AI-assisted Learning with ChatGPT and Large Language Models: Implications for Higher Education

Samuli Laato
Gamification Group
Tampere University
Tampere, Finland
samuli.laato@tuni.fi

Benedikt Morschheuser
Inst. of Information Systems
FAU Erlangen-Nürnberg
Nuremberg, Germany
benedikt.morschheuser@fau.de

Juho Hamari
Gamification Group
Tampre University
Tampere, Finland
juho.hamari@tuni.fi

Jari Björne
Dept. of Computing
University of Turku
Turku, Finland
jari.bjorne@utu.fi

***Abstract*—The recent progress in generative AI models, particularly large language models (LLMs), has brought about a transformation in the field of education. Conversational LLM services, such as Google's Bard and OpenAI's ChatGPT, offer students access to many abilities such as summarization and generation of text and code, and on-demand replies to questions on expert topics. In this paper, we observe ChatGPT to explore how LLM services impact learning and instruction in higher education. First, we mapped the capabilities of the system by reviewing the grey literature on ChatGPT and using the system ourselves for two months. Second, we selected a Bachelor level computer science curriculum from a Finnish university, and examined the impact of ChatGPT on the offered courses. As an outcome of this study, we highlight 13 implications for students' learning in higher education, and discuss the contemporary future of AI-assisted learning in universities and beyond.**

***Index Terms*—ChatGPT, Bard, GPT-4, generative language models, large language models, higher education, learning**

I. Introduction

"ChatGPT to me seems to weld an English professor's writing skills to an encyclopedia's knowledge base and a kindergartener's reasoning ability, producing impressive and highly polished nonsense." - Tim Sweeney, Founder and CEO of Epic games

In 2017, Vaswani et al. [1] introduced what we now call the Transformer architecture, a type of deep neural network that uses attention mechanisms to process sequential data such as text. The attention mechanism in the Transformer allows the model to capture long-range dependencies in the text data and has proven to be very effective for NLP tasks. This structure was used as a basis for OpenAI's first Generative Pre-training Transformer (GPT) model in 2018 [2], and is the backbone for the current generation of generative large language models (LLMs).

In November 2022 OpenAI released a GPT 3.5-powered web chatbot called *ChatGPT*, which, by the time of January 2023, already had 100 million monthly active users [3]. The success of ChatGPT was immediately followed by competitors announcing similar services. For example, Google revealed in February 2023 their own LLM chat service called *Bard*, which is based on their Language Model for Dialogue Applications (LaMDA) [4]. The capabilities of these LLM-based chatbots to understand natural language and accurately respond to it [5] has attracted significant attention from academia, industry and laypeople. As LLM-based chat services are still improving and becoming more commonplace, it is critical to also evaluate their potential impact on teaching and learning [2], [6]. Thus, in this paper, we focus on the impact of conversational LLM services for higher education. Thus, we explore the following research question (RQ):

RQ: *What implications LLM-based chat services such as ChatGPT have on students' learning processes in an existing Bachelor level computer science curriculum?*

By addressing the RQ we provide practical information for students and educators in higher education regarding the implications that LLMs have on learning and instruction. The rest of this study is structured as follows. First we describe the research method, followed up by our findings. We conclude the study by discussing the limitations and proposing directions for future research.

II. Method

For answering the RQ, we focus on OpenAI's popular LLM service ChatGPT [3]. We followed a three-step process as outlined below.

First (familiarization), the authors used ChatGPT daily for two months for various tasks and purposes to familiarize ourselves with its core functionalities and capabilities. The authors also actively followed the public discourse on ChatGPT on Twitter and traditional news outlets.

Second (data collection), we recorded all the potential use cases of ChatGPT for higher education learning and instruction that we came by during these two months, either by noodling with ChatGPT ourselves, or through discovering new use cases from social media or media articles. As our main source of information on ChatGPT's capabilities, we used OpenAI's

article about the system [5], but we also referred to other sources, particularly grey literature such as newspaper articles and blog posts on expert websites.

