GENERATIVE AI-CHATBOTS IN HIGHER EDUCATION: CHALLENGES AND OPPORTUNITIES IN STUDENT MOTIVATION AND AUTHENTIC ASSESSMENTS

Patricia Brockmann, & Heidi Schuhbauer
Computer Science Department, Nuremberg Institute of Technology (Germany)

Abstract

The rapid pace of development in generative artificial intelligence (AI) has resulted in the introduction of extremely advanced chatbots, such as ChatGPT, Google Bard and Copilot. These chatbots are capable of simulating human behavior quite successfully, because they have been trained on huge data sets with large language models (LLM). New avenues for application of generative AI-chatbots have arisen not only in industry, science and medicine, but also in education. The consideration of social and ethical consequences of this new technology has not kept pace with the speed of its development. This paper explores challenges and opportunities which have arisen in higher education with the emergence of generative AI-chatbots. For university students, the temptation to use chatbots to complete their assignments may seem like a good way to save time and effort. The disadvantage is that they may not learn anything by simply copying answers from a chatbot. Inexperienced students may trust the validity of answers which are incorrect, especially if they do not have prerequisite knowledge to evaluate the output of a chatbot. For educators, however, it can be extremely difficult to determine whether students have completed their assignments themselves, or whether they have submitted the output from a chatbot. Instructors have the responsibility to teach university students to avoid plagiarism by judiciously citing the sources used. Furthermore, competencies taught at the university level should go beyond mere reproduction of facts. The development of analytical capabilities and critical thinking often require hard work to learn from mistakes, so-called "productive failure". Creative teaching methods, such as gamification, may help to motivate students to engage with learning materials. A number of questions which have arisen from these challenges will be addressed in this paper. (1) Can educators determine whether student submissions are original or were written by a chatbot?; (2) Should students be allowed to use generative AI-chatbots to work on assessments? If so, how?; (3) Can creative teaching methods, such as gamification, motivate students to engage in learning?; (4) What kinds of assessments can help to evaluate whether students have achieved learning goals? Challenges encountered in exploring these questions when teaching undergraduate university courses will be presented. Opportunities to increase student engagement and design of authentic assessments will be discussed.

Keywords: Artificial intelligence, chatbots, motivation, gamification, assessment.

1. Introduction

Chatbots have come a long way since the original "chatter-bots", which were only able to answer the simplest of user questions (Mauldin, 1994). Because artificial intelligence itself was not yet as advanced as today, these early chatbots had to be explicitly pre-programmed to answer a limited number of pre-defined questions. Although they could only respond with simple, predetermined answers, they were considered useful to answer typical, rote questions, such as in customer service.

Recently, a quantum leap in chatbot technology was achieved by advances in deep learning methods, such as generative large-language models (LLM) (Floridi & Chiriatti, 2020). This new generation of chatbots, such as Chat-GPT (OpenAI, 2023) and Gemini (Google, 2023), can mimic human intelligence so convincingly, that it can be quite difficult to differentiate whether output has been produced by the chatbot or by a human. Generative AI chatbots have achieved a level of sophistication which enables their application in industry, science and medicine. Critical examination of social and ethical consequences of this new technology has not kept pace with the speed of its development.

The use and misuse of chatbots gives rise to a number of challenges in higher education. Answers to typical homework assignments, essays and even computer code can be generated by a

generative AI chatbot with comparatively little effort. The temptation to misuse chatbots to copy answers and thus avoid the effort necessary to actually learn a subject can be hard to resist. Especially inexperienced students may be more likely to trust incorrect answers output by a chatbot, because they lack the prerequisite knowledge to judge their validity. Educators often have difficulty determining whether assignments have been completed by students or have simply been copied from a chatbot.

An initial investigation of research questions is conducted, primarily using literature sources.

R1: Can educators determine whether student submissions are original or were written by a chatbot?

R2: Should students be allowed to use generative AI-chatbots to work on assessments? If so, how?

R3: Can creative teaching methods, such as gamification, motivate students to engage in learning?

R4: What kinds of assessments can help to evaluate whether students have achieved learning goals?

First, generative AI chatbots are described. Next, challenges experienced with AI chatbots in education are discussed. Possible solutions to these challenges, such as gamification and authentic assessments are considered. Finally, conclusions and plans for future work are presented.

