During the recent decades, the rise of digital technologies has reshaped the educational landscape. Evidence from the past 40 years on educational technologies impact consistently indicates positive benefits from its integration in programming courses or through innovative approaches (Higgins et al., 2012). However, while digital technologies might spark motivation and engagement in young students, its mere adoption does not ensure favorable outcomes. To fully embrace their potential, a deliberate pedagogical approach is essential, especially when transitioning from

traditional to modern educational approaches (Khaddage et al., 2021; Parker et al., 2020) .

As part of the larger digital transformation, Artificial-Intelligence (AI)-based systems have emerged as important players in the educational domain. These systems have found extensive use in administration, instruction, and learning (Chen et al., 2020), accompanied by a growing body of research around AI in Education (AIED), a distinct sub-field of digital learning research (Niemi et al., 2023). But despite the wide range of potential applications of AI-based systems, they have been mostly implemented to facilitate, or automate, mainstream learning approaches. While useful, this approach sidelines teachers agency, experience, and creativity to integrate such technologies into their unique practices and pedagogical designs. (Holmes et al., 2023)

A dominant trend has been that AIED implementations are predominantly designed by technical teams or departments, such as those focused in Science, Technology, Engineering, and Math (STEM), leading to a notable under-representation of educational or psychological perspectives in AIED research (Holmes & Tuomi, 2022; Zawacki-Richter et al., 2019). Among the probable causes of this we can highlight the lack of technical expertise, among others, which is a big barrier for educational adoption of technologies (Reid, 2014). However, as with other digital technologies, AI systems are evolving to become more accessible by reducing their technical entry barriers, both for casual use and for the design of new applications.

One of such evolution are Large Language Models (LLMs) such as OpenAI’s Generative Pre-Trained (GPT) model. They have significantly evolved into systems capable of performing various natural language tasks (Brown et al., 2020). Interestingly, they have shown emergent features as they are able to perform tasks that were not originally part, or expected, of the design (Wei et al., 2022). For example, GPT-3 could adapt to new tasks using an approach known as *In-context learning*—relying on natural language descriptions of the task—an ability that it’s predecessor struggles to perform consistently (Brown et al., 2020; Wei et al., 2022). This adaptation for different tasks, without requiring exhaustive re-design of their architectures, combined with their potential to serve as building blocks for task-specific AI-tools has led to some researchers to classify them as foundational models (Bommasani et al., 2022; Eloundou et al., 2023). Furthermore, its last iteration (GPT-4) brought a significant increase in performance across every measured metric (OpenAI, 2023a). These capabilities present wide opportunities in both design and application. Given that state-of-the-art models do not require advanced technical programming skills, professionals from different domains might now tailor customized tools that align closely with their contexts, needs, and specialized approaches (Cain, 2023). Building on this accessibility, OpenAI has recently announced the upcoming “GPT builder”, which will enable users to tailor specific web-based GPT applications for private or shared use (OpenAI, 2023b, 2023c). Altogether, a new generation of commercial, public, and private LLMs-based AI applications is on the horizon, with many speculated to be significant in shaping the near to mid-term future (Bommasani et al., 2022; Bubeck et al., 2023).

Due to its versatility and accessibility, OpenAI’s “ChatGPT” has become the fastest-growing application in history since its launch on late 2022 (Milmo, 2023). As

a conversational chatbot, it uses natural language processing to understand and generate human-like text in a dialectical fashion. By providing certain input (prompts) ChatGPT can answer different kinds of questions. Additionally, by using advanced prompting methods users can improve its performance in a wide range of tasks (Fernando et al., 2023; Wei et al., 2023). Potential educational uses involve teaching preparation (generation of course materials, providing suggestions); Assessment (generation of exercises or case scenarios, providing feedback); Learning support (answering questions, summarising information) among others (Lo, 2023; Montenegro-Rueda et al., 2023). However, educational research remains sparse and focuses predominantly in theoretically exploring its potential and limitations (Cain, 2023; Qadir, 2022), and assessing its performance on traditional assessment methods (Nisar & Aslam, 2023). Due to its novelty, a gap in exploratory empirical studies is evident.

We focused our study on exploring the interaction between students and GPT-4 for the creation of personalized course materials in a higher education doctoral course. Our research not only aims to pave the way for innovative designs that enrich and personalize in-person learning experiences using AI tools but also to inspire deeper exploration into the vast potentials of integrating AI in educational settings.

