Affordances of Large Language Models in Design Activity

David Prohofsky, South Dakota School of Mines and Technology
Dr. Micah Lande, South Dakota School of Mines and Technology

Micah Lande, PhD is an Assistant Professor and E.R. Stensaas Chair for Engineering Education in the Department of Mechanical Engineering at the South Dakota School of Mines & Technology. Dr. Lande directs the Holistic Engineering Lab & Observatory. He teaches human-centered engineering design, design thinking, and design innovation courses. Dr. Lande researches how technical and non-technical people learn and apply design thinking and making processes to their work. He is interested in the intersection of designerly epistemic identities and vocational pathways. Dr. Lande received his B.S. in Engineering (Product Design), M.A. in Education (Learning, Design and Technology) and Ph.D. in Mechanical Engineering (Design Education) from Stanford University.

Affordances of Large Language Models in Design Activity

Abstract

Large language models and AI tools such as ChatGPT have possible benefits within design process and design activity across design courses in higher education. With the advent and rise in use of large language models (LLM's) we seek to better understand the effects of these tools. This technology has widespread application throughout the STEM field from writing/correcting code in any language to brainstorming ideas for the next big project, or even producing fully written research papers (although not this one). However, LLM's are not well understood and while many students hail them as a quick way to finish homework assignments, there are perceptions from students for both the positive and negative roles there may be in learning and understanding core concepts.

This paper investigates the use of this technology and implications in and for the design process. In particular, large language models may be utilized to aid in generative forms of brainstorming as well as convergent synthesis of ideas, adding to its value in supporting design work. Additionally, the viewpoints of students who utilize LLM's and those who do not will be explored to see how they use LLM's to aid them. Policies pertaining to LLM's will be discussed to see if they align with student uses and what limitations are set. This includes interviews with students about how LLM's are commonly used.

While many people in academia are working on how to limit the use of these tools as they can be seen as either 'cheating' or a 'short cut' this paper aims to investigate the possible benefits of this technology in the design process as well as to imagine how we might live with its effects in the future. We aim to fill this gap and discuss ways to improve the usefulness of LLM's as well as to successfully integrate it into project expectations, as we consider the future of design learning and practice.

Introduction

This is a paper examining large language models and design implications as an *Intersection of Design and "X" research paper*.

A common plot in science fiction writing is a scenario where computers can act like humans, talking and performing tasks as only a human could do. This outlandish notion took a leap to becoming a reality in 2018 when ChatGPT-1 was released by the tech company OpenAI with a program that was able to autonomously answer text prompts without the need for preprogrammed responses. ChatGPT-1 was trained on a data set of 117 million parameters which allowed it to understand human language and answer questions. This offered basic functionality but was very limited in its use cases as the program was unreliable and highly inaccurate in its responses. Since then, the company has exponentially increased the functionality of this program with the most recent release of ChatGPT-4 [1]; trained with 1.76 trillion parameters (15,000% more parameters than ChatGPT-1) ChatGPT-4 offers massively improved functionality and is now able to perform language-based tasks at or near a human level or above [1]. Functionality

includes writing full papers, solve mathematical problems, processing complex problems, understanding images as an input, and more. For easy comparison, standardized tests such as AP exams and the LSAT were taken by ChatGPT-4 with it scoring above the average human on most (one especially good result is the Standardized Bar Exam with a score in the top 10% of test takers) [1].

While this prospect is scary on some levels, it also can provide some assistance in generative activities such as brainstorming and concept development across design activity. This has some potential to be helpful and beneficial to supporting design activity, but gives rise to some concern about how one might use this new technology.

LLM may be able to provide a grounding to the engineering design process, particularly helpful to generative steps like concept development or fashioning a list of questions to ask users. Having a person or design team to apply their judgment to select among options may be a necessary step to keep the humanity in engineering design activity.

Large Language Models

ChatGPT or "Generative Pre-trained Transformer" is an artificial intelligence (AI) based 'Large Language Model' (LLM) that through a pre-trained data set can interpret the user's input to answer prompts [1]. ChatGPT is just one of the available LLM's available to the general public with many other models offering similar features. Other popular options include PaLM 2 (also known as Bard) by Google, Llama 2 by Meta, and Microsoft Copilot (a modified form of GPT-3 optimized for coding). All these LLM's work in the same way with the primary difference being training data and what functions its optimized/tuned for. As such this paper will be focusing on the most popular of these at the time of writing which is ChatGPT-3.5 (as per student interviews).

With the great strides this type of software has made in recent years, it has become increasingly more useful for students to use these tools to generate ideas for projects as well as to filter existing ideas. As technology becomes more popular with students a deeper understanding of it becomes increasingly more necessary. This paper will outline common uses for the technology through semi-formal interviews with students as well as a discussion of the current capabilities and limitations for use in STEM education and how it may aid in projects.