Third (analysis), we engaged in reflection and discussion to estimate what implications the discovered use cases and capabilities of ChatGPT have on learning and instruction in higher education. In order to do this, we selected the educational curriculum of a Finnish University, focused on a Bachelor degree curriculum for computer science [7], and then critically evaluated how ChatGPT could assist students in passing the courses.

III. Findings

A. Overview of ChatGPT and its capabilities

ChatGPT represents the latest development in the evolution of ever larger language models. As in many fields, deep learning has had a profound impact on computational linguistics. Modern deep learning became first prominent with image analysis models developed on the ImageNet corpus [8]. Text mining applications followed within a few years, with methods such as convolutional and recurrent neural networks (in particular the LSTM architecture) applied to various computational linguistics tasks [9], [10]. However, a significant leap forward was introduced in 2018 with Google's BERT model, a general purpose LLM [11]. BERT was based on a transformer architecture, but its high performance was also due to having been trained on a massive text dataset. This large knowledge base encoded into the model allowed it to be utilized for various downstream tasks through the process of transfer learning [12]. A few years later, the GTP models took LLMs to a new level, with GPT-3 having 175 billion parameters compared to the 340 million parameters of the BERT_{LARGE} model. The ChatGPT model, which is the primary focus of the current paper, is a follow-up of the GPT-3 model [5].

Even on the most basic surface level LLMs such as ChatGPT offer a multitude of benefits for students. In Table I we list seven capabilities of ChatGPT, discovered through the previously described method. However, due to the chosen approach approach, Table I does not represent an exhaustive list of the capabilities, but rather, it highlights how at minimum ChatGPT and other similar LLM-based services are already disrupting multiple fields.

While the potential use cases of ChatGPT are manifold, currently it has various limitations [5]. For example, it sometimes "hallucinates" things or provides inaccurate information [14] and occasionally fails simple logical reasoning tasks such as math questions [15]. Keeping these limitations in mind, next we analyse what implications these have for learning and instruction in higher education.

TABLE I
Capabilities of ChatGPT and potential use cases for students in higher education.

Ability	**Use cases**
Conversational AI	As a LLM optimized for dialogue [5], ChatGPT helps users test their ideas, and reflect upon their thinking through receiving conversational feedback. It allows users to ask focused questions regarding topics they want more information about.
Language translation	Users can translate text back and forth from various languages, or simplify text, which can be used to, for example, read otherwise incomprehensible text.
Text generation	ChatGPT allows users to generate text via prompts. For example, students can generate entire abstracts for their academic papers or peer-review comments.
Code generation	Similarly to text generation, ChatGPT is able to produce programming code in multiple languages [13]. Though currently imperfect, this can accelerate development work. This opens new opportunities for students in both technical and non-technical fields.
Sentiment analysis	Students can conduct behavioral research more easily on datasets, for instance of social media posts, to understand how participants feel about a topic.
Summarizing essays and articles	Students can quickly obtain the main points of long articles without the need to get into the details. They can do this for transcripts, news articles or social media threads.
Communication improvement	Students can use ChatGPT to fix the grammar in their writing. ChatGPT can also more holistically improve the communication by reorganizing sentences and entire paragraphs, and suggesting words and phrases.

B. Implications of LLM services such as ChatGPT for learning and instruction

1) Conversational AI: ChatGPT can support students' thinking and idea generation in several ways. First, it can help in brainstorming. If a student is tasked with, for example, creating an application that leads to a more sustainable world, they could engage in a conversation about the topic with ChatGPT. This conversation could spark new ideas. Second, ChatGPT can help students reflect on their ideas and conceptions, and to see whether their understanding has some biases or flaws, or perhaps the students have involuntarily ignored some relevant information. For example, a student could share their initial thoughts on a topic with ChatGPT, and could receive suggestions of relevant connected information. This could be useful when preparing for oral exams, or simply for learning a more balanced view of the topic at hand. Third and finally, ChatGPT has concrete, as well as meta-functionalities, that guide students towards critical thinking. Through conversation and debate, ChatGPT can challenge students' ideas and broaden their horizons. Students can, for example, ask ChatGPT to take an opposing stance on some topic, and then engage in debate. However, even more broadly on the meta-level, ChatGPT will require students to constantly and critically evaluate the output of the system. This too, will contribute to students' learning of critical thinking.