2. Generative AI chatbots

Large language models (LLM) implemented as artificial neural networks possess advanced capabilities to generate natural language. Deep learning methods employ statistical pattern recognition methods to train these models on unstructured, unlabeled datasets, such as Wikipedia or the huge amount of text and books available on the internet (Floridi & Chiriatti, 2020). Generative artificial intelligence can go further than simply mimicking human-like behavior. Intelligent searches, summarization of text, generation of essays and computer code can be achieved with simple prompts (Tamkin et al., 2021). ChatGPT, first released as a free version in November of 2022, is the most widely known generative AI chatbot (OpenAI, 2023). Recently, Microsoft announced a major investment and cooperation with OpenAI. The intention is to integrate this technology into its Office software (Microsoft, 2023).

The increased use of generative AI chatbots in all aspects of life gives rise to a diverse range of opportunities for industry, science and society in general. Pividori and Green (2023) argued that large language models can improve productivity in writing and revising scientific papers. They demonstrated that a large language model was capable of suggesting appropriate revisions to academic papers. In their view, generative AI has the potential to revolutionize productivity in areas of knowledge work.

A major weakness in the validity of the results delivered by generative AI chatbots was discussed by Stokel-Walker and Van Noorden (2023). Because large language models are trained on content currently available on the internet, this also includes false, outdated and inherently biased information. Errors or misleading information are most often generated on topics with a low amount of training data. The editors of a major scientific journal stated that ChatGPT "cannot be trusted to get facts right or produce reliable references". For this reason, a number of scientific journals have forbidden the use of any text produced by generative AI (Thorpe, 2023).

The consideration of ethical aspects has failed to keep pace with the rapid development of generative AI. According to Brynjolfsson (2022), this is due to the sole focus on mimicking human intelligence to automate processes. Instead, technology should be used to find new ways to help humans enhance their productivity, to improve the well-being of employees and to create new business models. Farina et al. (2024) proposes two interventions to combat the negative effects of generative AI. To minimize factual errors, they suggest that LLMs should be supervised by humans, especially in sensitive areas, such as medicine, law and journalism. Second, they stress the importance of ensuring transparency and pluralism. Instead of relying on the "black box" of LLMs trained stochastically on an average spectrum of answers, cultural, political and value pluralism need to be maintained.

3. Challenges to education

3.1. Empirical experiences in education

The impact of generative AI chatbots on education was examined soon after the release of ChatGPT by Zhai (2022). He demonstrated the ease with which an apparently coherent-sounding, research paper could be generated within 2-3 hours, without any prior professional knowledge. The only problem was that the information used was not completely accurate. He concludes that while the use of generative AI can greatly increase the speed in conducting subject domain tasks, instructors need to adjust their learning goals for students. Creativity and critical thinking skills should be the focus of learning. Advantages and pitfalls when using ChatGPT in education was reported by Qadir (2023). He successfully used chatbots as intelligent virtual tutors to provide students with personalized, adaptive learning. One problem encountered was the lack of reliability of results and a tendency to output "hallucinatory"

misinformation. The differentiation between acceptable vs. non-acceptable use of chatbots in completing assignments was difficult to define.

The effects of AI chatbots were investigated for an undergraduate course in databases (Brockmann, 2023). Before the semester, the instructor entered the semester assignments as prompts into ChatGPT Version 3. This output was graded using the same grading scale as for students. ChatGPT achieved a grade almost an entire point higher than the students (B- vs. C-). At the end of the semester, students were assigned to test ChatGPT themselves by entering their assignments as prompts. They first expressed frustration that they had "wasted their time" solving the assignments themselves. During a detailed class discussion, however, they noticed that many of the chatbot answers which sounded correct were actually erroneous. Students realized that without first solving the assignments themselves, they would not have had the necessary knowledge to critically evaluate output to identify wrong answers.

This subsection shows that AI chatbots can aid in learning, if they are used carefully (R2: mixed).

3.2. Plagiarism

Plagiarism poses one of the major challenges when using chatbots in higher education. A study which tested 20 essays created by ChatGPT was analyzed using a popular plagiarism-checking software. They found that 16 of the 20 essays were judged to exhibit a high level of originality and less than 10% were flagged as possible plagiarism (Wadhwa & Joshi 2024). An opinion published by Liang et al. (2023) cautions that GPT detectors frequently misclassify non-native English writing as AI-generated. They raise serious ethical concerns that the inherent bias in training data can lead to the marginalization of non-native speakers of English. King (2023) prompted ChatGPT itself to write an editorial about the danger of students abusing ChatGPT to cheat on assignments. ChatGPT judged this behavior as unethical. It recommended that instructors should employ a wider variety of assessment methods, not just traditional essays. Oral presentations, group projects and other hands-on activities which give students the opportunity to demonstrate their skills may be more interactive and engaging and may reduce cheating.

This subsection shows that it is not possible to determine whether students have cheated (R1: negative).

