Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025 ASEE
Paper ID #47068

BOARD # 322: An NSF FMRG Supported Exploratory Study of Prompting Large Language Models for a Conversational Manufacturing Education Platform

Fatemeh Karimi Kenari, University of North Carolina at Charlotte
yasaswi bhumireddy, University of North Carolina at Charlotte
Xiaoliang Yan, Georgia Institute of Technology
Dr. Mahmoud Dinar, University of North Carolina at Charlotte

Mahmoud Dinar is an assistant professor of mechanical engineering. His main research interest involves integrating AI in multimodal computational frameworks to understand and aid design for manufacturing. His ongoing projects are creating a conversational platform to evaluate and teach manufacturing skills to the future workforce, and solving geometric puzzles to design sustainable and resilient products as a 3D composition of pieces.

Shreyes N Melkote, Georgia Institute of Technology

An NSF FMRG Supported exploratory Study of Prompting Large Language Models for a Conversational Manufacturing Education Platform

Introduction and Background

Manufacturing education is a foundation for technological and economic progress, providing individuals with the expertise and skills essential for fostering innovation and productivity. While fundamental knowledge of manufacturing processes, materials, and machinery, and general problem-solving skills remain central, the methods of delivery should adapt. Additionally, the evaluation of the learners' knowledge is key. Lecture-based teaching continues to dominate, yet the rapid advancement of artificial intelligence (AI) has enabled transformative applications across diverse fields, including education and manufacturing. In manufacturing, the increasing complexity of systems and the demand for efficiency have driven the adoption of AI tools to optimize processes, enhance decision-making, and foster human-machine collaboration. Similarly, in education, AI has the potential to enhance teaching approaches, e.g., with intelligent tutoring systems (ITS) and personalized learning platforms. AI chatbots stand out as for their 24/7 availability, and enhancing engagement through conversation. However, they should be examined as a reliable education tool in manufacturing, especially in adapting to different users. Here we present and evaluate an LLM-powered chatbot—the Manufacturing Adviser—in answering various types of manufacturing questions to 4 user levels from children to experts.

ITS are known for personalized learning, enabling students to progress at their own pace while receiving feedback. VanLehn [1] presented a meta-analytic review comparing the effectiveness of human tutoring, intelligent tutoring systems (ITS), other computer-based tutoring systems, and no tutoring in facilitating student learning. He found that the effect size of human tutoring was comparable to that of ITS. The effectiveness of tutoring is closely tied to the ability to adapt feedback and scaffolding to individual learners, rather than the type of tutoring system itself.

AI-powered chatbots mimic one-on-one human tutoring, dynamically adapting to students' individual needs and engagement levels [2]. Wang et al. [3] examined ChatGPT's capabilities across six domains including design (functional, conceptual, technical) and manufacturing (processes, systems, and production management). They categorized questions by difficulty and assessed performance on correctness, relevance, clarity, and comparability. Answers to basic knowledge and knowledge application questions were suitable. Creative capabilities were found to be comparable to top engineering students but did not surpass experienced professionals. Yang et al [4] identified the advantages of implementing AI chatbots in education as enhancing student engagement through interactive simulations, reducing workload for administrative staff by automating routine tasks, and personalizing education for diverse user needs.

Despite AI-powered chatbots' potential, the development process remains challenging. The lack of accessible tools and streamlined frameworks has created a gap in the effective adoption of this technology [5]. Shahriar et al [6], explored the evolution, capabilities, and limitations of ChatGPT, the state-of-the-art AI chatbot by OpenAI. The authors call for enhanced model training, ethical guidelines, and improved transparency to address these issues, emphasizing ChatGPT's potential as a versatile tool with significant room for refinement and responsible use.

System Design and Architecture

We created the Manufacturing Adviser chatbot to address the need for on-demand responses to manufacturing-related queries for different levels of users' knowledge. It is a web-based platform storing data about manufacturing-related documents, questions, users, and a history of their conversations to track progress. The interface has user authentication, and level selection.