These three dimensions of use cases (brainstorming, reflection, and critical-thinking) are all important aspects relevant to the computer science bachelor curriculum [7]. For example, all three dimensions are present (either explicitly or implicitly) during project, mentoring, seminar and work placement

courses. It is thus important that educators instruct students how to use ChatGPT and other conversational LLM services in a way that best supports their learning. Students should be given knowledge of the limitations of these systems as well as how they work, but should not be discouraged from using them due to the outstanding benefits. Educators should also instruct students about the topics they should engage in conversation with the chatbot. For example, during the lecture on the course on data structures and algorithms, the lecturer could tell students to talk with ChatGPT about the differences between different sorting algorithms, and the use cases for each. Taken together, we thus provide the following implications.

Implication #1: *As a conversational agent, ChatGPT supports students' idea formation and critical thinking.*

Implication #2: *Educators should instruct students about which topics related to the course they could engage in conversation with ChatGPT about.*

2) Language translation: ChatGPT can help students read text written in foreign languages. This is useful when some material is only available in a specific language. Of course there are multiple other translation tools also available, and thus, ChatGPT is not unique in this regard. However, there is a danger that ChatGPT can also be used for cheating, since many plagiarism tools such as Turnitin cannot accurately detect plagiarism through translation [16], and ChatGPT can be used to obfuscate the original authorship of text via multiple ways. This has enormous implications for not only students' learning, but also for academic research as a whole. Looking at the number of written tasks in the chosen curriculum [7], it is clear that this is a critical issue that requires immediate attention from the academia and educators.

OpenAI has already presented a classifier for indicating AI-written text[1], which currently has a modest accuracy of 26%. This system has been published to mitigate the concerns of plagiarism related to ChatGPT. Companies dedicated to development of plagiarism-detection software (such as Turnitin) have a clear incentive to help in detecting fraud based on generative AI models, so it is likely that detection methods will become incorporated into already widely used plagiarism detection systems. Conclusively solving this issue is beyond the scope of this paper, but we distil the following immediate implications of this for learning and instruction in higher education as follows:

Implication #3: *With ChatGPT students have an improved ability to access documents in foreign languages.*

Implication #4: *Students may use ChatGPT to circumvent copyrights and plagiarism detection software by translating text back-and-forth in other languages, and asking the system to reformulate paragraphs in new ways.*

3) Text generation: Even in technical fields, the educational curricula are filled with writing tasks. Looking at the selected curriculum of the Bachelor degree in computer science [7], there are at least the following written tasks: (1) The Bachelor thesis. This is a 10 credit course altogether involving the writing process of the entire thesis; (2) Course essays. Multiple courses currently require students to write reflective essays, study diaries or other written tasks; and (3) Presentations. Multiple courses currently require students to present their work, and ChatGPT can help generate the content for these. Even if not used to generate the whole task, ChatGPT supports the generation of snippets of text, such as paragraphs or sentences. Here it is important that the students using the tool are aware of the content of the output, and are able to evaluate whether to accept or reject it.

One implication of the ubiquitous use of ChatGPT for writing tasks is that all text produced in the world could slowly become more and more uniform, if it increasingly originates from the same source: ChatGPT. Of course ChatGPT will be able to mimic the styles of different writers to counter this, and it remains unclear that as new LLM services enter the market, how exactly will this convergence of the various writing styles around the world happen, or will it happen at all.

