Beyond the Hype: Towards a Critical Debate About AI Chatbots in Swedish Higher Education

Teresa Cerratto Pargman,* Elin Sporrong, Alexandra Farazouli, & Cormac McGrath
Stockholm University, Sweden

Interested in emerging technologies in higher education, we look at AI chatbots through the lens of human–technology mediations. We argue for shifting the focus from *what* higher education can do with AI chatbots to *why* AI chatbots are compelling for higher education's raison-d'être. We call for a critical debate examining the power of AI chatbots in configuring students as civic actors in an increasingly complex and digitalized society.

We welcome a continuous and rigorous examination of generative AI chatbots and their impact on teaching practices and student learning in higher education.

Keywords: ChatGPT, technological mediations, criticality, student writing, higher education practices

INTRODUCTION

Since the launch of Open AI's ChatGPT, large language models (LLMs) have attracted significant academic and press media attention, focusing on how spectacularly such AI chatbots perform and emulate human-like conversation. Such a focus has led to inflated hype (Leaver & Srdarov, 2023) and discourses considering how AI chatbots like ChatGPT act as a game-changer for higher education and academia (Stokel-Walker, 2022) with the potential to "revolutionize the way education is provided and support teachers in giving students more tailored and efficient learning experiences" (Pradana et al., 2023, p. 9). At the same time, critical concerns have emerged regarding AI chatbots. These include, for example, issues about (i) how higher education must respond to AI chatbots and whether they present an "inevitable change to which all must respond," including universities, to ensure their survival (Bearman et al., 2022), (ii) whether AI chatbots will replace teachers (Pradana et al., 2023), (iii) how AI might impact and alter teachers' authority across staff, machines, corporations, and students (Bearman et al., 2022, p. 369), and (iv) the legality of keeping teachers out of the loop (Colonna, 2023).

While these questions are compelling, they often overlook that digital technologies, like any tool, shape and mediate relationships constituting human experience and existence (Verbeek, 2005; 2011). As such, digital technologies constitute and are constitutive of emerging practices (Cerratto Pargman, 2020). In the education sector, LLMs-based AI chatbots mediate, for instance, student–teacher relationships and how students and teachers perceive the world they are part of and how they exist in it. Thus, while AI chatbots are tools that may save time, accelerate performance, and increase efficiency in academia, they also form social and material conditions that actively configure and shape students as citizens in an increasingly complex and digitalized society.

Building on our ongoing work on AI in Swedish higher education (Cerratto Pargman & McGrath, 2021; Farazouli et al., 2023; McGrath et al., 2023, 2024), we ask how AI chatbots come to make up new practices and values in higher education, how education practice shapes them, and how we, actors in the higher education sector, become in our entanglement with them (Lindgren, 2023a; Verbeek, 2005; 2011).

In light of the pressing discussions on AI chatbots in the Swedish higher education sector, we see the need to add criticality to discourses on how AI chatbots may operate in an academic space and transform core practices, such as assessment. In this sense, we seek to stimulate debate about using AI chatbots in Swedish higher education by presenting selected insights and questions for higher education institutions and teachers to reflect on.

SIX INSIGHTS ON AI CHATBOTS FOR DISCUSSION IN HIGHER EDUCATION

The following section presents six insights we have gained from our readings on AI chatbots and the questions we wish to raise to contribute to a critical discussion on AI chatbot mediation in higher education.

AI chatbots are computationally trained on what language looks like, not what it means

Focusing on the materiality of the language models on which AI chatbots are built on, we note that "LMs [language models] are trained only on linguistic form and don't have access to meaning (Bender & Koller, 2020, p. 5185). More specifically, they are "trained on string prediction tasks: that is, predicting the likelihood of a token (character, word or string) given either its preceding context or (in bidirectional and masked language models) its surrounding context" (Bender et al., 2021, p. 611). When deployed, such systems are unsupervised and take a text as input, commonly outputting scores or string predictions (Bender et al., 2021). In this sense, the generated outputs are not based on a model capable of understanding words, characters, or meaning. Instead, the model operates statistically, independently of the semantics involved, at the level of the linguistic form of the human language. In other words, it focuses on what the words look like and the probability of them occurring in relation to each other instead of what they mean for someone.