The key architectural components of the Adviser include an attention-based context analyzer [7] to maintain conversational continuity, and tailor responses based on the user's expertise level, similar to how an ITS tracks learners' progress, i.e., it adjusts future responses based on conversation history, and account for the user's existing knowledge. The Adviser also incorporates user-level personalization, dynamically adjusting language and the depth of information to align with different user levels. Additionally, Knowledge Retrieval Augmented Generation (RAG) [8] integrates knowledge retrieval from manufacturing documents with Large-Language-Model's generation capabilities (ChatGPT in this case) to provide contextually relevant responses. Manufacturing documents are divided into smaller chunks of 500 words. Each chunk is transformed into a numerical representation (embedding), capturing semantic information for similarity-based retrieval. Figure 1 shows the Adviser system's workflow. When users pose questions, the system transforms the user's query into a numerical form, generates a query embedding, matches it with stored document embeddings, and retrieves the most relevant chunks, which are ranked and assembled for response generation. Dynamic response generation utilizes OpenAI's language models to combine retrieved information with the user's expertise level and conversational history.

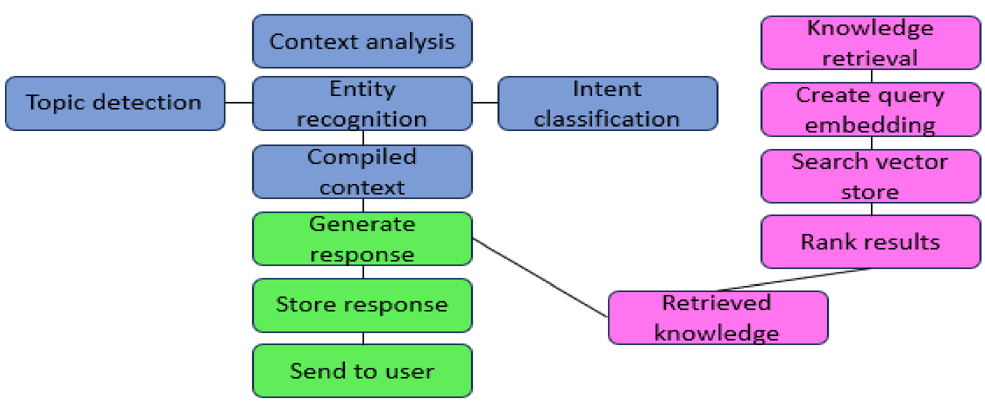

Figure 1: Backend System's Workflow

Methodology and Results

The evaluation of the Adviser focused on its ability to provide contextually appropriate responses across different levels of expertise. A set of 30 questions was selected from Kalpakian and Schmid's textbook on manufacturing processes [9]. These questions were divided into three categories: General Process Questions**:** Broad questions that provide an overview of manufacturing concepts and techniques. (e.g. What is casting?) Sub-Process Questions: Detailed questions that focus on specific stages within manufacturing processes. (e.g. How does ultrasonic machining differ from traditional machining?) Process Parameter Questions: Technical questions that delve into parameters such as temperature, pressure, and material properties. (e.g. What are the thermomechanical effects of hot rolling on alloyed steel?). We passed the questions with explicitly desired response levels (child, teenager, college student, or expert) to ChatGPT API.

The Adviser's performance was assessed for the 4 user levels quantitatively using cosine similarity between embedding vectors of responses and questions derived by OpenAI's text-embedding-ada-002 model. Chunk embeddings were independent of phrase length. The cosine angle of the vectors ranges between 1 (perfect alignment) and -1 (opposing meanings). Figure 2 shows the Average Cosine similarity per user level and cosine similarity between responses for different users. The left figure shows that Expert level has the highest cosine similarity between questions and responses. Child and Teen show lower similarity which is likely due to a narrower range of words for simpler and less nuanced answers to questions. The right table shows the pair-wise cosine similarity among the responses for different user levels. Responses for children and experts have the least similarity indicating that the Adviser generates responses with the most differences at the two ends of the spectrum. The highest similarity is between responses for teens and college students implying a middle ground for a baseline response.