4. Possible solutions

4.1. Gamification

One possible teaching method to increase motivation of students to learn a subject is gamification. The term "gamification" is defined by Deterding et al. (2011) as the "use of game mechanics in non-gaming contexts". Positive, intrinsically motivating psychological reward systems, similar to those experienced when playing games, can help to strengthen engagement and encourage behavioral patterns (Huotari & Hamari, 2012). Gamification has been successfully implemented in a variety of different contexts, such as commerce, health, intra-organizational systems, sustainable consumption, human resources, innovation, data gathering and education (Hamari et al., 2014).

Caponetto et al. (2014) conducted an extensive review on gamification in education. They found that gamification makes learning more attractive, captivating and thus more effective. Gamification can encourage behaviors such as collaboration, creativity and self-guided study. An empirical study conducted Chapman & Rich (2018) found that a majority of their students reported increased motivation when taking part in a gamified course. The most motivating game elements (as rated by students) were the opportunity to earn points for assignments, due date bonuses and penalties, due date flexibility, a current grade indicator, a course map and a leaderboard. A systematic literature review by Popp & Schuhbauer (2023) conducted a meta-analysis on the effectiveness of gaming elements on e-learning platforms. They investigated which gamification elements should be used to increase learning motivation in asynchronous e-learning. Badges, leaderboards, level up systems and progress bars showed the most promise. A further increase in users' motivation could be achieved by combining badges with leaderboards or level up systems.

In contrast to these mostly positive results, negative effects associated with gamification in education were identified in a systematic mapping study conducted by Toda et al. (2018). Four negative effects associated with gamification were identified: indifference, loss of performance, undesired behavior and declining effects. Loss of performance, where gamification hinders the students' learning process, was the most commonly reported issue. Students who were most active in the gamified activity focused more on the game itself and scored lower than their peers on the transfer skills test. Undesired behavior was the second most common negative aspect found. This was often caused by demotivation due to excessive competition. Leaderboards were strongly associated with negative effects, such as loss of performance and undesired behavior. They postulate that this finding correlates with psychological

literature showing the negative effects of ranking systems within learning environments. When designing a gamification component to increase student motivation, possible negative effects should be considered.

This subsection demonstrates that gamification can be used to motivate students (R3: positive).

4.2. Authentic assessments

Instead of adapting a policing approach to generative AI chatbots, Rudolph et al. (2023) advocates a student-centric approach to pedagogy and assessments to support learning. They postulate that digital literacy in education should include AI tools. Memorization of information does not teach students the type of problem-solving skills required in our modern society. The role of faculty is to engage and motivate students to learn. Assessments should encourage students' creative and critical thinking abilities. Especially during the pandemic, many educators turned to online exams to reduce infection risk. Since then, the increasing use of AI chatbots threatens the legitimacy of online exams, due to the potential for cheating. A reevaluation of testing methods is necessary in the era of AI chatbots. The need for authentic assessments, which concentrate on higher cognitive skills, problem-solving and creative thinking coincides with evolving pedagogical learning goals of the 21st century (Ifelebuegu, 2024).

Wiggins (1990) was one of the first authors to make the case for authentic assessments. In contrast to traditional tests, students should not simply recall acquired knowledge. Instead, higher competencies on worthy intellectual tasks should be demonstrated: collaborating with others, writing, revising and discussing papers and even conducting research. Murphy et al. (2017) conducted a literature review on authentic assessments in higher education. They found that authentic assessments can encourage active student learning, improved achievement and greater retention. Students benefit from real-world experiences in safe, supportive environments. Although there was some resistance when working in large groups with time and resource constraints, careful planning and inclusive student consultation helped.

The subsection shows that authentic assessments can evaluate achievement of learning outcomes (R4: positive).

5. Conclusions and future work

This initial exploration of the research literature delivers first insights to the research questions.

R1: Can educators determine whether student submissions are original or were written by a chatbot?
Negative. It is extremely difficult to determine whether students have completed their own work.

R2: Should students be allowed to use generative AI-chatbots to work on assessments? If so, how?
Mixed. Activities which require critical thinking and adequate citation can aid learning.

R3: Can creative teaching methods, such as gamification, motivate students to engage in learning?
Positive. Gamification can motivate students to learn, if the activity is designed carefully.

R4: What kinds of assessments can help to evaluate whether students have achieved learning goals?
Positive. Authentic assessments can aid in evaluating achievement of higher order learning goals, especially real-world problem solving, creativity, critical thinking and collaboration.

