Capstone Design
CONFERENCE 2024
JUNE 3–5, 2024 ▶ KNOXVILLE, TENNESSEE

How Would a Chatbot Fare in Capstone?

Elizabeth A. DeBartolo
Rochester Institute of Technology

The availability of online file repositories, solution manuals, and homework "help" has forced professors to think carefully about the assignments they give and how they assess their students' work. Now with the widespread availability of online chatbots and image generators, the problem is magnified. Capstone instructors, whose students are always working on new problems that do not come from textbooks, have not needed to worry about copies of tests and homework answer keys available online, but could an online AI be used to "cheat" in a capstone design course? This paper examines a sample design project, specifically focused on the act of problem definition and basic project management.

Keywords: ChatGPT, AI, academic honesty, assessment

Corresponding Author: Elizabeth A. DeBartolo, eademe@rit.edu

Introduction

Easy access to internet resources, such as file shares, and homework help and solution manuals from services such as Chegg have led to an ongoing fight by educators to stay one step ahead and make sure their students are learning content – as opposed to learning how to use online cheating tools. For example, Broemer & Recktenwald published an in-depth investigation of the use of Chegg during a two-hour exam with data from Chegg showing extensive use of the service to post and look for answers to exam questions.[1] This was paired with suggestions for how to reduce the use of Chegg during exams. DeGoede compared the relationship between homework and test grades in the pre- and post-Chegg eras, with homework averages increasing but test scores remaining the same.[2] The recommendation from this paper was more robust: "*The goal of this discussion-based HW model was not to repair the broken correlation between HW and exams, but rather to increase engagement with the material and shift the focus of the student work from getting the correct answer on the problems to developing skills*."

As capstone instructors, we confidently sat back through all of this and told our students: "Go ahead, use what's available from non-pirated sources (just cite your work)! Our class is immune to these threats because our projects are unique, real-world problems to be solved! Borrow inspiration from elsewhere! Reference industry standards that practically write your test requirements!" We maintained that the value proposition of the design engineer was the ability to identify the problem to be solved, creatively develop design concepts, and critically evaluate the solutions available. The emergence of free Generative Pre-trained Transformers (GPTs), based on large language models (LLMs), could change all that. This paper focuses specifically on ChatGPT3.5 [3], which uses information from January 2022 and earlier.

Although ChatGPT was only released to the public in November 2022, several studies have already evaluated its ability to complete a variety of different tasks.[4–6] Notably, a recent literature search collected the results of studies across different domains and reported "critical and higher order thinking" as the highest-scoring performance area. Engineering was not evaluated, but programming scored well ("outstanding to satisfactory") and mathematics and software testing scored poorly ("unsatisfactory").[4] The tool was labeled potentially useful for summarizing information and answering questions, but was flagged for generating incorrect information, using biased data to form responses, and lacking the most up to date knowledge. A study focused specifically on programming found that it provides functional code anywhere from 75% to 95% of the time, and can be a useful tool in optimizing code or fixing known errors.[5] ChatGPT has also taken the Fundamentals of Engineering (FE) exam for Environmental Engineering.[6] While it is difficult to determine whether the AI would have passed the exam, ChatGPT3.5 answered more than half of the questions correctly across the entire test: 58.95%, increasing to 62.68% with refined prompts. For a tool that is actively being trained and improved, this is an impressive showing.

In this paper, a capstone design project submission is submitted to ChatGPT3.5 with a series of prompts to step through the early stages of the design process to assess (1) whether a student could use a chatbot to "cheat" in a capstone design course and (2) whether ChatGPT3.5 could provide targeted (and legitimate) aid for students taking a capstone course.

The Problem

A client submitted a non-confidential project proposal to our system, asking for the following:

> "We would like an aquatic wheelchair that will allow for accessibility for individuals that require trunk control and have significant physical limitations. Design a chair with a PVC base, able to be totally submerged and water proof with counter balancing for anti tipping; a seat with waterproof mesh material that can conform to an individual body as opposed to a hard plastic 90° seat; capability to adjust the recline of the back of the seat to up to 120°; adjustable foot and leg support to allow for 90° bend in the knees or to support the legs/feet off the ground; have 2 different adjustment points-one for the back rest and one for the foot/leg support; a butterfly or H shaped harness that zips in the front like a vest and supports the trunk and a seatbelt around the waist. All materials need to be completely waterproof. Weight limit up to 250 lbs."

This is the information initially available to students, so the experiment starts here.

Methodology

ChatGPT works in a conversational format, where the user enters prompts, and ChatGPT provides responses. Each chat has the flow of a conversation, so later prompts can reference earlier prompts and responses, allowing the user to refine prompts where desired. The prompts in this case were based on the framework we provide for our students in the course. The students in our two-semester multidisciplinary course follow a system design process, beginning with problem definition, then system level design, and detailed design. In the second semester, teams start with component- and system-level building and testing, and progress to system integration. For the purposes of this experiment, only the Problem Definition phase was examined in detail; cursory work for the design and analysis phases will also be presented. Prompts were refined only if the response was of low enough quality that the response would have been flagged by an instructor.

