Students' perception and experience using ChatGPT (Large Language Models)

Nicole O'Brien   Michael Dunlop

Suffolk University

Abstract: Artificial Intelligence (AI) tools and Large Language Models (LLMs), such as ChatGPT, have impacted higher education. The future expectation is that LLMs will pose significant challenges and opportunities. This study addressed students' perception and experience using LLMs in a course assignment administered in an undergraduate Data Science course. A sample of 18 students (n=18) was included in the study. Although ChatGPT, and other LLMs, were previously used by 94% of students, this study found mixed results on the accuracy and time saving benefits when using these types of AI. Students reported using LLMs in several academic areas, including idea generation, structuring outlines, obtaining data/information, fact checking, coding, as well as spelling and grammar in a written document. The results in this study found that LLMs met expectations from 61% of students, whereas the other 39% did not have their expectations met. The accuracy and time to fact check the output of LLMs were polarized based on student responses. Further research is recommended in the perception, use, and expectations using LLMs/ChatGPT in the classroom. Student integrity, academic dishonesty, and off-loading to AI are other areas also needing further investigation.

Keywords: ChatGPT, Large Language Models, Artificial Intelligence, Chatbots

Artificial Intelligence (AI) has dramatically impacted higher education. The AI chatbot that quickly rose to prominence in 2023 was the release of the San Francisco-based artificial intelligence lab's OpenAI ChatGPT (Dickson, 2023). ChatGPT stands for "Chat Generative Pre-Trained Transformer," which means it is trained to predict the next word in a given sequence on large amounts of textual data (Wilson, 2024). Mollick and Mollick (2023), drew an analogy to the autocomplete feature in search bars to predict what word should populate next. However, instead of one word or a string of a few words, when something is typed into ChatGPT, or another Large Language Model (LLM), it will extend the prompt logically based on its training to create a substantial amount of textual data (Mollick and Mollick, 2023). Bill Gates (2023), believed that there are several ways that AI can assist teachers, such as assessing a student's understanding of a subject.

AI and LLMs are becoming prevalent in environments outside of academia. In fact, the Big Four audit firm, PricewaterhouseCoopers LLP, teamed up with ChatGPT owner OpenAI to offer clients advice generated by AI as the audit firm looks to cut costs and boost productivity (Anghel, 2023). Yet, as organizations explore the use of generative AI in learning and development, they must maintain a strong sense of responsibility that requires adherence to legal standards and an ongoing commitment to adaptability in respect to the evolving technological landscapes (Dellarocas, 2023).

Utilizing AI, and LLMs in particular, must be navigated carefully due to concerns about potentially nefarious uses, such as allowing students to cheat on tests and facilitating the spread of misinformation (Dickson, 2023). Within higher education, LLMs are a mechanism for student dishonesty. Bill Gates (2023) watched in awe this past September when the ChatGPT AI model answered 59 out of 60 Advanced Placement Biology exam questions correctly with exceptional open-ended answers to six exam related questions. Impressive as they are, generative AI tools are inherently probabilistic, meaning that they are prone to being wrong with the danger that what is produced is inaccurate, unfair, or offensive (Branscombe, 2023). This concern is particularly relevant to global organizations where cultural diversity is important (Dellarocas, 2023). Dellarocas (2023), the Assoc Provost of Digital Learning & Innovation at Boston University,

further declared that it may be prudent to form ethics committees, invest in AI literacy and awareness programs, and establish clear guidelines and processes for AI governance. It is imperative to expect results that cannot be used right away and always be alert in recognizing the ways generative AI can be usefully wrong (Branscombe, 2023). With that in mind, Microsoft has adopted Copilot rather than autopilot for most of its generative AI tools, which allows humans to be in control with the AI, like ChatGPT, being the copilot (Branscombe, 2023). Ask Copilot for 20 suggestions for an appealing presentation title and within seconds, a relevant suggestion list will be provided, many won't be what is being looked for, but a handful may be relevant or at least spark some creative inspiration (Microsoft Worklab, 2023).