Research Methods

As the main goal of the interviews was to find student's opinionated views on the technology as well as individual use cases a semi-structured interview was selected. This interview type aligned with the primary purpose stated above as well as preventing narrowing of student's ideas by overly restrictive questioning. Additionally, interview questions where left open-ended with the interviewer encouraging the participants to discuss their thoughts further to promote the participant's own views on the topic. Students were selected out of those who regularly use LLM's, and selections were also limited to mechanical engineering students and computer science students. These limitations were imposed because both of these degrees involved a large number of projects where a similar methodology is used in idea generation and selection.

Additionally, students who commonly used Chat-GPT or were very familiar with its use were selected as their personal experience allows for a firsthand account of the use cases of this software.

In the end, six students were selected for interviews with two mechanical engineering students and two computer science majors questioned. These students were all upper classmen who had experience working on these types of projects using LLM's as well as work on previous projects before the popularization of this software. This higher level of experience allowed for comparing and contrasting the different methods used for project idea generation and selection.

One additional mechanical engineering professor was interviewed. This professor was selected due to his interest in and work with mechanical engineering projects. The same basic interview questions were used however further discussion was conducted with the professor to further ascertain their views on this subject with both student use and their personal use. The responses given were lumped in with the student responses and further discussed in a later paragraph. The interview questions in Table 1 were created to get a general idea of the student's thoughts on the software. Additional questions were added to further probe into the student's views on current and future uses for this type of technology tools, as applied to engineering projects among additional contexts.

Table 1: Interview Questions

What is your major?
How often do you use LLM's?
What is your primary use of LLM's?
What are the primary challenges for you in class projects?
In what areas do you think LLM's could aid you in engineering projects?
How do you currently use them and ideally how might you want to use them?
What barriers prevent this ideal use case and where might you want to see LLM's expanded upon in the future?
What process do you currently employ to generate ideas for projects?
How may this be improved using AI tools?
How do you narrow down your ideas to come to a final design decision?
Is this a situation that you might use AI as well?
Do you consider the information provided by LLM's to be trustworthy?

Results

The comments provided by the students are summarized through tables 2 and 3 with positive and negative aspects being separated out into the different tables. These tables were left generalized with further explanation below.

Table 2: Student Responses for common use cases

Student Responses	Interest in use case
Debugging code	High
Generating plans/guides	Medium
Generating lists	Medium
Writing simple emails/documentation	Low
Generating discussion questions	Low
Used similar to google.com however allows for more specific/tailored questions and answers	Low
Finding external resources to learn topics	Medium
Clarification of vague or obscure standards (such as ASME or laws)	Low
Changing the tone of the written text	Medium
“getting the ball rolling” generating initial ideas	Low
Using Chat-GPT to give what needs background research	Low
Grammar check/rewording	High
Summarizing content for easier consumption	Medium

Table 3: Student Responses desired improvements/short comings

Inconsistencies or errors in math and code generation
Improvements in prompt recognition/response to better answer questions
Lacks varied opinions as other group members would have
Ability to improve upon your work
Ability to adapt to the writer's ‘voice’ and mimic it in its writing

Student responses explained

Students traditionally generated ideas for projects by brainstorming a wide range of thoughts in a list and then narrowing them down by both practicality and interest in the ideas. Students also talked about how they preferred to perform the previous steps in small groups to incorporate divergent opinions to come out with a better overall final product.

The common theme students identified for using LLM’s was generating lists of ideas to be sorted by the user. Students pointed out that due to the software being very quick and its ability to generate simple ideas that can effectively cover the requirements of the prompt this makes it particularly useful for the generation stage of brainstorming. Students did not however commonly use LLM’s in the convergent aspect of projects preferring to select these themselves.

Faculty response detailed

The interviewed faculty had a large amount of interest in this technology as they focused on the student design side of engineering. This faculty member conducts multiple class projects where the use of this software would prove to vary the process or parameters of the student's work as well as the grading of these projects. They discussed how in their personal use LLM’s often proved useful for editing and increasing the speed at which they were able to compose emails and supplementary coursework for students.

The Professor expressed caution about the overuse of this technology specifically pertaining to false/miss applied information in written work. While he was excited for its use in aiding the writing process for reports and documentation.

Additional Methods

Additionally, further use cases were discovered through traditional methods of research and the use of the "chat" tool in ChatGPT 3.5.

The researchers used ChatGPT's own functionality to uncover further use cases through a variety of prompts as listed below. The responses listed are in the same format as listed by the LLM's output however each bullet point was further explained by the program which was truncated by the authors.