The Prompts: Defining the Problem

Based on assignments, the following prompts were used:

- <u>Methods</u>: "What methods might I use to define this problem. Write in the voice of a student learning about the design process. For example, would you conduct interviews or focus groups with particular people?"
- <u>Interview the customer:</u> "A customer has requested the following: [pasted from proposal] What clarifying questions might I ask them to better define the problem?"
- <u>Clarify the problem and write customer requirements</u> [information provided by the customer in a phone call to the course instructor]: "The wheelchair will be used in a zero-entry therapy pool, with a focus on children under the age of 18. The harness provides trunk support for children with physical disabilities. Given this, write a preliminary set of customer requirements and a set of measurable and testable engineering requirements."
- <u>Write engineering requirements</u>: "Convert these customer requirements to a set of measurable and testable engineering requirements."
- <u>Standards</u>: "What relevant industry standards should I consider in both the design and testing of this wheelchair?"
- <u>Constraints</u>: "List 4-5 constraints for this project"
- <u>Use Cases</u>: "What are some sample use cases for such a wheelchair?"
- <u>Initial risk assessment</u>: "What are the highest risks for this project?"
- <u>Preliminary project plan</u>: "Write a list of tasks required to design this product. Aim for a list of about 20-30 tasks, suitable for a team of 4 engineers."
- <u>Problem Statement</u>: "Write a 1-paragraph problem statement for this project. It should include current state, desired state, key goals and deliverables, and key constraints."

Refined Prompts

In some cases, the prompts lacked enough specificity to give useful answers and required refinement:

- <u>Customer requirements</u>: "Rewrite the Customer Requirements listed above, replacing priority words such as "must" and should" with importance ranks 1, 3, and 9, where 9 is most important, 3 is less important, and 1 isn't important but would be nice to have. At most, half of the requirements can be ranked 9"
- <u>Engineering requirements</u>: "For each of these engineering requirements, what is the unit of measure, what is the marginally acceptable value, the ideal value, and which direction should the value be driven. Provide the results in a format that can be copied and pasted into an Excel spreadsheet, one requirement per row"
- <u>Use Case</u>: "Rewrite the first use case listed above (Aquatic Therapy Sessions: The wheelchair can be used during aquatic therapy sessions for children with physical disabilities, providing buoyancy and support as therapists guide them through exercises in the water) so they can be laid out in a typical swimlane diagram. Include the actions of relevant stakeholders and a sequence of events for the use case. Format this so that it can be copied and pasted into an excel spreadsheet, with each actor and all of that actor's actions in a single row. The columns represent time, so that I can read the

table left to right and get a sense of the sequence of events."

- Project plan: "Modify this schedule if the team is following a systems engineering process"

Results

The grading system used in our class is a holistic evaluation of student work by phase, where A-level is for high quality work, supported by evidence and critical thinking, B-level is for work that is generally good but lacking in insight, and C-level is for work that has significant gaps and haphazard use of design tools. We do not typically grade individual artifacts, so the grades reported here are estimates.

Methods: A-
The list of suggestions included diverse and relevant stakeholders in diverse and relevant situations. For example, "observing children during actual aquatic therapy sessions to understand their movements, interactions with existing equipment, and the challenges they face" and "organizing focus groups with a diverse set of stakeholders, including children with different types of physical disabilities, parents, caregivers, therapists, and aquatic therapy specialists" would both provide valuable insights.

Interview the customer: B+
The list of questions for the customer was thorough and even included some questions about how "firm" some requirements were, such as the 250 lb capacity. Some questions built on solutions presented by the customer ("are there any preferences for the type of waterproof mesh material?") when it may have been better to question whether those solutions were truly necessary. For a team struggling to get started thinking about what to ask, this would provide a good start.

Write customer requirements: C
Initial customer requirements were poor, with nearly everything listed as the highest importance. Most implied priority with ("must" and "should" statements). The requirements were simple reframing of the input problem and did not show any insight into the project. Refining the prompt improved them slightly, but a team relying on these requirements would potentially miss features important to the customer and prioritize needs incorrectly. If the prompts had been augmented with data from customer interviews, the quality should improve.

Write engineering requirements: C+
These were similarly superficial, but each included a one-sentence "testing" bullet. Refining the prompt resulted in about half of the requirements including units of measure and ideal/target values. For example, "Engineering Requirement: The wheelchair shall support a static weight of up to 250 lbs without permanent deformation or structural failure. // Testing: Apply static loads incrementally to the wheelchair, measuring and assessing any deformations or failures." Like customer requirements, these could have been improved by augmenting inputs with customer interview information.

Standards: A
The list of standards was far more thorough than what most student teams can identify. This is a useful way for teams to conversationally identify relevant standards.

Constraints: B-
Three of the five constraints identified were as expected: time, budget, and regulatory compliance. The other two, accessibility & inclusivity and usability & user-friendliness, aren't true constraints that limit the set of design options available.