Given that LLMs have been pre-trained on large amounts of information scraped from either specific databases or the available documents on the internet, they are capable of many tasks across many fields, yet there is no clear-cut instruction manual nor is there any guarantee that the information is accurate or unbiased. This makes it difficult to know what tasks they are good or bad at without more investigation since LLMs, like ChatGPT (or simply chatbots) often make mistakes (Mollick and Mollick, 2023). "Like humans, chatbots also make mistakes, commonly referred to as 'hallucinations,' and will write things that are inaccurate and odd (Franklin, 2023, AI standards differ, para. 6)." LLMs will complete tasks they are given and even make-up answers that are outside the available data set to complete the task. Dellarocas (2023) asserted that continuous monitoring and validation where learners can report discrepancies or confusion are essential to ensure the quality of the AI output content. It is up to the user to verify the accuracy of the LLMs outputs, since like humans, they are prone to mistakes.

Further investigation on LLMs and the perception, use, and future prevalence in higher education is important to explore. This study is of particular importance since it addresses a gap in the literature in the emerging trends AI will present in the higher education classroom. The wealth of data that is available to students and the AI tools and techniques that have evolved to translate the data into useful information is constantly evolving, with mixed results. A study involving students and how they interface with AI, and specifically the level of accuracy of AI, is critical to understand.

Method

To assess the impact of AI, and more specifically LLMs (e.g., ChatGPT) in the classroom. An exploratory study was conducted during the fall 2023 academic semester with 18 students enrolled in a data and decisions analysis course. The students researched a country on their own and then performed the same research utilizing an LLM. Next, students compared and contrasted the results from each approach. Finally, students answered open-ended questions (see Appendix A) based on their experience using the LLM as it pertained to the information created, as well as for fact checking. Eighteen students (n=18) completed an anonymous survey. Since this study was an exploratory case study, the sample size, though small, was enough to understand how students utilize LLMs. According to Myers (2013), small sample sizes are still effective when the research is exploratory in nature. The incentive for students to fill out the survey was a five percent increase in their grade for the project, which many students chose. This research was granted IRB approval in November 2023 with the data collected in December 2023.

The participants (students) were mainly second year students with a few being in their third or fourth year in a bachelor's in business (a four-year) degree with various majors represented. The institution is a mid-sized private university with about 8,000 undergraduate and graduate students.
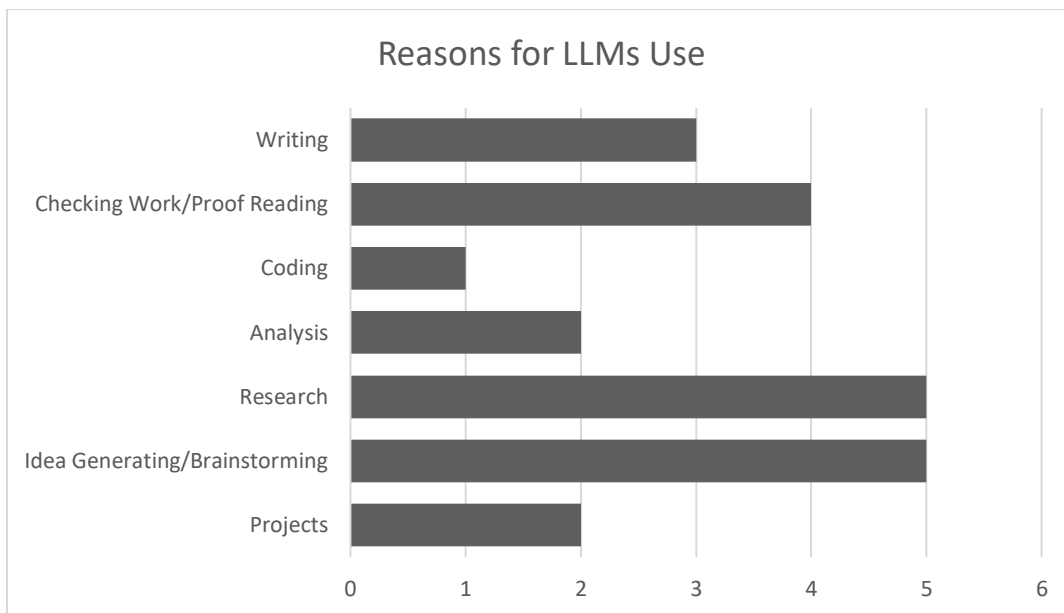
<div align="center">**Results**</div>

Usage of LLMs before the assignment.