Table 3: ChatGPT Queries and Summarized Responses

Question:	Summarized Response:
How can ChatGPT be used to improve effectivity in STEM-based projects at a university level?	1. Research Assistance 2. Programming Help 3. Idea Generation 4. Drafting and Editing Papers 5. Conceptual Understanding 6. Learning and Explaining 7. Customization through fine-tuning 8. Collaborative problem solving
What use cases are expanded upon by the 4.0 update?	1. Better handling of instructions 2. Reduced biases 3. Fine-tuning for specific use cases 4. Advanced Control [over the model's behavior]
What are the potential effects of LLM's and the use of LLM's on STEM University Education?	Positive Effects: 1. Quick access to information 2. Programming assistance 3. Conceptual understanding 4. Idea generation 5. 24/7 Availability Potential Challenges: 1. Quality of information 2. Over reliance on technology 3. Lack of practical application 4. Ethical considerations 5. Bias and limitations 6. Customization challenges

Of addition interest is the disclaimer added by the ChatGPT at the bottom of the 1st question: "*While ChatGPT is a powerful resource, it's essential to validate information obtained from it and use it as a complement to traditional research methods. Additionally, being aware of the model's limitations and biases is crucial when applying it to STEM projects in an educational context."*

This matches and improves on the official disclaimer at the bottom of the program (written by OpenAI) *"ChatGPT can make mistakes. Consider checking important information."*
This statement implies that the program is not considered a trusted source such as traditional research sources and needs the implementation of further fact-checking methods by the operator. The author will further discuss this statement below as similar sentiments were voiced by other researchers and users of the program.

Findings

Common Uses and Benefits

The most common uses for LLM's as given by university students can be characterized into 3 categories:

Generation of new information/ideas based on a given users prompt

This use case is unique to LLM's as a technology. Many students are turning to LLM's due to the speed and ease of generating many ideas. This effectively augments or even replacing traditional brainstorming methods. The software is able to generate a large list of ideas based on the user's prompt and the user is able to improve/build upon these through additional prompts. While it excels at generation, narrowing down of ideas is often done by the user as they are choosing based on their own interest in the topic. Students said they preferred selecting ideas from the list generated by hand as the software does not have the same interests and opinions on what makes the topic interesting. Due to this personalization of this process, it is unlikely that a tool can replace it.

Searching for or clarifying existing information for the user

This case is quite similar to existing technology such as Google.com or other popular search engines. However, the change is analogous to switching from a library to the internet in terms of how detailed, quick, and specific the information retrieved can be. LLM's are able to tailor the answers far more specifically to the user's description than a standard search engine is able to do. It's ability to extrapolate data allows searching for specific use cases that may not have resources attached to it already. Alternatively, the user can quickly find many existing external sources through the software. By asking it to cite the information or recommend resources the user can find more reliable sources quickly as opposed to looking through a database for the specific source to match your topic or question.

Creation of technical documentation such as writing code or emails

This form of generation is mainly used by students to either edit existing text and code to achieve the desired response and level of polish or to generate a rough draft version of it for the user to refine. Rarely, is this feature used to create complete works instead its used to accelerate the existing process of creating the blocks of text or code where the user only copies the parts they like. Whether it's used for the initial generation or refinement of the user's text seems to depend on the intent for the content and to make up for weaknesses in the students' skillset [2].

Problems and Issues

Problems with the software as identified by students that prevented them from fully utilizing the functionality of the LLM include inherent biases, false information, and limitations to calculations.

Inherent Biases

Inherent biases were not brought up by students in interviews, but this may further illustrate why this is an issue. LLM's contain an inherent bias based on the data training set that was used, this is reflected in the results or writing created by it. "ChatGPT is known to perpetuate stereotypes such as nurses being female and doctors being male…" [2], many of these biases are included in human writing which is then reflected by the program however the identifiable source of these biases are lost when in this form making it harder to identify. While many of the other problems can be solved through increasing the data set of the AI model, this problem will have to be carefully considered by the AI companies if it can be solved at all.

False Information

'Hallucination' or falsely presenting information can be an issue. While the software excels at the generation of documents it is prone to falsely presenting information or making it up [3]. This phenomenon is not currently well understood as to the cause however the effect is it requires a higher level of supervision from the user to catch these additional error's that are introduced by the program. This effect is currently being minimized as the technology advances with GPT-4 showing "large improvements over GPT-3.5… GPT-4 resists selecting common sayings… however it still can miss subtle details," [1]. These improvements are predicted scale with the model and will likely become less of an issue in the future.