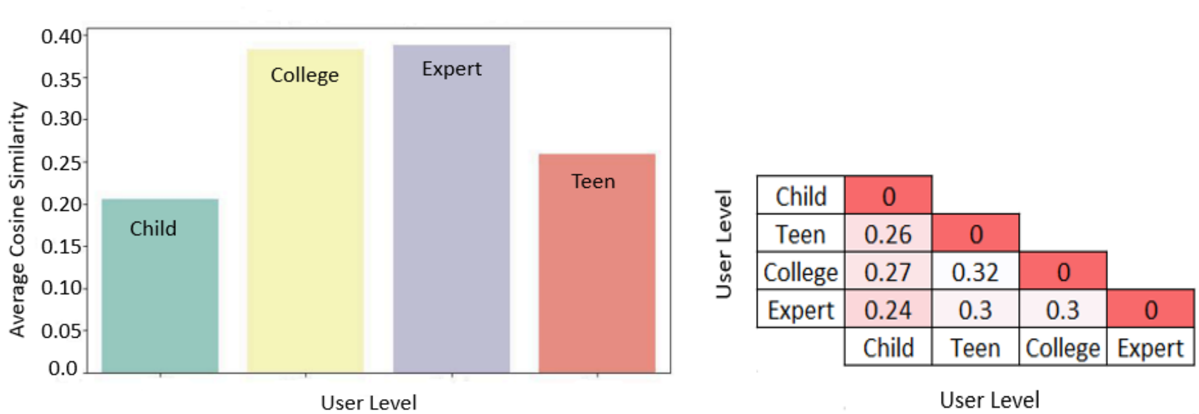

Figure 2: Cosine similarity per user level and cosine similarity between every two levels.

The Adviser's responses were evaluated qualitatively for correctness, relevance, and suitability for the user's level of understanding of manufacturing. One researcher manually searched for answers to the 30 questions using textbooks and papers to assess the Adviser's responses' relevance. The results of the evaluation revealed several insights into the Adviser's performance. The Adviser demonstrated an ability to generate contextually relevant responses. For example, responses for children were simplified and explanatory, using examples from day-to-day objects that children would see, while those for experts were detailed and technical. Figure 3 shows an example of an input question and the four levels of responses generated by the Adviser. As can be seen, the complexity and details of responses increase with the level of expertise of the user. This could be seen more in general process questions since the sub-process and process parameters questions are more suitable for college and expert level in general. The user prompt is also crucial to have a reliable answer for that level. For example, a response to a sub-process question such as “How do cooling rates affect the hardness of forged components?” does not have a simplified answer that is understandable for a child.