Of the sample, 17 students had previous experience with using LLMs, while one did not. The students that had previous experience with LLMs had used them for various reasons (see Figure 1).

Figure 1

Reasons for LLMs Use

LLMs were mostly used as a tool by the students to aid in the creation of work from helping with idea generation, structuring outlines, for data/information, and for cleaning up spelling and grammar in a written document. Others used the LLMs for advice in using different programs, to help with coding, and to summarize material. As stated by a student: "I used AI before to go through and structure outlines, clean up essays and writing assignments, and help with coding." This student's observation highlights the variety of uses students make of AI programs across all elements of assignment production.

Expectations of LLMs

The expectations of LLMs were mixed, some students expected the tool to be an "all knowing super-intelligent force." Some expected a base level of information, with the information given being truthful and accurate. Following are some student excerpts:

"I was expecting it to be able to find Base level information for me."

"AI is often portrayed as an all-knowing, super-intelligent force."

"To be truthful and very accurate with its information given."

If students expect the AI to be accurate, it will lead to problems. Our goal as instructors is to help them temper these expectations of accuracy.

Other students, however, were less optimistic that LLMs were always accurate.

Following are some student excerpts:

"That it's a good tool but isn't always 100% reliable."

"I thought it would be helpful, but I understand that it is not always 100% accurate."

"They were pretty accurate although some of the information it provides sometimes isn't true or relevant *to* today."

Of the 18 survey participants, six were pessimistic as to the accuracy, reliability, and time involved using LLMs, as indicated by the following student excerpts:

"That AI isn't exactly reliable."

"That it would take longer for the AI to give me what I was looking for."

These less optimistic observations show students are applying their critical thinking skills to the results of their AI prompts. Instructors can leverage this skepticism to encourage students to take a more cautious approach to AI results.

The students' expectations after use were also assessed with 61% stating that the LLMs met or exceeded their initial expectations, whereas the other 39% found that their expectations were not met as the LLMs made mistakes or just didn't perform at the level they had hoped. Here are two student excerpts illustrating the contrasting views on the LLMs:

> "Yes, it was quicker than I expected. I thought that what I was asking had to be precise in order for AI to deliver results and that it would require multiple attempts to get right. This wasn't the case."

> "My expectations were different because I thought AI usually did not make any mistakes, but I found mistakes while using them."

Asking the students to reflect on their experiences with the LLMs helped them be aware of the contrast of their expectations and the reality of using the LLMs.

User's Opinions of LLMs

The students were asked to compare the LLMs' output with what they had produced. They reported that LLM was good for some applications as it remembers grammar rules and the writing was better, such that the writing in some cases was more sophisticated. It was found that the information may be better for some purposes like editing, yet in composing text, it was lacking. Students noted the following observations:

> "I found that the writing of AI can sound more sophisticated, but the writing of the group would be more direct."

> "It was more detailed and often contained really accurate information."

> "While the AI gave a good general report, the writing that was created by the group included more specific points along with statistical figures."

Asking the students to critique the results of using the LLMs seemed to result in a greater awareness of the strengths and weaknesses of the LLMs.

The level of accuracy for some was found to be an issue as students were asked to fact check the information/statements created by the LLMs. Depending on the questions asked, the accuracy of the LLM answers were mixed. One student reported that the LLMs contradicted itself, which happens when it cannot find an answer within its database. Others found that the LLMs gave very general information, while others stated the information given was accurate and more in depth than they had expected. Some students also noted that it was difficult to fact check as the sources that the LLMs used were not cited, thereby making it difficult to check its sources, which indicates a misunderstanding of how the LLMs compile data. LLMs do not compile information from specific sources, rather they synthesize statistical samples of patterns of data from huge data sets. Students commented on the following topics:

"Lots of errors with AI, and info that doesn't apply."

"Incredibly hard to fact check. Each AI gave different answers."