Moreover, while AI chatbots generate text, they are not additive, creating original content; instead, they re-produce existing text based on the training data (Saunders, 2023).

In this context, some questions of relevance relate to: If AI chatbots can generate syntactically coherent human-like text, what are the implications for how teachers engage with teaching and student assessment practices? What are the consequences for student learning processes?

Such questions require higher education institutions to discuss the difference between "linguistic form" and "meaning" (Bender & Koller, 2020) openly and to put effort into designing program activities that cultivate students' criticality and ability to discern nonsense text from knowledge in the AI era.

AI chatbots compute responses based on text available on the Internet

Following the material constitution of AI chatbots and how it may impact student construction of meaning, we ask where the data used to feed these large language models comes from and whether they are reliable. AI chatbot developers use a corpus of text data from the Internet to train their language model (Bender et al., 2021). While OpenAI, for example, states the company uses information that is publicly available on the Internet, information that they license from third parties, and information that their users or human trainers provide (OpenAI, 2024),

the company has not disclosed the exact sources of this data for giving their model a broad understanding of language. As generated responses to users' prompts are based on the corpus available to the large language model (Bender et al., 2021), the bias already existing in the corpus used to train the model will result in bias in responses given by AI chatbots. On this note, Noble (2018), who studied the problem of relying on the information accessed on the Internet by search engines like Google, has a point when she argues that "information organization is a matter of sociopolitical and historical processes that serve particular interests" (p. 138) and that "search does not merely present pages, but structures knowledge, and the results retrieved in a commercial engine creates their particular material reality" (p. 148). Considering this observation, we ask how students' knowledge will be configured by interacting with AI chatbots like ChatGPT.

Relevant questions for higher education institutions and university teachers include: How deeply are students engaging with the material that matters when they use AI chatbots? Will the potential bias in the text-generated responses become an incentive to boost the development of academic argumentative skills or result in reproducing biases? How can higher education institutions inform students on how to use tools that are opaque with their sources when source criticism is so quintessential to the higher education experience?

AI chatbots are developed for purposes other than education

AI chatbots are not only an example of a recent technological development or the result of the operation of computational techniques and machine learning. Broadly, they are general purpose tools not developed for use in education contexts specifically. They are also products designed and developed by people, that is, groups of developers, designers, programmers, data, and computer scientists who are led by private companies' commercial interests which may have great ambitions for technological and scientific progress but a limited understanding of the implications of their developments for critical sectors in society such as higher education. On this note, Gilliard and Rorabaugh (2023) observe that developers of chatbots may perhaps not primarily be "concerned with educators very much. It has come, disrupted, and walked away with no afterthought about what schools should do with the program. Now is a time for educators to understand how, if at all, such innovations fit in the education sector" (Gilliard & Rorabaugh, 2023, p. 1).

These observations echo Rahm's (2019) critical piece on the need to change teaching practices every time a new digital technology emerges in society. Accordingly, the education sector is allocated the task of providing citizens' development of competencies, in this case, AI literacy, to indirectly address, among others, the needs and ambitions of international companies to make their artifacts available. In this context, what task should university programs assume every time a novel technology becomes available on the market?

As discussed earlier in our work (Farazouli et al., 2023), AI chatbots have a potentially unsettling impact on how teachers view students' work, throwing into sharp relief the potential breakdown in trust between university teachers and students. On this note, universities must engage in dialogue with students and colleagues about what kind of authentic problems AI chatbots solve in their teaching context. Bergviken-Rensfeldt and Rahm (2023) "explain that while AI may promise to replace 'routine and tedious' educational work with automation, research demonstrates that automation may paradoxically lead to the opposite result: an increased workload for humans as well as a restructuring of how the work is performed" (p. 117). Here, we need to consider important questions, such as: What problems do AI chatbots solve in the context

of higher education practice? Why are AI chatbots viewed as the "next thing" that universities must embrace? What would happen if universities critically researched these technologies before adopting and introducing them into everyday practice?