Implication #5: *Students can prompt ChatGPT to generate paragraphs of text for essays, presentations and their thesis.*

Implication #6: *The ubiquitous use of LLMs may contribute to all written text becoming more heterogeneous in style throughout academia and beyond.*

4) Code generation: ChatGPT can generate simple code snippets in response to specific prompts or questions, such as writing a function to perform a specific task. However, the generated code is not always correct or optimal, and often requires manual editing and debugging. Furthermore, ChatGPT is not equipped with the ability to understand or write code in all programming languages, so it may not be able to generate code for more complex or specialized tasks. Despite these limitations, it remains a powerful tool for assisting in various programming tasks and statistical analyses.

In our example curriculum [7], the majority of programming courses contained homework programming assignments, whether they were about constructing simple methods, designing algorithms, implementing pseudo code sorting algorithms in specific languages, solving specific tasks or generating complete software. ChatGPT can aid students in the simpler tasks, and can help them progress faster and further. However, this raises the question of how students' learning will be affected when they are offloading some workload to services such as ChatGPT. To some degree, ChatGPT is replacing forum-based coding assistance websites such as Stack Overflow [17]. In comparison to Stack Overflow, ChatGPT provides more precise and personalised responses, which can be helpful for learning in many instances.

We also tested ChatGPT on sit-in exam questions with pre-existing grading criteria in the context of a university machine learning course. For simple term definition questions, the ChatGPT answers were consistently correct. For longer essay questions the system wrote extensive, high quality answers, but

[1]https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text/

also with omissions of specific topics emphasized in the course material, which almost always appear in real student answers. There seemed to be more errors the more applied the answer was required to be. In other words, ChatGPT works best when it can regurgitate "encyclopedia knowledge", common definitions of well-known concepts. While such questions are common in exams, homework should, and usually does contain more applied tasks with a strong personal component of contribution unique to each student, tasks on which ChatGPT is less likely to be able to provide complete answers.

Implication #7: *ChatGPT allows students to find precise and quick answers for various (mostly simple) programming tasks.*

Implication #8: *Over-reliance on ChatGPT for programming can hinder students' learning of basic routine programming skills.*

Implication #9: *ChatGPT allows programmers and non-technical people to make applications faster, resulting in more complete IT-products.*

5) Sentiment analysis: Sentiment analysis is a powerful tool for marketing research [18], but it also has a wide variety of purposes. Looking at the case curriculum [7], it could be utilized by students as a research approach when they are doing their Bachelor's thesis, or as a smaller tool when preparing presentations for computer science seminars and projects.

Implication #10: *ChatGPT offers students an easy access to an otherwise elusive form of text analysis: sentiment analysis.*

6) Summarizing essays and articles: Almost all university courses in the observed curriculum contain reading tasks, whether these are research papers, books, presentation slides or other course material [7]. For this reason, the ability of ChatGPT to consume course material and distil it into manageable bits can be considered relevant for learning in higher education. However, there are a few shortcomings we need to discuss. First, LLMs may summarize or understand some parts of the material wrongly. Adding an extra layer of interpretation between the source material and the reader can lead to inaccuracies. Second, since ChatGPT and other LLM services are currently commercial products, uploading copyrighted course material there for processing may be an ethical or a legal issue. Third, reading summaries as opposed to full texts may hinder the ability to understand the nuances in the given topic.

With these limitations in mind, ChatGPT can help students in, for example, by answering spot-on questions about topics students are confused about while reading. For the purpose of this research, the authors ran multiple queries on ChatGPT over the course of multiple weeks about computer science topics ranging from microprocessors to specific questions on various programming languages. While there were inaccuracies and even mistakes, in general the information provided was accurate.

Implication #11: *ChatGPT's ability to provide on demand answers to questions that are troubling students contributes to more personalized learning.*

Implication #12: *ChatGPT offers students improved options in finding and presenting information in an understandable and digestible way.*

7) Communication improvement: Since the curriculum observed in this work [7] contains several written tasks, the communication improvement help of ChatGPT can be useful. However, there are some potential issues. For example, if students blindly trust a LLM to re-word their text, they may not learn to critically focus on specific word choices, or the outcome may not completely reflect their intention. While further research is needed on the specific impacts of this on learning, we are currently able to formulate the following implication.