Use Cases: B-
The extensive list of use cases identified covered many that were relevant and some that were not ("Independence and Mobility Training: Using the wheelchair in water allows children to work on independence and mobility skills, encouraging them to explore movement in a controlled and supportive setting"). A refined prompt led to identification of specific actors and actions, properly formatted as a partially accurate swimlane-style diagram (Table 1).

Table 1: Sample Use Case

Time	Child w/Physical Disabilities	Caregiver	Therapist
1	Positioned at the poolside		
2	Transferred to aquatic wheelchair	Assisted	
3		Adjusted	
4			Engaged
5			Guided
6		Observed	
7	Completed therapy session	Assisted	
8	Exited pool and returned	Assisted	
9			Provided feedback

Initial risk assessment: B+
The preliminary list of risks was surprisingly thorough, and even included mitigation strategies without being prompted. The risk list included some that students tend to forget about early on (e.g., user acceptance and

manufacturability). The mitigation strategies were sometimes useful ("Involve potential users, caregivers, and healthcare professionals in the design process through user trials and feedback sessions.") and sometimes not ("Conduct rigorous testing of all mechanical components, including repeated adjustment cycles and load testing, to ensure reliability and longevity" is a reactive strategy rather than proactive).

Preliminary project plan: B
The initial project plan was far too broad, including tasks from research through product launch. Refining the prompt to focus on minimum viable product and use a systems design process yielded a meaningful list of tasks, although they were not granular enough to help a team on a day-to-day basis. The assignment to team members appeared to be random. For a team struggling to get started with their project plan, this may give them a start.

Problem Statement: A
This summary of information into a given format was well written and addressed all required points.

Beyond Problem Definition

A cursory exploration of work beyond the problem definition stage was done, but generally yielded mediocre results. ChatGPT.35 was able to suggest a variety of solutions to a specific prompt. For example, when asked "What mechanism might be used to enable adjustment of the foot and leg support", 10 options were provided, all of them generic means of making mechanical position adjustments, although some (scissor lift, electronic linear actuator) would not be feasible in this situation. Our students are instructed to favor quantity over quality for initial brainstorming, so the responses were acceptable.

Questions about which parts of the system would experience the highest stress require engineering insight to answer. The response, "In the design of the aquatic wheelchair for therapy pool use, one part that is likely to be subjected to the highest stresses is the adjustable foot and leg support mechanism," identified a relatively low-stress subsystem of the project.

Discussion

Overall, the performance of ChatGPT3.5 was about equivalent to what we would expect from an inexperienced student, or one who did not pay close attention to instructional modules preparing them to complete this work. A student could not rely entirely on this chatbot to pass the capstone course, because our team projects require students to collaborate in real time. A student not contributing to discussion and only sharing text-based work would quickly get flagged by their project advisor or teammates as a low contributor.

For a student intimidated by the open-ended, ill-defined nature of a capstone project, ChatGPT3.5 could be a useful tool in helping to combat the fear of a blank page. The interconnected nature of the various deliverables required in a systems design problem means information must be interpreted and repackaged (e.g. translating customer to engineering requirements, identifying risks and tasks related to the problem). Summarizing information is a task that chatbots tend to do well,[4] so this tool may help students to connect seemingly disparate pieces of information they gather.

This tool is continually "learning", so prompts that give unsatisfactory responses today may yield better answers tomorrow, or next semester, or next year. Engineers using ChatGPT3.5 must apply critical thinking to review and improve on the responses to prompts, and must be aware of confidentiality limits, but this can still be a useful tool for preliminary or exploratory work. Capstone instructors would do well to embrace, rather than ban its use.

Conclusions

A student relying on ChatGPT3.5 in our capstone class would most likely be able to contribute to team discussions using AI-generated responses. Their work on the individual assignments showing contributions toward team deliverables would put them on a list of students to keep an eye on to make sure they were learning and practicing good design process, but they may benefit from the starting point it provides.

References

1. Broemer, E. & Recktenwald, G. Cheating and Chegg: a Retrospective. in *2021 ASEE Virtual Annual Conference Content Access Proceedings* (ASEE Conferences). doi:10.18260/1-2--36792.
2. DeGoede, K. A Chegg® Era Model for HW. in *2020 ASEE Virtual Annual Conference Content Access Proceedings* (ASEE Conferences). doi:10.18260/1-2--33979.
3. OpenAI. ChatGPT: Conversational AI Platform. https://chatgpt.com/ (2023).
4. Lo, C. K. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Educ Sci (Basel)* **13**, 410 (2023).
5. Rahman, M. M. & Watanobe, Y. ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Applied Sciences* **13**, 5783 (2023).
6. Pursnani, V., Sermet, Y., Kurt, M. & Demir, I. Performance of ChatGPT on the US fundamentals of engineering exam: Comprehensive assessment of proficiency and potential implications for professional environmental engineering practice. *Computers and Education: Artificial Intelligence* **5**, (2023).