AI chatbots redefine current understandings of student writing competence
The phenomena of copy-pasting a response generated by an AI chatbot may be problematic not only from an academic integrity point of view but also concerning students' development and mobilization of various literacies and competencies. What do students learn when they ask AI chatbots to answer questions given in assignments? Especially when students may, at times, be more concerned about passing exams than meeting aimed learning objectives, there is a risk that chatbots will be used for performing through getting good grades rather than developing literacies, abilities, and competence (Byrnes, 2023). In this regard, it is not farfetched to ask if student-AI chatbot mediation will increase the gap between the students already writing critically and feeling self-confident in writing their responses and those struggling with writing and falling behind. While one group may be able to distinguish between passing an assignment and learning from an assignment, there is a risk that others might not be able to consider such a distinction. Amid performance pressures, AI chatbots may seem attractive despite the risk of lost literacy development opportunities. On this note, Lindgren (2023b) observes that while technology development has historically developed to liberate us from physical work, large language models and associated AI chatbots propose to "liberate" us from the creative tasks demanding idea creation, planning, argumentation, and other critical literary activities.

We ask educators how AI chatbots might change our understanding of student writing competence standards. What do we gain in higher education with novel interpretations of literacy? What do we lose? What kind of tasks will AI chatbots do now and in the future, and what are the implications for human creativity and knowledge? Will engaging with text generated by chatbots generate a depth and passion for learning?

AI chatbots produce disembodied text without a point of view
While spectacular, AI chatbot's responses come from standard data science practices that most often attempt to "see everything from nowhere," giving the impression of providing universal truths that are pure and objective (Haraway, 1988, p. 581). In this sense, understanding the voices and positionality involved in especially contested societal topics for students becomes difficult as responses provided by the chatbot are not grounded in human experience and existence. As such, the response cannot belong to a specific geographical place or specific felt time or enact a subjective "partial perspective" (Haraway, 1988) because computational linguistic techniques know nothing about belonging, feelings, or experience. As neither the content nor the structure is designed by a particular someone, the question is: What competencies do students need to develop to be able to argue with AI chatbots and develop critical literacy?

If accessing information today is about forming and developing one's own critical point of view to act on the complex dilemmas affecting our societies, how are AI chatbots transforming students' strategies to think critically and construct meaning? (Sporrong & Tikkanen, 2016).

Here, we identify several core questions, including: Who answers to whom in the text generated by AI chatbots? What does it entail to read information from unknown sources? How can students distinguish critical voices and identify where discourses on pressing social issues have emerged? How could the interests, ideologies, and values behind AI chatbots' answers be deconstructed as part of a student's learning journey?

AI chatbots may be developed through unethical data science practices
Practices in higher education are fundamentally relational and ethical concerning the very nature of engagement between teachers and students. Teachers should introduce subjects truthfully, treat students respectfully, and examine without prejudice (SULF, 2019).

Previous work suggests that university teachers want to engage ethically with AI (McGrath et al., 2023) and that learning technologies, some may include AI, could be deployed with equity ambitions in mind and be used by groups of students in the margins of the higher education system (Cerratto Pargman et al., 2023). However, we note that AI tools and systems may have been developed unethically (Bender et al., 2021) in terms of the resources they use (Luccioni et al., 2023), the data they are sourced on (Bender et al., 2021), and their apparent contempt for intellectual property (Grynbaum & Mac, 2023), and this gives cause for concern.

AI chatbots come with a massive environmental cost (Li et al., 2023). In particular, Strubell et al. (2019) show that neural network models for natural language processing "are costly to train and develop, both financially, due to the cost of hardware and electricity or cloud compute time, and environmentally, due to the carbon footprint required to fuel modern tensor processing hardware" (p. 1). Based on the Strubell's et al., (2019), evidence, Bender et al., (2021) emphasize the "environmental cost" of such models by arguing that "increasing the environmental and financial costs of these models doubly punishes marginalized communities that are least likely to benefit from the progress achieved by large language models and most likely to be harmed by negative environmental consequences of its resource consumption" (p. 610).

Furthermore, part of the training and fine-tuning involved in developing AI chatbots requires human intervention in the form of reinforcement learning from human feedback, where humans assess the degree of accuracy or the desired fairness in AI chatbot output. The model adjusts its output to human responses and learns over time (Frey & Osborne, 2023). This involves engaging humans to shift through the material to remove violent, sexist, and racist material.

