Toward LLM-Powered Robots In Engineering Education

C. Estelle Smith
Colorado School of Mines,
Department of Computer Science
Golden, Colorado, USA
estellesmith@mines.edu

Alemitu Bezabih
Colorado School of Mines,
Department of Computer Science
Golden, CO, USA
alemitubezabih@mines.edu

Shadi Nourriz
Colorado School of Mines,
Department of Computer Science
Golden, CO, USA
shadinourriz@mines.edu

Bo Wu
Colorado School of Mines,
Department of Computer Science
Golden, CO, USA
bwu@mines.edu

Gabe Fierro
Colorado School of Mines,
Department of Computer Science
Golden, CO, USA
gtfierro@mines.edu

ABSTRACT

We present initial results from two projects studying the emerging role of LLMs at a small, engineering-focused R1 university: (1) the development of an LLM-based virtual teaching assistant; and (2) a survey of all students at the university. Our virtual TA project suggests that custom implementations of course-specific LLM agents may be more desirable and effective for supporting student learning than general-purpose conversational agents such as ChatGPT. Our survey also provides evidence that despite some concerns about LLMs like ChatGPT, 36.8% of students had tried them at least a few times and 30.4% had already incorporated routine use into their academic pursuits. We provide a taxonomy of emergent LLM use cases, and show that the more students use LLMs, the more they are perceived as beneficial. We conclude by discussing implications of this work for HRI and LLM-powered robots in engineering education.

CCS CONCEPTS

• **Computer systems organization → Robotics**; • **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

Large language models, engineering, education, social robotics, human-robot interaction

ACM Reference Format:
C. Estelle Smith, Alemitu Bezabih, Shadi Nourriz, Bo Wu, and Gabe Fierro. 2024. Toward LLM-Powered Robots In Engineering Education. In *Proceedings of Companion of the ACM/IEEE International Conference on Human Robot Interaction (HRI) (HRI '24 Companion) (HRI '24 Companion).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

HRI '24 Companion, March 11–15, 2024, Boulder, Colorado

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The increasing accessibility of Large Language Models (LLMs) presents engineering students with transformative opportunities for learning and career development, offering a broad range of content generation capabilities outside traditional educational contexts. While previous educational tools like interactive tutoring systems (ITS) [2, 3, 10, 13] have catered to specific tasks within curated learning experiences, the general-purpose nature of LLMs allows for versatile and interactive content creation through natural language prompts. Concerns and opportunities have arisen regarding the potential impact of LLMs on traditional learning processes and career development, prompting the need for collaborative efforts to align AI development with modern pedagogical principles [5, 7, 9].

HRI has long considered prospective roles of social robots in education. Social robots have garnered attention for their ability to improve the educational experience, serving as tutors and peer learners, with the potential to become everyday educational tools [4, 11]. Importantly, their physical presence and interactive social capabilities offer advantages over UI/UX-based learning platforms [12]. Integrating social robots into education presents significant challenges in AI and robotics [4], including the development of specialized content for robot-based learning and the achievement of seamless social interaction, such as generating verbal responses. LLMs could be integrated within educational robots to support convincing natural language dialogues. However, researchers must first develop a clear understanding of concomitant use cases, interaction paradigms, educational settings, and opportunities *v.s.* risks.

During the Spring23 semester, we initiated two projects at our university: (1) the development of an LLM-based virtual teaching assistant (Sec. 2); and (2) a survey of all students that collected baseline information on student usage of LLMs as well as opinions and concerns on the role of LLMs in engineering education (Sec. 3). In this paper, we briefly summarize preliminary insights from these two projects. Although these data do not specifically investigate robots, they provide invaluable context that is already influencing the education of roboticists and HRI researchers, as well as engineering disciplines more broadly. We conclude by discussing open questions these data suggest around the evolving nature of integrating LLMs within embodied educational robots (Sec. 4).