Approach

The study was conducted in an on-campus doctoral course titled “Basics of AI in education”, aimed at exploring historical and contemporary AIED developments. We covered an overview of the historical technical evolution of AIED systems, examination of current popular systems, like performance predictors, AI-tutors and LLMs, a review of the intersection between AI and cognitive sciences, as well as a discussion of emerging ethical concerns and regulatory developments.

We designed the course assignments to explore hypothetical scenarios about AI implementations in education. For this purpose the students were asked to individually design a study, and write a reflective essay about a specific chosen scenario. Consequently, the course topics were broad and suggestive, not focused on any pre-defined scenario, giving the students freedom to choose a scenario of their own interest. Complementing the course contents with a practical experience, we guided the students to interactively generate their hypothetical scenarios with the support of AI.

To facilitate this, we designed a guided interaction to co-create personalized hypothetical scenarios through GPT-4, which involved pre-defined prompting templates and sets of conceptual keywords to construct final scenario generation prompts. We guided our design to address two main objectives.

Firstly, to accommodate students with diverse levels of experience with GPT. Due to its novelty, we had to guide students without experience in interacting with GPT before. By providing clear guidelines and a set of actions to choose from, outlined by prompt templates, we expected that students lacking prior experience could complete their co-creation task without major technical challenges. Secondly, to enable a systematic and reproducible examination of interactions with GPT by establishing a controlled interaction protocol. Not only it standardizes how students interact with GPT but it can be tailored to capture specific elements according to our research interests, particularly the behavioral dimensions of students interaction.

Upon reflecting on the conversational dynamics of ChatGPT in the context of our scenario co-creation task, we opted to guide our prompt design with insights from behavioral models used in Information Foraging Theory. Given that ChatGPT has contextual memory of the conversation, it enables not only to generate scenarios but also to delve deeper by prompting for additional details, thus enriching the scenario with more information. This dynamic closely mirrors two core aspects of Information Foraging Theory: Exploration, where new information landscapes are searched (scenarios), and Exploitation, where agents decide to utilize a landscape to its fullest potential—in this case, further elaborating an interesting scenario. (Cohen et al., 2007; Hills et al., 2015; Todd & Hills, 2020). Therefore we shape our prompt design around these distinct foraging actions, conceptualizing the interactions as an information foraging task.

Furthermore, we deemed worthy to focus our analysis in examining the role of Computational Thinking Skills in modulating students interactions with ChatGPT. In a seminal contribution, Wing (2006) underscored the importance of Computational Thinking skills, which enables people to solve tasks in a similar way that computer algorithms work and are beneficial across most professional domains. Since then, Computational Thinking Skills have grown into a cornerstone of research in digital education and different researchers argue that they correlate with more confident and efficient use of digital technologies (Cansu & Cansu, 2019; Grover & Pea, 2013; Shute et al., 2017). Regarding its influence over Human-AI Interactions, Celik (2023) explored the determinants of AI literacy, which encompass the knowledge for using, recognizing and evaluating AI-based tools, and reported a significant correlation with Computational Thinking Skills. Building on these contributions, we aim to delve deeper into the practical manifestations of such association and investigate whether the unique nature of LLMs, distinct from other AI systems, might reveal new insights into Computational Thinking Skills within the evolving landscape of education.

Therefore, in our study we focused our analysis in exploring the following aspects.

RQ1. How does the students' Computational Thinking Skills influence their information foraging behavior when co-creating scenarios with GPT-4?

Alternatives:

- (i) Through which behaviors do students showcase their Computational Thinking skills during their engagement with GPT-4?
- (ii) How are students' Computational Thinking skills reflected in their interactions with GPT-4?
- (iii) In what ways do students demonstrate Computational Thinking skills when engaging with GPT-4 in a structured setting?

Note, alternatives chosen by a set of alternatives co-created with GPT-4

2 Methods

Guided interaction Design

As mentioned above, our design purpose consists of orienting the students in their interaction and enabling a methodical analysis of the process. It consists of a prompting template, comprising six modifiable prompts, a set of keyword cards, and an interaction protocol. With them, students crafted a diversity of prompts tailored to their interests. Subsequently, we used these crafted prompts with ChatGPT to create and elaborate their hypothetical scenarios.