As mentioned, AI chatbots do not inform us where they source their data. This has led to several ongoing cases of litigation where publishing houses, creative professionals, and authors are suing OpenAI and other companies involved in developing large language model-based AI chatbots over the theft of intellectual material (Creamer, 2023).

Against this background, we ask the reader to consider: What does it mean to learn at such a high cost to the environment? What does it mean to use tools that were developed where humans must siphon through material that is violent, racist, and discriminatory? And how can university teachers, in good faith, ask students to engage with material that has been potentially unlawfully and unethically sourced?

CONCLUSION
Interested in emerging technologies in higher education, we argue for shifting the focus from *what* higher education can do with AI chatbots to *why* AI chatbots are compelling for higher education's raison-d'être. Based on key readings on AI chatbots, we contribute six insights to add criticality to the current unnuanced debate on AI chatbots in higher education.

We call for a debate examining the power of AI chatbots in configuring students as civic actors in an increasingly complex and digitalized society.

We welcome a continuous and rigorous examination of generative AI chatbots and their impact on teaching practices and student learning in higher education.

AUTHOR BIOGRAPHIES

Teresa Cerratto Pargman is Professor of human-computer interaction at Stockholm University. Currently she leads the project Ethical and Legal Challenges in Relationship to AI-driven Practices in Higher Education funded by the WASP-HS (Wallenberg Foundations). Teresa is also an associate director for outreach at Digital Futures in Sweden and an associated researcher at the Weizenbaum Institute in Berlin, Germany.

Elin Sporrong is a Doctoral Student at the Department of Computer and Systems Sciences at Stockholm university. Her work, that focuses on ethical aspects of anticipations of AI in higher education, is conducted within the project Ethical and Legal Challenges in Relationship to AI-driven Practices in Higher Education funded by the WASP-HS (Wallenberg Foundations).

Alexandra Farazouli is a Doctoral Student at the Department of Education at Stockholm University, studying emerging AI-driven technologies and their ethical implications for educational practices in higher education, within the project Ethical and Legal Challenges in Relationship to AI-driven Practices in Higher Education funded by the WASP-HS (Wallenberg Foundations).

Cormac McGrath is Associate professor of Education at the Department of Education, Stockholm University, and is co-PI for two projects addressing the impacts of AI in education funded by the WASP-HS (Wallenberg Foundations).

REFERENCES

Bearman, M., Ryan, J., & Ajjawi, R. (2022). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education*, 1–17.

Bender, E., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *FAccT '21: Proceedings of the 2021 ACM conerence on fairness, accountability and transparency* (pp. 610–623). ACM. https://doi.org/10.1145/3442188.3445922

Bender, E. M., & Koller, A. (2020, July). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 5185–5198). Association for Computational Linguistics.

Bergviken Rensfeldt, A., & Rahm, L. (2023). Automating teacher work? A history of the politics of automation and artificial intelligence in education. *Postdigital Science and Education*, *5*, 25–43.

Byrnes, G. (2023). *Learning to live with ChatGPT*. University World News. Retrieved January 9, 2024, from https://www.universityworldnews.com/post.php?story=20230222112438466

Cerratto-Pargman, T. (2020). Practice as a concept in educational technology. In M. A. Peters & R. Heraud (Eds.), *Encyclopedia of educational innovation* (pp. 1–5). Springer.

Cerratto Pargman, T., Lindberg, Y., & Buch, A. (2023). Automation is coming! Exploring future(s)-oriented methods in education. *Postdigital Science and Education*, *5*, 171–194. https://doi.org/10.1007/s42438-022-00349-6

Cerratto Pargman, T., McGrath, C., Viberg, O., & Knight, S. (2023). New vistas on responsible learning analytics: A data feminist perspective. *Journal of Learning Analytics*, *10*(1), 133–148.