2 DEVELOPING AN LLM-BASED VIRTUAL TA

Design Process. During Spring23, our team experimented with various prototype LLMs for academic use, including one trained on all of the university's URL domains and another that summarized scientific literature, before settling on a virtual teaching assistant, called "HiTA", trained on course materials. This platform integrates LLMs with Retrieval Augmented Generation (RAG) technology to enhance the educational process by embedding educators within the LLM-assisted learning loop. It is uniquely designed to leverage the expertise of educators, allowing them to supervise AI interactions with students and ensuring that the generated content aligns with teaching goals and course materials. It features a multi-modal prompt processing subsystem including modes for general inquiries, homework assistance, and practice questions. Furthermore, HiTA enables educators to customize the system by uploading course-specific documents and setting guidelines for response generation, ensuring that responses are pertinent and educationally relevant. The web interface facilitates easy navigation for both students and educators to promote an efficient and engaging learning experience.

Preliminary Insights. In its pilot deployment, HiTA was used by 6 instructors and 400 students across 4 courses during Spring23. Students submitted over 14,000 questions. Feedback collected from users highlighted HiTA's utility, with a majority finding it helpful and preferring it over existing AI tools like ChatGPT for educational purposes. The comparison of student performances in courses before and after its introduction showed a positive impact on learning efficacy. The pilot run suggests that LLM-assisted tutoring tools have the potential to improve engineering education.

3 SURVEY OF ENGINEERING STUDENTS

Survey Methods. With ethics board approval, we surveyed all students at our USA-based engineering-focused R1 university in May 2023. Enrollments that semester included 6,938 students (5,350 undergraduate, 1,588 graduate). We received permission from the provost to use two separate list-servs to directly email a recruitment message to all undergraduate and graduate students. To encourage honesty, students could take the survey anonymously without providing any contact or identifying information, or they could fill in a separate form to opt-in to a drawing for one of four $25 e-gift cards. 827 eligible students began the survey and 593 completed it (72.3% completion rate). Our response rate (based on completed responses) was 8.5% of the student body; this sample is statistically representative of the university at a 98% confidence level with a 5% margin of error. We computed descriptive statistics and statistical tests and used directed content analysis [6] to inductively develop codebooks to capture emergent themes from free response questions. Our codebooks include complete code definitions along with data examples and are available at bit.ly/GenAICodebook.[1]

Survey Results. Only 32.8% of respondents had never tried LLMs; 36.8% had tried them at least once or twice; 30.4% used them regularly or daily. Figure 1 shows all use cases coded through analysis of

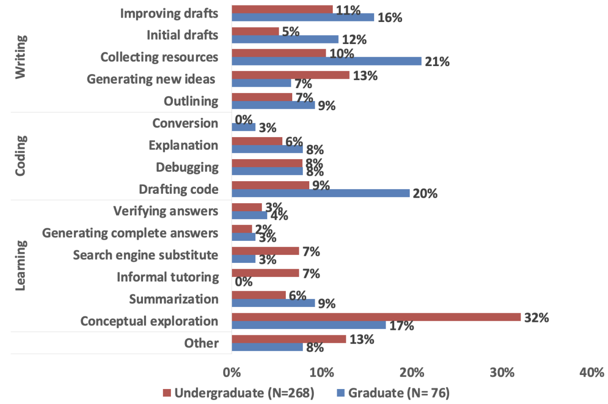

Figure 1: Emergent Student Use Cases of LLMs.

an optional free response question (answered by $N = 344$ students) that asked how students were using LLM-based tools in classes, research, or professional efforts. Exceedingly few students described behaviors that appeared to be intentional cheating (*i.e.* "generating complete answers"). Rather, students described a variety of use cases that they claimed deepened learning (*esp.* "conceptual exploration"), helped with coding (*e.g.,* drafting, debugging, or explaining code), or assisted with writing (*e.g.,* collecting resources, improving drafts, or generating new ideas).