Implication #13: *Educators need to emphasise that even if students use ChatGPT to reorganize their text, they will ultimately be responsible for the entirety of what they hand in.*

IV. Discussion

A. Key Findings

The discovered 13 implications are summarized in Table II. On the meta level, these implications provide substantial evidence that conversational LLM services are powerful tools for supporting learning. Particularly relevant for learning is the dialogue between the chatbot and the user. In the academic literature, there are multiple examples of how children effectively learn by asking their parents questions (see e.g., [19]). In formal classroom-based education it is not possible to have a personal "sensei" for all the students. However, ChatGPT not only provides an actor from whom questions can be asked from, but also contains an enormous library of knowledge. While chatbots and conversational agents have existed prior to the era of ChatGPT, nothing equally powerful has been made available to the general public in this scale before. Teachers in higher education are now be forced to reorient their teaching approaches to make use of these newly available modes of learning.

B. Limitations

LLM-based chat services such as ChatGPT and Bard are still relatively new to the large population, and while there is a substantial body of academic literature supporting the technical basis of these systems, the literature on the use cases and implications is currently in its infancy. In this work, we focused on the context of teaching and learning in higher education, but the LLM systems have various other implications for society. Furthermore, our analysis was based largely on ChatGPT powered by GPT-3.5 and GPT-4, and in the near future, we may see new capabilities such as multimodality, which will require further examination.

TABLE II
SUMMARY OF IMPLICATIONS OF CONVERSATIONAL LLM SERVICES ON LEARNING AND INSTRUCTION IN HIGHER EDUCATION

#	Implication for teaching and learning
1	Conversation with fine-tuned LLMs supports students' idea formation and critical thinking.
2	Educators can instruct students about which topics, related to the course, they should discuss about with LLM services.
3	With LLMs, students have an improved ability to access documents in foreign languages.
4	Educators and evaluators need to be aware of cheating opportunities arising from LLM services.
5	Students can prompt LLMs to generate paragraphs of text for them for essays, presentations and their thesis.
6	Ubiquitous use of LLMs for text generation and editing may contribute to all written text becoming more heterogeneous.
7	LLM services allow students to find answers quickly for various programming tasks.
8	Reliance on LLMs can hinder students' learning of basic routine skills. Impacts of this need to be examined.
9	LLM services allow programmers and non-technical people to make applications faster.
10	Students can use LLMs for automating otherwise time-consuming analyses, such as a social media post sentiment analysis.
11	LLM services' ability to provide on demand answers to questions is groundbreaking for personalized learning.
12	With LLM services students can more easily seek information and transform it to a digestible format.
13	Educators should emphasise, that students are ultimately responsible for the authenticity and correctness of the text they hand in, even if they use LLMs to generate it.

C. Future research agenda and conclusion

There are multiple alternative methods that could be used to probe this research topic. For example, interviews with key stakeholders (teachers and students) could offer a more practice-grounded view. Alternatively, case studies of actual uses and reactions could be valuable. Overall, since LLM services are disrupting multiple aspects of teaching and learning in higher education, it is critical to continue research on this topic, assessing the constantly developing technologies, and to evaluate whether entire educational systems require re-design.

We conclude with an anecdote that directs us to think about our AI-driven contemporary future. In 1997, the world witnessed a highly public chess match between Garry Kasparov, the reigning world champion, and IBM's computer chess program, Deep Blue [20]. Kasparov ended up losing the match, but his team raised questions about potential human interference in the computer's moves. While the combination of human and machine at the time resulted in a stronger chess player than either alone, today's neural network-based chess AIs, such as DeepMind's Alpha Zero [21], have surpassed human capabilities in every aspect of chess, eliminating the need for any human involvement. As we witness the remarkable abilities of LLM services such as ChatGPT, they are still best operated by a human. However, it is clear that the 20-year journey from Deep Blue to Alpha Zero has similarities to the current situation. We are now witnessing a metaphoric rematch between Kasparov and Deep Blue, only this time it is not about chess, but the entire processes of human language and reasoning.

