Developing Explainable AI Systems to Support Feedback for Students

Hongming Li
University of Florida
hli3@ufl.edu

Anthony F. Botelho
University of Florida
abotelho@coe.ufl.edu

ABSTRACT

We present a research plan focused on developing explainable AI systems powered by large language models (LLMs) to provide safe, reliable, and effective feedback for students. The research integrates several early-stage exploratory analyses, including understanding differences between human-written and AI-generated feedback, building a taxonomy of effective feedback types, examining qualities of peer discourse that support engagement, and investigating student interactions with LLMs. The overarching goal is to create systems that validate LLM-generated content and improve the way feedback is provided through generative AI in educational contexts. The methodology involves a series of interconnected studies utilizing natural language processing, machine learning, and qualitative research techniques. By addressing the current limitations and concerns surrounding AI in education, this work aims to contribute to the responsible integration of AI-powered tools that genuinely support learning and complement human instruction, ultimately promoting educational equity and student success.

Keywords

Explainable AI, Large Language Models, Feedback, AI in Education, Machine Learning

1. INTRODUCTION

The rapid advancement of artificial intelligence (AI) has opened up new possibilities for enhancing educational experiences and supporting student learning. In particular, the emergence of large language models (LLMs) like GPT and their integration into educational technologies [2, 3] has the potential to transform the way feedback is provided to students [11]. These AI systems offer benefits such as automated assessment and personalized feedback for students. However, concerns remain about the accuracy, effectiveness, and reliability of these AI systems compared to teacher assessments, as well as risks like generating false information or plagiarized content [16, 9]. To harness the full potential of LLMs for educational purposes while addressing these concerns, it is crucial to develop explainable AI approaches that can generate meaningful, pedagogically sound, and contextually relevant feedback that is also safe and reliable. Existing research has explored various aspects of AI in education, including personalized task assignment [4], adaptive digital learning environments [7], and AI chatbots for facilitating student-AI interactions [18, 1, 8]. While these studies have demonstrated the potential benefits of AI in learning, there is a need for further investigation into the specific application of LLMs for providing feedback to students.

Feedback is a crucial element in the learning process as it assists students in recognizing areas for enhancement, clarifying misunderstandings, and directing their advancement [10]. Nonetheless, delivering personalized and prompt feedback can be complex for teachers [12, 13], particularly in large class setups. AI-driven feedback systems can be beneficial in this regard, utilizing LLMs to produce tailored and focused feedback on a large scale. To create efficient AI-based feedback systems, it is vital to comprehend the attributes that set apart high-quality, human-like feedback. Previous studies have stressed the significance of feedback that is precise, actionable, and in line with educational goals [15, 5]. Moreover, research has underlined the importance of integrating emotional and motivational components into feedback to boost student involvement and persistence [14, 17]. Expanding on these findings, our study aims to develop transparent AI systems that can provide secure, dependable, and efficient feedback to students resembling the best practices of human instructors. By harnessing the capabilities of LLMs and merging them with a profound understanding of effective feedback methodologies, we aim to design AI-powered tools that can deliver targeted, meaningful, and educationally sound assistance to learners.

The overarching research question guiding this work is: How can explainable AI systems powered by large language models be developed to provide safe, reliable, and effective feedback for students? This question encompasses several subquestions. 1) What are the key differences between human-written and AI-generated feedback, and how can these insights inform the development of LLM-powered feedback systems? 2) What types of feedback are most effective for supporting student learning, and how can LLMs be trained or evaluated to emulate these feedback types? 3) What factors influence the way students interact with LLMs, and how can these insights inform the design of AI-powered tools that

H. Li and A. F. Botelho. Developing explainable ai systems to support feedback for students. In B. Paaßen and C. D. Epp, editors, *Proceedings of the 17th International Conference on Educational Data Mining*, pages 998–1002, Atlanta, Georgia, USA, July 2024. International Educational Data Mining Society.

https://doi.org/10.5281/zenodo.12730029

effectively support and enhance learning?

2. PROGRESS TO DATE

In our pursuit of developing explainable AI systems to support feedback for students, we have conducted a series of exploratory analyses that lay the groundwork for this research direction. These analyses span several key areas, including understanding the differences between human-written and AI-generated feedback, building a taxonomy of effective feedback types, examining peer interactions in collaborative settings, and investigating student interactions with large language models (LLMs).

2.1 Analyzing Human-Written Content and AI-Generated Content

As part of the project to develop explainable AI systems for student feedback, we focuses on analyzing the operation of AI content detectors and aims to gain a deeper understanding of the subtle differences between human-written and AI-generated content, with the hope of applying these insights to educational feedback. By collecting a dataset of both types of text content and applying natural language processing techniques and machine learning models, we aim to uncover the key characteristics that set human written content apart. This analysis will enhance our understanding of the unique qualities of human feedback and inform the development of AI systems that can more closely emulate the effectiveness of teacher-written feedback. The insights gained from this work will directly contribute to creating explainable AI systems that provide meaningful and pedagogically sound, and humanely warm feedback.