Two required free response questions asked about the role of LLMs in engineering education. We report percentages of undergraduate/graduates as UG%/G%. 51%/53% believe that LLMs should be conditionally permitted in engineering education; 28%/31% think they should be generally permitted without restriction; 13%/12% feel they should be banned or discouraged. 43%/46% are concerned that LLMs may *damage* learning, while 25%/18% feel they can *improve* learning. 25%/25% anticipate using LLMs in their future career and 22%/21% believe they can improve efficiency. However students also expressed concerns about misinformation (18%/28%), societal issues (9%/12%), and job replacement (15%/12%). To cope with these issues, 11%/13% felt that Generative AI should be incorporated in engineering curricula, while a larger proportion (38%/44%) want to be taught specific career-relevant uses of Generative AI.

Students provided Likert ratings of the benefit of LLMs to their discipline on a scale of 1 (extremely damaging) to 10 (extremely beneficial). Ratings had a mean value of 6.5 ($SD = 2.28$) pointing toward general optimism, albeit with differences across departments. Interestingly, students who used LLMs more frequently were more likely to rate the benefit of GenAI more highly ($p < 0.0008$) and to encourage its use in education ($p < 0.002$). Students who have not used LLMs provided lower ratings of its benefits ($p < 0.0001$) and were more likely to discourage use of LLMs in education ($p < 0.0003$).

4 DISCUSSION OF IMPLICATIONS FOR HRI

The use of GenAI, particularly LLM-based chatbots, in engineering education is reminiscent of educational literature on ITS. Although prior attempts to make ITS more sophisticated and conversational did not always result in better student outcomes [8], the anecdotal success and student acceptance of systems like "HiTA" suggest we

[1]The Computer Science department is the second largest department on campus, but received the largest overall number of participants. Therefore we performed a preliminary analysis on the CS subset ($N = 133$ students), with a manuscript now under review [1]. In this paper, we share preliminary insights from across the entire university for the first time, with intent to publish a complete analysis soon.

should rethink these prior results. At the same time, the ease with which GenAI enables creation of familiar conversational interfaces is similar to the way that statistical language models largely supplanted algorithmic approaches to natural language processing. The resulting outputs seemingly capture nuance and semantics more readily than their expert system counterparts, but are also harder to interpret and manage. How can the design of future human-LLM interactions in education learn from findings from the HRI literature? Do the same design principles and philosophies hold, given the independent interactions students are already having with LLMs?

Our survey revealed that students readily embraced and adapted LLMs for various learning purposes (Figure 1). This supports the notion that embedding LLMs within educational robots might also be accepted by students, esp. if they can offer personalized learning experiences catered to individual needs and course content. By providing real-time, context-aware assistance and feedback, such robots could have the potential to overcome challenges identified by prior HRI work [4] and contribute to long-term learning outcomes. Envisioning robots not only as teachers but also as peer companions capable of learning alongside students introduces a novel approach to fostering camaraderie and collaborative learning, possibly mitigating the sense of isolation often associated with traditional educational methods. However, substantial consideration must be invested in determining the design of and access to such robots. Should instructors program large humanoid robot(s) to be available on campus in designated study rooms? Should students purchase their own small robots or interactive virtual agents that can be programmed into speakers (*e.g.,* Alexa)? Should robots be available for the **same** types of use cases that HiTA was built to support, or that emerged organically following the release of ChatGPT, or are there more sophisticated paths ahead? For example, many of the students in our sample used LLMs for coding or writing assistance, but raw code or drafts of written documents are not highly amenable to verbalization, whereas a use case like "concept exploration" might be highly portable between a chat interface and a robot. New types of prompting and fine-tuning will need to be developed in order to embody LLMs in robots, so that they can respond to student queries with appropriate, useful, and educational replies. In other words, future research should consider the emergent adoption of LLMs as valuable context without necessarily "overfitting" to these use cases.