REFERENCES

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[2] L. Floridi and M. Chiriatti, "Gpt-3: Its nature, scope, limits, and consequences," *Minds and Machines*, vol. 30, pp. 681–694, 2020.

[3] K. Hu, "Chatgpt sets record for fastest-growing user base - analyst note," *[ONLINE], available at: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/, visited February 9, 2023*, 2023.

[4] S. Pichai, "An important next step on our ai journey," *[ONLINE], available at https://blog.google/technology/ai/bard-google-ai-search-updates/*, 2023.

[5] OpenAI, "Chatgpt: Optimizing language models for dialogue," *[ONLINE], available at: https://openai.com/blog/chatgpt/, visited January 20, 2023*, 2022.

[6] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier *et al.*, "Chatgpt for good? on opportunities and challenges of large language models for education," 2023.

[7] U. of Turku, "Study guide of information and communication technology bachelor degree 2022-2024," *[ONLINE], available at: https://opas.peppi.utu.fi/en/degree-studies/faculty-of-technology/14002/33054, visited February 9, 2023*, 2023.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[9] H. Liang, X. Sun, Y. Sun, and Y. Gao, "Text feature extraction based on deep learning: a review," *EURASIP journal on wireless communications and networking*, vol. 2017, no. 1, pp. 1–12, 2017.

[10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[12] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big data*, vol. 3, no. 1, pp. 1–40, 2016.

[13] B. Marr, "How chatgpt and natural language technology might affect your job if you are a computer programmer," *[ONLINE], available at: https://www.forbes.com/sites/bernardmarr/2023/01/23/how-chatgpt-and-natural-language-technology-might-affect-your-job-if-you-are-a-computer-programmer/?sh=3ccdba86174b, visited February 10, 2023*, 2023.

[14] C. Thorbecke, "Google's ai chatbot bard had an inaccurate response in public demo," *[ONLINE], available at: https://edition.cnn.com/2023/02/08/tech/google-ai-bard-demo-error/index.html, visited February 9, 2023*, 2023.

[15] T. I. Express, "What happens when chatgpt has to solve a basic math problem? check out its response," *[ONLINE], available at: https://indianexpress.com/article/technology/tech-news-technology/chatgpt-messing-up-a-basic-math-problem-apologising-and-correcting-itself-8346602/, visited February 9, 2023*, 2022.

[16] M. Jones and L. Sheridan, "Back translation: an emerging sophisticated cyber strategy to subvert advances in 'digital age'plagiarism detection and prevention," *Assessment & Evaluation in Higher Education*, vol. 40, no. 5, pp. 712–724, 2015.

[17] I. Isaac, "Is chat gpt bad business for stack overflow?" *[ONLINE], available at: https://frontdreams.com.ng/insight/is-chat-gpt-bad-business-for-stack-overflow, visited January 20, 2023*, 2022.

[18] M. Rambocas and B. G. Pacheco, "Online sentiment analysis in marketing research: a review," *Journal of Research in Interactive Marketing*, 2018.

[19] E. A. Jant, C. A. Haden, D. H. Uttal, and E. Babcock, "Conversation and object manipulation influence children's learning in a museum," *Child development*, vol. 85, no. 5, pp. 2029–2045, 2014.

[20] M. Campbell, A. J. Hoane Jr, and F.-h. Hsu, "Deep blue," *Artificial intelligence*, vol. 134, no. 1-2, pp. 57–83, 2002.

[21] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel *et al.*, "A general reinforcement learning algorithm that masters chess, shogi, and go through self-play," *Science*, vol. 362, no. 6419, pp. 1140–1144, 2018.