"I just had to run questions multiple times, to make sure AI wasn't contradicting itself (which happens often)."

"I did not find any errors in the way AI did the work. It was very easy to fact check of the work of the AI."

This variety of student commentary highlights how important the skill of reviewing LLMs results can be. Part of student debriefing should be to assess what prompts result in accurate results and what prompts require intensive review for accuracy.

Further, the students' opinions on willingness to work with LLMs again were positive. Nevertheless, when asked if they would be expected to fact-check the work of the LLMs, students were less willing to use the tool for work purposes. Thirteen still wanted to use the LLMs, while five decided it was not an efficient use of their time. Of the 13 students who were still willing to use LLMs, the amount of time they gave to fact-check was between ten minutes to an hour. Student comments included:

"Yes, for important things it would be worth it, but it should be easy to fact check. Maybe 10 minutes fact checking for every 1 hour of work."

"It would depend on what I was using the AI for. If it was to help code or to write an email, I'd be willing to do the work. But if it's for a more academic writing piece, then no. I'm willing to spend a decent amount of time on this."

"I've found that it's actually quicker to not use AI because of all the cross referencing I am constantly doing. That's why I only would use it on something more general but never on like a test or serious essay."

These student comments indicate that they had developed a more sophisticated approach to using the LLMs for different kinds of work because of analyzing and reflecting on the results of their work.

Conclusions and Recommendations

Given the swift emergence of AI in higher education, this exploratory study found that students perceive benefits and drawbacks from incorporating LLMs in the classroom. The study found that students used LLMs for multiple reasons. This new tool has limitations such as hallucinations or poor-quality responses (Kasneci et al, 2023; Kumar et al., 2023). Some students were aware that LLM responses needed to be examined and fact checked, but the many seemed to believe that this technology was infallible. Thus, even after use, most students felt that the technology met their expectations while a smaller portion felt it did not. Kasneci et al. (2023) suggested learners may rely on this technology heavily without understanding its inherent limitations, which means students must be made aware when using LLMs. Current LLMs have limitations for school and work projects. Further, students need to be made aware of the issues with LLMs and taught how to critically assess the information created.

Some students in the study realized that the work produced by the LLMs needs to be verified and corrected, with some saying that use of LLMs was too much work, while others stated that they would use it depending on how much work it would save them overall. Finding

the correct balance of when to use LLMs is critical. For instance, chatbots can be an efficient tool for cognitive offloading, thereby reducing the demands it takes to complete a task (Franklin, 2023). Allowing students to complete, or offload, lower-level tasks may free up student time for higher-level thinking and tasks (Franklin, 2023). Nevertheless, it is also important to acknowledge that many courses, especially foundational courses, require students to master lower-level skills and information to draw upon in learning higher-level thinking. Additionally, in the future, when LLMs are more prevalent, individuals will be expected to know what should be offloaded to the LLMs and what needs the individual's attention (Kasneci et al., 2023; Meyer, J. et al., 2023). A significant opportunity for organizations to transform their learning and development strategies, making them more efficient and effective with the rapidly evolving skills landscape, will occur as this technology matures (Dellarocas, 2023). Determining what can and should be offloaded to AI needs additional analysis.

Another area to explore further pertains to student integrity and academic dishonesty policies. Having a statement in a course syllabus on the usage of LLMs is a good starting point. For example, the University of Washington (2024) provides some basic syllabus language for faculty on how students may use AI-based tools in their courses. Stating instructor expectations in the syllabus is important as expectations may differ from course to course. Issues arise with LLMs as to what constitutes cheating or over-reliance on technology. Cheating with LLMs is a form of academic dishonesty. Franklin (2023) would not allege the use of LLMs as student plagiarism since plagiarism is using an actual person's idea or language without proper citation. Instead, students are typically charged with falsifying academic information. Yet, plagiarism issues could arise with the use of LLMs as the technology produces work that is not the student's work, which could be considered plagiarism (Kasneci et al., 2023).

There are some limitations to this study. As the sample size is small, further research should be conducted to understand how students utilize LLMs. Further research should examine what skills students need to critically assess when using the technology along with the exercises/learning material best suited to aid students' critical thinking skills regarding LLMs. The students in this study were all business students in a technology related course in which they

were provided with guidance on LLMs. Research on non-business students and non-technical academic disciplines and how they can utilize LLMs should be further explored.