Colonna, L. (2023). Teachers in the loop? An analysis of automatic assessment systems under Article 22 GDPR. *International Data Privacy Law*, *14*(1), 3–18. https://doi.org/10.1093/idpl/ipad024

Creamer, E. (2023, 5 July). Authors file a lawsuit against OpenAI for unlawfully 'ingesting' their books. *The Guardian*. https://www.theguardian.com/books/2023/jul/05/authors-file-a-lawsuit-against-openai-for-unlawfully-ingesting-their-books

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2023). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education*, 1–13. https://doi.org/10.1080/02602938. 2023.2241676

Frey, C. B., & Osborne, M. (2023). Generative AI and the future of work: A reappraisal. *Brown Journal of World Affairs*, 1–12.

Gilliard, C., & Rorabaugh, P. (2023, February 3). *You're not going to like how colleges respond to ChatGPT*. Slate. https://slate.com/technology/2023/02/chat-gpt-cheating-college-ai-detection.html

Grynbaum, M. M. & Mac, R. (2023, 27 December). The Times sues OpenAI and Microsoft over A.I. use of copyrighted work. *The New York Times*. https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, *14*(3), 575–599. https://doi.org/10.2307/3178066

Leaver, T., & Srdarov, S., (2023). ChatGPT isn't magic: The hype and hypocrisy of generative artificial intelligence (AI) rhetoric. *M/C Journal*, *26*(5). https://doi.org/10.5204/mcj.3004

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. https://doi.org/10.48550/arXiv.2304.03271

Lindgren, S. (2023a). *Handbook of critical studies of artificial intelligence*. Edward Elgar Publishing.

Lindgren, H. (2023b, January 28). Vad gör de kreativa AI-tjänsterna med oss? [What are the creative AI-services doing to us?] *Svenska Dagbladet*. https://www.svd.se/a/dwL37X/vad-gor-de-kreativa-ai-tjansterna-med-oss

Luccioni, A. S., Jernite, Y., & Strubell, E. (2023). *Power hungry processing: Watts driving the cost of AI deployment?* arXiv. https://doi.org/10.48550/arXiv.2311.16863

McGrath, C., Cerratto Pargman, T., Juth, N., & Palmgren, P. J. (2023). University teachers' perceptions of responsibility and artificial intelligence in higher education – An experimental philosophical study. *Computers and Education: Artificial Intelligence*, *4*, Article 100139.

McGrath, C., Farazouli, A., Cerratto-Pargman, T. (2024). *AI chatbots in higher education. A state-of-the-art review of an emerging research area*. Reserach Square. https://doi.org/10.48550/arXiv.2311.16863

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press. https://doi.org/10.2307/j.ctt1pwt9w5

Open AI. (2024). *How ChatGPT and our language models are developed*. Retrieved 9 January 2024 from https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed

Pradana, M., Elisa, H. P. & Syarifuddin, S. (2023). Discussing ChatGPT in education: A literature review and bibliometric analysis. *Cogent Education*, *10*(2), Article 2243134. https://doi.org/10.1080/2331 186X.2023.2243134

Strubell, E., Ganesh, A., & McCallum, A. (2019). *Energy and policy considerations for deep learning in NLP* arXiv. https://doi.org/10.48550/arXiv.1906.02243

SULF. (n.d.). *Ethical Guidelines for university teachers*. Retrieved 15 January 2023 from https://sulf.se/en/work-salary-and-benefits/ethics/

Sporrong, E., & Westin Tikkanen, K. (2016). *Kritiskt tänkande – i teori och praktik* [Critical thinking – In theory and practice]. Studentlitteratur.

Rahm, L. (2019). *Educational imaginaries: A genealogy of the digital citizen* Retrieved 15 January 2023 from. https://doi.org/10.3384/diss.diva-154017

Saunders, S. (2023). *Rather than ban generative AI, universities must learn from the past*. University World News. https://www.universityworldnews.com/post.php?story=20230221075136410 Accessed January 13 2023.

Stokel-Walker, C. (2022). AI bot ChatGPT writes smart essays-should academics worry? *Nature*. https://doi.org/10.1038/d41586-022-04397-7

Verbeek, P.-P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. Pennsylvania State University Press.

Verbeek, P.-P. (2011). *Moralizing technology: Understanding and designing the morality of things*. The University of Chicago Press.

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). *Emergent abilities of large language models*. https://openreview.net/forum?id=yzkSU5zdwD