ACKNOWLEDGMENTS

We sincerely thank all student participants in our survey. We also thank all of the student research assistants who have contributed to the development of HiTA and the student authors who are working on manuscripts describing our survey results. We would also like to thank the Trefny Innovation Instruction Center at the Colorado School of Mines, especially Deb Jordan, Megan Sanders, and Carter Moulton, for helping to pilot and refine the survey and dispensing the lottery prize. Finally, we thank the Office of the Provost for allowing us to use the global undergraduate and graduate list-servs to distribute our survey.

REFERENCES

[1] 2024. Early Adoption of Generative Artificial Intelligence in Computing Education: Emergent Student Use Cases and Perspectives in 2023.. In *Currently In Submission and Under Review.*

[2] John R. Anderson, C. Franklin Boyle, and Brian J. Reiser. 1985. Intelligent Tutoring Systems. *Science* 228, 4698 (April 1985), 456–462. https://doi.org/10.1126/science.228.4698.456

[3] John R. Anderson, Albert T. Corbett, Kenneth R. Koedinger, and Ray. Pelletier. 1995. Cognitive Tutors: Lessons Learned. *Journal of the Learning Sciences* 4, 2 (April 1995), 167–207. https://doi.org/10.1207/s15327809jls0402_2

[4] Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. 2018. Social robots for education: A review. *Science Robotics* 3, 21 (Aug. 2018), eaat5954. https://doi.org/10.1126/scirobotics.aat5954 Publisher: American Association for the Advancement of Science.

[5] Paul Denny, James Prather, Brett A. Becker, James Finnie-Ansley, Arto Hellas, Juho Leinonen, Andrew Luxton-Reilly, Brent N. Reeves, Eddie Antonio Santos, and Sami Sarsa. 2023. Computing Education in the Era of Generative AI. https://doi.org/10.48550/arXiv.2306.02608

[6] Hsiu-Fang Hsieh and Sarah E. Shannon. 2005. Three approaches to qualitative content analysis. *Qualitative Health Research* 15, 9 (Nov. 2005), 1277–1288. https://doi.org/10.1177/1049732305276687

[7] Sal Khan. 2023. Harnessing GPT-4 so that all students benefit. A nonprofit approach for equal access! https://blog.khanacademy.org/harnessing-ai-so-that-all-students-benefit-a-nonprofit-approach-for-equal-access/ Publication Title: Khan Academy Blog.

[8] James A. Kulik and J. D. Fletcher. 2016. Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. *Review of Educational Research* 86, 1 (March 2016), 42–78. https://doi.org/10.3102/0034654315581420 Publisher: American Educational Research Association.

[9] Weng Marc Lim, Asanka Gunasekara, Jessica Leigh Pallant, Jason Ian Pallant, and Ekaterina Pechenkina. 2023. Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education* 21, 2 (July 2023), 100790. https://doi.org/10.1016/j.ijme.2023.100790

[10] Douglas C. Merrill, Brian J. Reiser, Michael Ranney, and J. Gregory Trafton. 1992. Effective Tutoring Techniques: A Comparison of Human Tutors and Intelligent Tutoring Systems. *Journal of the Learning Sciences* 2, 3 (July 1992), 277–305. https://doi.org/10.1207/s15327809jls0203_2

[11] David P. Miller and Illah Nourbakhsh. 2016. Robotics for Education. In *Springer Handbook of Robotics*, Bruno Siciliano and Oussama Khatib (Eds.). Springer International Publishing, Cham, 2115–2134. https://doi.org/10.1007/978-3-319-32552-1_79

[12] Omar Mubin, Catherine J. Stevens, Suleman Shahid, Abdullah Al Mahmud, and Jian-Jie Dong. 2013. A Review of the Applicability of Robots in Education. *Technology for Education and Learning* 1, 1 (2013). https://doi.org/10.2316/Journal.209.2013.1.209-0015

[13] Thomas Price, Rui Zhi, and Tiffany Barnes. 2017. *Evaluation of a Data-Driven Feedback Algorithm for Open-Ended Programming.* Technical Report. International Educational Data Mining Society. https://eric.ed.gov/?id=ED596568