References

Anghel, L. (2023, October 17). *PwC offers advice from bots in deal with ChatGPT firm OpenAI.* Bloomberg News. https://www.bnnbloomberg.ca/pwc-offers-advice-from-bots-in-deal-with-chatgpt-firm-openai-1.1985564

Branscombe, M. (2023, December 6). *6 generative AI hazards IT leaders should avoid.* CIO Magazine. https://www.cio.com/article/1251205/6-generative-ai-hazards-it-leaders-should- avoid.html

Dellarocas, C. (2023, December 8). *How GenAI Could Accelerate Employee Learning and Development.* Harvard Business Review. https://hbr.org/2023/12/how-genai-could-accelerate-employee-learning-and-development

Dickson, B. (2023, June 5). *What is ChatGPT? A basic explainer.* PC Magazine. https://www.pcmag.com/how-to/what-is-chatgpt-a-basic-explainer

Franklin, A.J. (2023, August 21). *Off to college with AI: The good, bad and prospects for cheating.* The Buffalo News. https://buffalonews.com/news/local/education/off-to-college- with-ai-the-good-bad-and-prospects-for-cheating/article_4b2f77de-2fa6-11ee-b94b-57a493d0ed6c.html

Gates, B. (2023, March 21). *The Age of AI has begun.* Gates Notes: The Blog of Bill Gates. https://www.gatesnotes.com/The-Age-of-AI-Has-Begun

Kasneci, E., Sessler, K., Kuchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U.,

Groh, G., Gunnemann, S., Hullermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., and   Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large

language models for education. *Learning and Individual Differences.* Volume 103, DOI:
10.1016/j.lindif.2023.102274

Kumar, M., Mani, U. A., Tripathi, P., Saalim, M., and Roy, S. (2023). Artificial Hallucinations by
Google Bard: Think Before You Leap. *Cureus*, 15(8)

Meyer, J., Urbanowicz, R., Martin, P., O'Connor, K., Li, R., Peng, P.-C., Bright, T., Tatonetti, N.,
Won, K., Gonzalez-Hernandez, G., and Moore, J. (2023). *ChatGPT and large language
models in academia: opportunities and challenges.* BioData Mining.
https://biodatamining.biomedcentral.com/articles/10.1186/s13040-023-00339-9

Microsoft Worklab, (n.d.). *What we mean when we say AI is "Usefully Wrong."* Microsoft
WorkLab Newsletter. https://www.microsoft.com/en-us/worklab/what-we-mean-
when-we-say-ai-is-usefully-wrong

Mollick, E. and Mollick, L. (2023, September 25). *Student Use Cases for AI.* Harvard Business
School Publishing. https://hbsp.harvard.edu/inspiring-minds/student-use-cases-for-ai

Myers, M. (2013). Qualitative Research in Business & Management 2nd edition, London, *Sage*.

University of Washington (2024, January 8). *Sample syllabus statements regarding student use
of artificial intelligence.* Teaching at UW. https://teaching.washington.edu/course-
design/chatgpt/sample-ai-syllabus-statements/

Wilson, M. (2024, February 25). *ChatGPT explained – everything you need to know about the AI
chatbot.* TechRadar. https://www.techradar.com/news/chatgpt-explained

APPENDIX A: AI Use – Six Question Survey

Thinking of your use of AI both for the term report for ISOM 201 and previous use of AI (such as ChatGPT) answer the below questions to the best of your ability.

1. Did you use AI before the assignment? If yes, how often? List some of the uses of AI that you have performed.

2. What were your expectations of using AI?

3. Were your expectations different than what you found using AI?

4. Did you find that the writing of AI was better than that was created by the group? *Note: Students worked either individually or in small groups (i.e., 2 to 4 students).*

5. Did you find that there were errors or facts that were not correct in the AI? How hard was it to fact check the work of the AI?

6. If you had to fact check every time you used AI, would you be willing to do the work? If so, how much time would you be willing to spend on this?