2.2 Developing a Taxonomy of Effective Feedback Types

Building upon our analysis of human-written feedback, we are working on a project that focuses on constructing a taxonomy of effective feedback types. This taxonomy aims to categorize and understand the diverse ways in which teachers provide feedback to students, with the ultimate goal of training LLMs to generate feedback that aligns with these effective strategies. By analyzing a large corpus of teacher-written feedback, we are identifying the key dimensions and characteristics that define high-quality feedback. This taxonomic framework will serve as a foundation for evaluating the performance of AI-generated feedback and guiding the development of LLMs that can provide targeted and impactful support to students. The insights gained from this work will directly contribute to the creation of explainable AI systems that emulate the best practices of human educators in providing feedback.

2.3 Exploring the Dynamics of Peer Discourse and Collaboration in Collaborative Settings

Recognizing the importance of peer interactions in the learning process, we are also examining how students engage with each other in collaborative settings, providing students with ongoing peer feedback during their learning process in a different way. By analyzing log data from a collaborative mathematics learning platform, we are uncovering interaction patterns that support student engagement and help them overcome challenges. Through techniques such as frequency analysis, cluster analysis, we are identifying distinct profiles of collaboration and specific behaviors that characterize productive peer interactions. This initial exploratory work lays the foundation for our planned empirical analyses. By understanding these dynamics, we aim to develop AI systems that facilitate and enhance peer feedback, ultimately supporting students in their educational journeys.

2.4 Investigating Student Interactions with LLMs

A key component of our research on developing explainable AI systems for student feedback involves understanding how students interact with LLMs. In a recent study, we explored how students utilize the chat functionality of ChatGPT to supplement their learning in a computer science education graduate course. By examining the dialogue history between students and the AI, we aimed to understand the factors that influence the way students engage with LLMs and how these interactions impact their learning experiences.

Figure 1 presents a set of visualizations derived from our initial analysis, offering a view of the longitudinal trends across extracted dialogue features. While we did not observe statistically significant differences between the high and low experience groups, likely due to the limited sample size, the visualizations reveal several interesting patterns. Students with higher prior experience tended to engage in more dialogue rounds, using more words and characters per message, although the complexity of the language used remained similar across both groups. These trends suggest that while prior CS experience may share a relationship with the quantity of interactions students have with ChatGPT, it does not necessarily correlate with the quality or complexity of the language used. We further examined the correlations between students' prior experience and the various dialogue features using a Spearman correlation heatmap (Figure 2). The heatmap suggests that higher confidence and experience in computer science and programming is associated with a slightly lower likelihood of starting conversations with a question, as well as the use of shorter words on average. Conversely, higher experience was positively correlated with longer conversations and a greater number of topics discussed. These findings align with the trends observed in the longitudinal analysis and provide additional support for the notion that prior CS experience may influence the way students interact with ChatGPT, even if the effects are not strongly pronounced.

This study provides insights into how students perceive and leverage AI tools, informing the design of explainable AI systems that can effectively support and enhance student learning. As we continue to refine and expand upon these lines of research, we are confident that they will converge to inform the development of AI-powered tools that provide targeted, meaningful, and pedagogically sound support to learners. The study also revealed some limitations, such as the potentially rigid nature of the guidance provided in the course materials. This aligns with the perspectives mentioned in previous research, suggesting that inflexible and non-encouraging prompts may hinder students' opportunities to genuinely explore and demonstrate their action characteristics [6]. Consequently, in our current ongoing work, we offer more diverse assignment guidance and set open-

ended research questions, encouraging students to interact with ChatGPT using spontaneous strategies. This approach has the potential to foster deeper reflection and novel insights.

3. DISCUSSION

The exploratory analyses conducted as part of our research on developing explainable AI systems for student feedback have insights. By examining the nuances between human-written and AI-generated content, we have identified key characteristics that distinguish human feedback, informing the development of AI systems that can more closely emulate the effectiveness of teacher-written feedback. This aligns with our first research question, which seeks to understand these differences and leverage them to create more human-like AI-powered feedback systems.

Our work on constructing a taxonomy of effective feedback types addresses our second research question, which aims to identify the most effective feedback strategies for supporting student learning. By categorizing and analyzing a large corpus of teacher-written feedback, we are developing a framework that can guide the training and evaluation of LLMs to generate feedback that aligns with best practices in education. This taxonomy will serve as a foundation for creating AI systems that provide targeted and impactful feedback to students.