Future work will include the development of a learning game for first semester students. As a capstone project, final year students participate in each phase of a research project to develop a learning game about a startup company. This research project starts with requirements analysis, continues through the development and testing phases to determine which elements of gamification can increase student motivation. As an authentic assessment, the goal of this project is to write and publish a conference paper.

References

Brockmann, P. (2023). Challenges and Opportunities Presented by Generative AI-Chatbots: Effects on an Undergraduate Course on Databases. *Proceedings of the International Conference on Information Communication Technologies in Education.*

Brynjolfsson, E. (2022). The Turing trap: the promise & peril or human-like Artificial Intelligence. *Daedelus, 151*(2), 272-287.

Caponetto, I., Earp, J., & Ott, M. (2014). Gamification and education: A literature review. *European Conference on Games Based Learning*, *1,* 50-57.

Chapman, J. R., & Rich, P. J., (2018). Does educational gamification improve students' motivation? If so, which game elements work best? *Journal of Education for Business, 93*(7), 314-321.

Deterding, S., Dixon, D., Rhaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining "gamification". *Proceedings of the 15ᵗʰ international academic MindTrek conference: Envisioning future media environments,* 9-15.

Farina, M., Yu, X., & Lavazza, A. (2024). Ethical considerations and policy interventions concerning the impact of generative AI tools in the economy and in society. *AI and Ethics.* https://doi.org/10.1007/s43681-023-00405-2

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope limits and consequences. *Minds and Machines, 30*, 681-694.

Google (2023). *Welcome to the Gemini era*. https://deepmind.google/technologies/gemini/#introduction

Guerra-Pujol, F. E. (2023). How I learned to Love GPT-3. *UCF Faculty Focus, 22*(1), 25-26.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? A literature review of empirical studies on gamification. *47ᵗʰ Hawaii International Conference on System Sciences,* 3024-3034.

Huotari, K., & Hamari, J. (2012). Defining gamification: a service marketing perspective. *Proceedings of the 16ᵗʰ International Academic MindTrek Conference*, 17- 22.

Ifelebuegu, A. (2023). Rethinking online assessment strategies: Authenticity versus AI chatbot intervention. *Journal of Applied Learning and Teaching, 6*(2), 385-392.

King, M. R., & ChatGPT. (2023). A conversation on artificial intelligence, chatbots, and plagiarism in higher education. *Cellular and Molecular Bioengineering*, 16(1), 1-2.

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns, CellPress, 4*(7). https://doi.org/10.1016/j.patter.2023.100779

Mauldin, M. L. (1994, August). Chatterbots, Tinymuds, and the Turing Test: Entering the Loebner prize competition. *AAAI-94 Proceedings* (Vol. 94, pp. 16-21).

Microsoft (2023, January 23). *Microsoft and OpenAI extend partnership*. https://blogs.microsoft.com/blog/2023/01/23/microsoftandopenaiextendpartnership/

Murphy, V., Fox, J., Freeman, S., & Hughes, N. (2017). "Keeping it Real": A review of the benefits, challenges and steps towards implementing authentic assessment. *All Ireland Journal of Higher Education, 9*(3).

OpenAI (2023). *Introducing ChatGPT*. https://openai.com/blog/chatgpt.

Pividori, M., & Greene, C. S. (2023). A publishing infrastructure for AI-assisted academic authoring. *bioRxiv: The preprint server for biology*. doi: 10.1101/2023.01.21.525030.

Popp, T. H. & Schuhbauer, S. (2023). Systematic Literature Review of the Effort of Gaming Elements on E-Learning Platforms. *Proceedings of the 20ᵗʰ International Conference on Cognition and Exploratory Learning in Digital Age.*

Qadir, J. (2023). Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education. *IEEE Global Engineering Education Conference (EDUCON) 2023.*

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, *6*(1).

Stokel-Walker, C., & Van Noorden, R. (2023). What ChatGPT and generative AI mean for science. *Nature*, *614*(7947), 214-216.

Toda, A., Valle, P. H. D., & Isotani, S. (2018). The Dark Side of Gamification: An Overview of Negative Effects of Gamification in Education. In A. Cristea, I. Bitencourt, & F. Lima (Eds.), *Higher Education for All. From Challenges to Novel Technology-Enhanced Solutions. Communications in Computer and Information Science*, Vol. 832. Cham: Springer.

Wadhwa, V. & Joshi, J. K. (2024). AI Chatbot Generated Content and Similarity Index: Challenge for Plagiarism Detection Software. *International Journal of Library Information Networks and Knowledge, 9*(1), 70-81.

Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research, and Evaluation*, *2*(2). https://doi.org/10.7275/ffb1-mm19