Limit to Calculations

Another issue is an inability to correctly do complex math functions. While ChatGPT-3.5 was able to explain how to solve math problems it struggled to correctly execute this process and directly solve them. This results in many incorrect answers furthering the issues of LLM's being 'confidently incorrect'. GPT-4.0 has improved its ability to solve this type of problem through a method of double-checking results however it is still not foolproof. GPT-4's improvements are shown through simulated test scores of the AP Calculus BC test where GPT-4 scored 43-59th percentile instead of the 0-7th percentile that GPT-3.5 showed. This is likely to continue to improve in future iterations of the software and will likely become less of a problem for users.

Future Considerations

There are two types of knowledge groups that need to be addressed for further understanding of the usefulness of LLM's. Declarative and procedural (interpretive) knowledge where declarative can be aided the most by Chat-GPT. Declarative knowledge is the simplest of the two types and refers to facts that can be stated and memorized. LLM's are able to aid in the user through its vast database of pre-learned information that can then be drawn upon and stated for the user. Due

to the factual nature of declarative knowledge, a user can look up this information (in many cases) either through a reference book, search engines, or more quickly through LLM's. Procedural knowledge provides more complications for LLM's as it requires "commonsense knowledge about the real world… [such as] procedures or sequences of actions for achieving [a] particular goals," [4]. While procedures can be generated by the program and even refined through further prompts the user must be able to use their own knowledge to fully understand and utilize these printed instructions. As such LLM's have limited use cases in the type of advantages they offer for users.

To make use of the above knowledge types requires different types of expertise that the user must possess. There are two primary types of expertise that LLM's are able provide to the user. Expertise can be split up into two primary groups, routine, and adaptive expertise. Routine expertise is the ability of a person to perform their standard function and job, while adaptive expertise is that person's ability to "…deal effectively with novel situations and problems" [5]. Adaptive expertise is becoming increasingly important in the workspace as jobs require more information from outside of one's domain of work [5], this is an area where LLM's are able to supplement the information known by the user to allow them to quickly work outside of their area of expertise. LLM's are especially well suited for this case as they contain a large body of information in every field of study and is able to place this into the context of the user's routine knowledge base. This does pose an additional challenge as the user is no longer able to identify when the information given is correct.

Conclusions and Implications

LLM's such as Chat-GPT in its current stage is analogous to early stages of Wikipedia. Many of the same concerns were raised as Wikipedia became popular including worries of accuracy, biases, and lack of reliable sources. However, as Wikipedia has matured many of these issues have been solved resulting in it becoming a far more respected source. LLM's will likely follow a similar path becoming more popular, a more reliable source of information, and being able to more reliably generate new ideas for the user following their intent.

Not all the problems in the use of this software are possible to fix as every tool will have inherent issues. Due to the power of these tools and the black box nature of their functionality, it becomes more important that the user have the ability and knowledge to effectively utilize this. It has become important for schools to teach students (and educators) how to use LLM's as well as the inherent flaws and biases discussed above [6].

This tool is incredibly powerful and able to affect many aspects of life. While many may seek to prevent the spread of this it is likely that the LLM's will only become more popular and further integrated into everyday life. We must learn how to properly use these tools as well as how to navigate an educational and work environment where LLM's will be used.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant number #2010696. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

References

1. OpenAI et al., "GPT-4 Technical Report," arXiv, Mar. 2023. [Online]. Available: http://arxiv.org/abs/2303.08774.

2. J. G. Meyer et al., "ChatGPT and large language models in academia: opportunities and challenges," *BioData Mining*, vol. 16, no. 1, p. 20, 2023. [Online]. Available: https://doi.org/10.1186/s13040-023-00339-9.

3. B. A. Huberman and S. Mukherjee, "Hallucinations and Emergence in Large Language Models," SSRN, Sep. 18, 2023. [Online]. Available: https://doi.org/10.2139/ssrn.4676180.

4. M. P. Georgeff and A. L. Lansky, "Procedural knowledge," *Proceedings of the IEEE*, vol. 74, no. 10, pp. 1383-1398, Oct. 1986. [Online]. Available: https://doi.org/10.1109/PROC.1986.13639.

5. K. Bohle Carbonell et al., "How experts deal with novel situations: A review of adaptive expertise," *Educational Research Review*, vol. 12, pp. 14–29, 2014. [Online]. Available: https://doi.org/10.1016/j.edurev.2014.03.001.

6. M. A. Ayanwale et al., "Teachers' readiness and intention to teach artificial intelligence in Schools," *Computers and Education: Artificial Intelligence*, vol. 3, 2022, Art. no. 100099. [Online]. Available: https://doi.org/10.1016/j.caeai.2022.100099.