The investigation of peer interactions in collaborative settings has revealed distinct profiles of collaboration and specific behaviors that characterize productive peer feedback. While our initial exploratory work has provided valuable insights, we acknowledge the need for further empirical analyses that incorporate richer data on student discourse and expression. By collecting and analyzing this data, we can gain a deeper understanding of the dynamics that support engagement and help students overcome challenges in collaborative learning environments. This will inform the development of AI systems that facilitate and enhance peer feedback, addressing an important aspect of our research goals.

Our study on student interactions with LLMs has shed light on the factors that influence how students engage with these tools and the impact on their learning experiences. The longitudinal trends and correlations observed in our analysis suggest that prior experience in computer science and programming may influence the quantity and nature of student interactions with ChatGPT. These findings contribute to the growing body of literature on student-AI interactions [18, 1, 8] and highlight the importance of considering individual differences in the design of AI-powered learning tools. However, we recognize the limitations of our current study, particularly the small sample size, which may have hindered the detection of statistically significant differences between experience groups. Future research should aim to recruit a larger and more diverse sample to validate and extend our findings, as recommended by recent studies [3].

4. CONCLUSION AND FUTURE WORKS

In conclusion, our research on developing explainable AI systems to support feedback for students has made significant progress through a series of exploratory analyses. Moving forward, our research will focus on several key areas to advance the development of explainable AI systems for student feedback. One of our primary goals is to develop the Verification and Evaluation Tool for Targeting Invalid Narrative Generation (VETTING), which aims to validate the content produced by LLMs and ensure its accuracy, reliability, and appropriateness for educational contexts. VETTING will incorporate insights from our exploratory analyses, leveraging techniques such as natural language processing, machine learning, and educational theory to create a robust system for evaluating AI-generated feedback.

Another crucial next step in our research is to develop systems that can validate the content produced by LLMs. This aligns with our larger goal of improving how feedback is given through generative AI. By creating mechanisms to assess the quality, relevance, and pedagogical soundness of LLM-generated feedback, we can ensure that students receive accurate and effective support that enhances their learning experiences.

To achieve these goals, we plan to conduct further empirical studies that build upon our exploratory analyses. This will involve collecting and analyzing data from a wider range of educational settings, incorporating diverse student populations and subject domains. By expanding the scope of our research, we can refine and validate our findings, ensuring that the AI systems we develop are applicable and effective across various educational contexts. We intend to collaborate with educators and domain experts to gather insights and feedback on our proposed AI systems. This will help us align our research with the practical needs and challenges faced by teachers and students in real-world educational settings. By engaging stakeholders throughout the development process, we can create AI tools that are not only technically advanced but also pedagogically sound and user-friendly.

5. ACKNOWLEDGMENTS

We would like to thank NSF (e.g., 2331379, 1903304, 1822830, 1724889), as well as IES (R305B230007), Schmidt Futures, MathNet, and OpenAI.

6. REFERENCES

[1] A. Almusaed, A. Almssad, I. Yitmen, and R. Z. Homod. Enhancing student engagement: Harnessing "aied"'s power in hybrid education—a review analysis. *Education Sciences*, 13(7):632, 2023.

[2] S. Baral, A. F. Botelho, J. A. Erickson, P. Benachamardi, and N. T. Heffernan. Improving automated scoring of student open responses in mathematics. *International Educational Data Mining Society*, 2021.

[3] A. F. Botelho, S. Baral, J. A. Erickson, P. Benachamardi, and N. T. Heffernan. Leveraging natural language processing to support automated assessment and feedback for student open responses in mathematics. *Journal of Computer Assisted Learning*, 39(3):823–840, 2023.

[4] A. F. Botelho, A. Varatharaj, T. Patikorn, D. Doherty, S. A. Adjei, and J. E. Beck. Developing early detectors of student attrition and wheel spinning