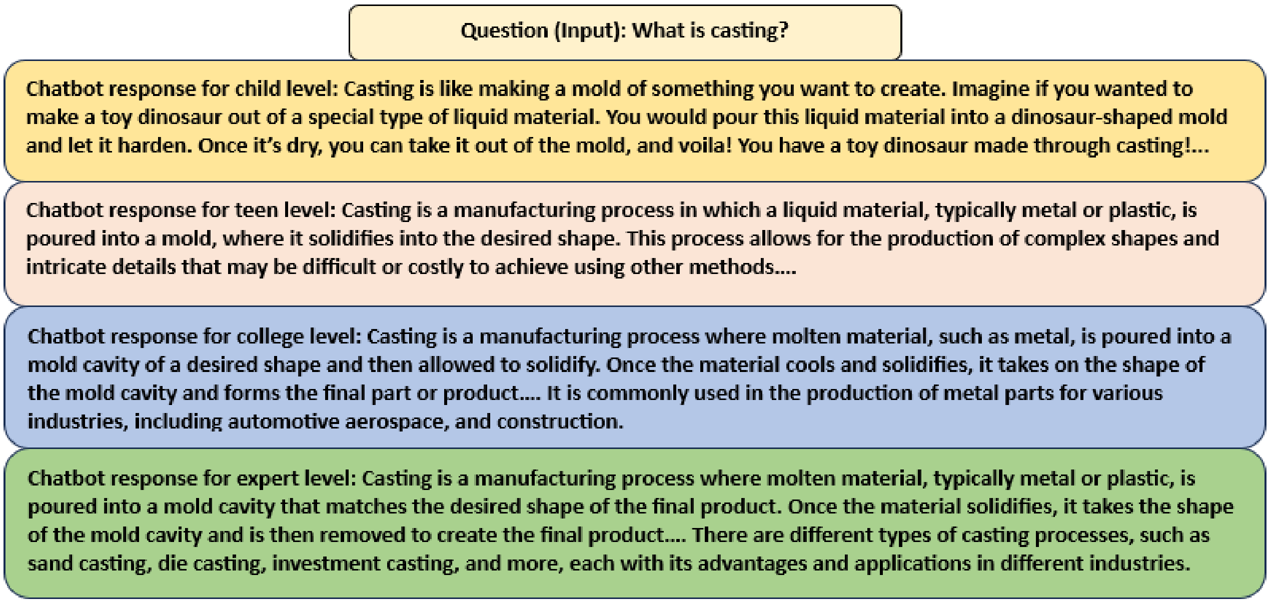

Figure 3: Highlights of responses of the Adviser for a question at four different levels.

Conclusion and Future Work

The advancement of new technologies, especially AI, has created opportunities in education. Employing this approach to educating the next generation, regarding the skill gap in manufacturing, as one of the important aspects of society and economics, is worth considering. In this paper, we conducted an exploratory study in creating a ChatGPT-powered Manufacturing Adviser and evaluated its capability in educating users of different knowledge levels. Our results indicate that the Adviser could provide contextually relevant responses to different question types in manufacturing, either general process questions or more detailed questions about process parameters, and change the answers based on the users’ levels. We used cosine similarity to measure the semantic overlap of the questions and their responses. Our results show the highest cosine similarity at the expert level due to the more detailed and wide range of used words at this level. Using the attention mechanism (storing user-specific context) helped tracking the user’s progress to base the future responses on what they already know. Early findings highlight the Adviser’s potential as a tool for manufacturing education. By using the Retrieval Augmented Generation, combining LLMs with manufacturing-specific knowledge bases, the Adviser provides context-relevant answers.

Despite its potential strength, the system faces challenges in adapting some responses to the user’s level due to the unsuitability of the prompt questions for that level. Future work could address these limitations by using other self-attention mechanisms like BERT for NLP tasks. Other LLMs such as Gemini and specialized Manufacturing LLMs could be explored for better accuracy in manufacturing queries. Incorporating visual language models (VLMs) will enhance interaction with the Adviser. For example, integrating CAD models and process diagrams could enable the Adviser to provide real-time feedback on manufacturability and design considerations. Furthermore, a reverse Q and A (asking users questions about manufacturing and evaluating their responses) would be suitable to assess the user's understanding level. For example, just because someone is at college does not mean that their manufacturing knowledge is aligned with their academic level. This feature could be more useful for beginners or children's level since they may not have enough knowledge to know what questions to ask the Adviser in the first place. Expanding question types can be an additional approach to enhance the usefulness of the Adviser, especially with questions related to design for manufacturing (DFM), i.e., modifications to improve manufacturability and selecting the optimal manufacturing process for a design.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. 2229260. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] K. VanLehn, "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems, " *Educational Psychologist,* vol.46, no. 4, pp. 197-221, 2.11.
[2] A. Celepcikay, and Y. Yildirim, "Artificial intelligence and machine learning applications in education," *Eurasian Journal of Higher Education*, vol. 2, no. 4, 2021.
[3] X. Wang, N. Anwer, Y. Dai and A. Liu, "ChatGPT for design, manufacturing, and education, "Proceedings of the 33rd CIRP Design Conference, 2023, pp. 7–14.
[4] S. Yang, and C. Evans, "Opportunities and challenges of using AI Advisers in higher education, " *Proceedings of the 7th International Conference on Information and Education Technology (ICIET 2019)*, 2020, pp. 79–83.
[5] D. Ramandanis, and S. Xinogalos, "Designing a Chatbot for contemporary education: A systematic literature review, " *Information, vol. 14, no. 503,2023.*
[6] S. Shahriar and K. Hayawi, "Let's have a chat! A conversation with ChatGPT: Technology, applications, and limitations, " Artificial Intelligence and Applications, pp. 1-5, Jun. 2023.
[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need, " *Proceedings of the 31st Conference on Neural Information Processing Systems,* 2017, Long Beach, CA, USA.
[8] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks, " *Proceedings of the 34th Conference on Neural Information Processing Systems*, 2020, Vancouver, Canada.
[9] S. Kalpakjian, and S. R. Schmid, *Manufacturing processes for engineering materials* (6th ed.), Pearson, 2016.