Prompting templates and Set of Keywords

The prompting templates were divided into two main categories, according to the information foraging theoretical approach described above. Firstly, we created three different Exploration prompts aimed at creating unique scenarios. Secondly, we created three Exploitation prompts to further elaborate already generated scenarios. They consisted in a fixed text written in black representing unchangeable elements, along with variable spaces written in colored texts. These colored spaces corresponded to different categories where students could select and combine various concepts. These concepts were selected from the set of keywords. This distinction between the two types of prompting templates facilitates our analysis by enabling us to clearly describe their use as either exploration or exploitation actions.

To construct the set of keywords, we envisioned six categories that generally described AI-influenced educational scenarios. These categories were: Actors, AI Tools, Educational Process, Environment, Subject, and Activity. Additionally, we created an "Effects" category to elicit both positive and negative scenarios. Afterwards, aiming of examining the students' personal interests in driving the interaction, these categories were populated based on an initial course assignment, in which students were asked to outline their interests in relation to the aforementioned categories. The final result was a diverse set of keywords, reflecting the varying perspectives and interests that students brought to the course.

Both the prompting template and the keywords were later printed and laminated. The keywords were represented as a series of color-coded cards, each card corresponding to an individual keyword and colored by category. The color-coding matched the ones used in the prompting template, to enable a more intuitive understanding of the allowed combinations.

Interaction Protocol

The interactions were realized individually with each student. The main author acted as intermediary between the students and ChatGPT, guiding them over the co-creation process according to the following interaction protocol:

- Prompting templates were displayed in a table, all of the options visible for them. The keywords cards were facing down, distributed almost equally in 5 rows and divided by category.

- I reiterated the aim of the task—creating an interesting scenario for their assignments—and introduced the purpose of the prompting templates and the set of keywords. Additionally, I provided an example of a generation prompt that included example keywords.
- If the aim and rules were clear, students proceeded to uncover the first row of cards and started crafting a prompt by selecting one template and keywords of interest. When completed, I prompted ChatGPT and displayed the generated text. After students finalized reading, I reminded them their different choices, either generate other scenarios (explore), further elaborate the generated scenario, or end the task by choosing the scenario in front of them.
- When the participant found that any scenario was interesting enough for them to choose it and end the task, the full transcript of the conversation for the selected scenario was provided to them.

2.1 Quantitative analysis - Network Analysis

2.2 Qualitative analysis- Content Analysis

3 Results

4 Discussion

Discussions should be brief and focused. In some disciplines use of Discussion or ‘Conclusion’ is interchangeable. It is not mandatory to use both. Some journals prefer a section ‘Results and Discussion’ followed by a section ‘Conclusion’. Please refer to Journal-level guidance for any specific requirements.

4.1 Limitations

4.2 Further Lines of Research

4.3 Conclusion

Conclusions may be used to restate your hypothesis or research question, restate your major findings, explain the relevance and the added value of your work, highlight any limitations of your study, describe future directions for research and recommendations.

In some disciplines use of Discussion or ‘Conclusion’ is interchangeable. It is not mandatory to use both. Please refer to Journal-level guidance for any specific requirements.

References

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2022, July). On the Opportunities and Risks of

- Foundation Models [arXiv:2108.07258 [cs]]. <https://doi.org/10.48550/arXiv.2108.07258>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Nee-lakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodeli, D. (2020, July). Language Models are Few-Shot Learners [arXiv:2005.14165 [cs]]. <https://doi.org/10.48550/arXiv.2005.14165>
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023, March). Sparks of Artificial General Intelligence: Early experiments with GPT-4 [arXiv:2303.12712 [cs]]. Retrieved March 31, 2023, from <http://arxiv.org/abs/2303.12712>
- Cain, W. (2023). GPTeammate: A Design Fiction on the Use of Variants of the GPT Language Model as Cognitive Partners for Active Learning in Higher Education, 1293–1298. Retrieved April 18, 2023, from <https://www.learntechlib.org/primary/p/221996/>
- Cansu, F. K., & Cansu, S. K. (2019). An Overview of Computational Thinking [Number: 1]. *International Journal of Computer Science Education in Schools*, 3(1), 17–30. <https://doi.org/10.21585/ijcses.v3i1.53>
- Celik, I. (2023). Exploring the Determinants of Artificial Intelligence (AI) Literacy: Digital Divide, Computational Thinking, Cognitive Absorption. *Telematics and Informatics*, 83, 102026. <https://doi.org/10.1016/j.tele.2023.102026>
- Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, 1, 100002. <https://doi.org/10.1016/j.caeai.2020.100002>
- Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration [Publisher: Royal Society]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481), 933–942. <https://doi.org/10.1098/rstb.2007.2098>
- Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023, March). GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models [arXiv:2303.10130 [cs, econ, q-fin]]. <https://doi.org/10.48550/arXiv.2303.10130>
- Fernando, C., Banarse, D., Michalewski, H., Osindero, S., & Rocktäschel, T. (2023, September). Promptbreeder: Self-Referential Self-Improvement Via Prompt Evolution [arXiv:2309.16797 [cs]]. <https://doi.org/10.48550/arXiv.2309.16797>
- Grover, S., & Pea, R. (2013). Computational Thinking in K–12: A Review of the State of the Field. *Educational Researcher*, 42(1), 38–43. <https://doi.org/10.3102/0013189X12463051>
- Higgins, S., Xiao, Z., & Katsipataki, M. (2012, November). *The Impact of Digital Technology on Learning: A Summary for the Education Endowment Foundation. Full Report* (tech. rep.) (Publication Title: Education Endowment Foundation ERIC Number: ED612174). Education Endowment Foundation. Retrieved November 1, 2023, from <https://eric.ed.gov/?id=ED612174>

- Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., & Couzin, I. D. (2015). Exploration versus exploitation in space, mind, and society [Publisher: Elsevier]. *Trends in Cognitive Sciences*, 19(1), 46–54. <https://doi.org/10.1016/j.tics.2014.10.004>
- Holmes, W., Bialik, M., & Fadel, C. (2023, January). Artificial intelligence in education. <https://doi.org/10.58863/20.500.12424/4276068>
- Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57(4), 542–570. <https://doi.org/10.1111/ejed.12533>
- Khaddage, F., Norris, C., Soloway, E., & Davidson, A.-L. (2021). Bridging the Gap Between Traditional Pedagogies and Digital Technologies “Challenges and Misalignments”, 80–87. Retrieved November 6, 2023, from <https://www.learntechlib.org/primary/p/219642/>
- Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature [Number: 4 Publisher: Multidisciplinary Digital Publishing Institute]. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>
- Milmo, D. (2023). ChatGPT reaches 100 million users two months after launch. *The Guardian*. Retrieved November 6, 2023, from <https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app>
- Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J. M., & López-Meneses, E. (2023). Impact of the Implementation of ChatGPT in Education: A Systematic Review [Number: 8 Publisher: Multidisciplinary Digital Publishing Institute]. *Computers*, 12(8), 153. <https://doi.org/10.3390/computers12080153>
- Niemi, H., Pea, R. D., & Lu, Y. (Eds.). (2023). *AI in Learning: Designing the Future*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-09687-7>
- Nisar, S., & Aslam, M. S. (2023, January). Is ChatGPT a Good Tool for T&CM Students in Studying Pharmacology? <https://doi.org/10.2139/ssrn.4324310>
- OpenAI. (2023a). Gpt-4 technical report.
- OpenAI. (2023b, November). Introducing GPTs. Retrieved November 9, 2023, from <https://openai.com/blog/introducing-gpts>
- OpenAI. (2023c, November 6). *Openai devday, opening keynote*. <https://youtu.be/U9mJuUkhUzk?t=1552>
- Parker, J., Maor, D., & Herrington, J. (2020). Authentic online learning: Aligning learner needs, pedagogy and technology [Publisher: Institutes for Educational Research in NSW, SA and WA]. *Issues in Educational Research*, 23(2), 227–241. <https://doi.org/10.3316/ielapa.354613892620670>
- Qadir, J. (2022, December). Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education. <https://doi.org/10.36227/techrxiv.21789434.v1>
- Reid, P. (2014). Categories for barriers to adoption of instructional technologies. *Education and Information Technologies*, 19(2), 383–407. <https://doi.org/10.1007/s10639-012-9222-z>

- Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, 22, 142–158. <https://doi.org/10.1016/j.edurev.2017.09.003>
- Todd, P. M., & Hills, T. T. (2020). Foraging in Mind [Publisher: SAGE Publications Inc]. *Current Directions in Psychological Science*, 29(3), 309–315. <https://doi.org/10.1177/0963721420915861>
- Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., & Fedus, W. (2022, October). Emergent Abilities of Large Language Models [arXiv:2206.07682 [cs]]. Retrieved March 8, 2023, from <http://arxiv.org/abs/2206.07682>
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2023, January). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models [arXiv:2201.11903 [cs]]. <https://doi.org/10.48550/arXiv.2201.11903>
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–35. <https://doi.org/10.1145/1118178.1118215>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 39. <https://doi.org/10.1186/s41239-019-0171-0>