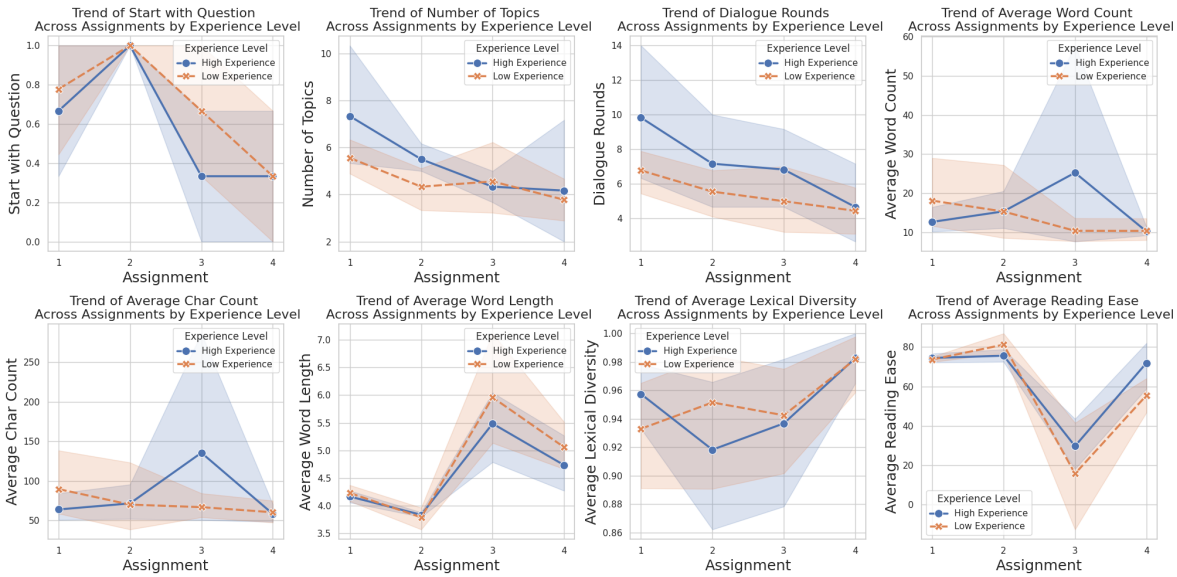

Figure 1: Sequential Assignment: Longitudinal Trends in Dialogue Features

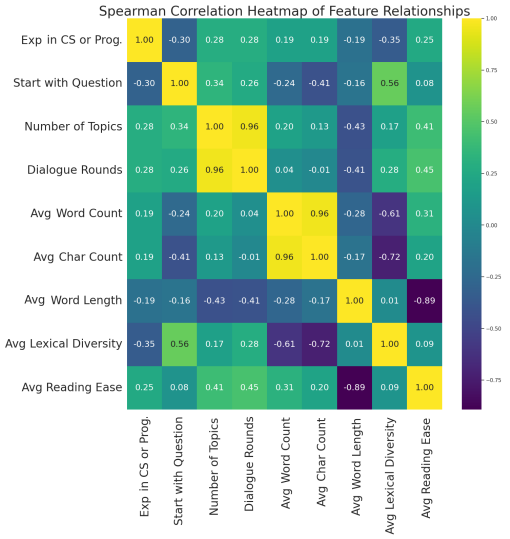

Figure 2: Spearman Correlation Heatmap: Influence of CS Programming Experience on Dialogue Features

using deep learning. *IEEE Transactions on Learning Technologies*, 12(2):158–170, 2019.

[5] M. D. Cannon and R. Witherspoon. Actionable feedback: Unlocking the power of learning and performance improvement. *Academy of Management Perspectives*, 19(2):120–134, 2005.

[6] S. Chen, F. Ouyang, and P. Jiao. Promoting student engagement in online collaborative writing through a student-facing social learning analytics tool. *Journal of Computer Assisted Learning*, 38(1):192–208, 2022.

[7] T. K. Chiu, Q. Xia, X. Zhou, C. S. Chai, and M. Cheng. Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 4:100118, 2023.

[8] Y. Dai, A. Liu, and C. P. Lim. Reconceptualizing chatgpt and generative ai as a student-driven innovation in higher education. *Procedia CIRP*, 119:84–90, 2023.

[9] E. Francke and A. Bennett. The potential influence of artificial intelligence on plagiarism: A higher education perspective. In *European Conference on the Impact of Artificial Intelligence and Robotics (ECIAIR 2019)*, volume 31, pages 131–140, 2019.

[10] J. Hattie and H. Timperley. The power of feedback. *Review of educational research*, 77(1):81–112, 2007.

[11] H. Khosravi, S. B. Shum, G. Chen, C. Conati, Y.-S. Tsai, J. Kay, S. Knight, R. Martinez-Maldonado, S. Sadiq, and D. Gašević. Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 3:100074, 2022.

[12] F. Kibaru. Supporting faculty to face challenges in design and delivery of quality courses in virtual learning environments. *Turkish Online Journal of Distance Education*, 19(4):176–197, 2018.

[13] L. Major and G. A. Francis. Technology-supported personalised learning: a rapid evidence review. 2020.

[14] S. Narciss, S. Sosnovsky, L. Schnaubert, E. Andrès, A. Eichelmann, G. Goguadze, and E. Melis. Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education*, 71:56–76, 2014.

[15] D. J. Nicol and D. Macfarlane-Dick. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2):199–218, 2006.

[16] V. J. Owan, K. B. Abang, D. O. Idika, E. O. Etta, and B. A. Bassey. Exploring the potential of artificial intelligence tools in educational measurement and

assessment. *EURASIA Journal of Mathematics, Science and Technology Education*, 19(8):em2307, 2023.

[17] E. A. Skinner and J. R. Pitzer. Developmental dynamics of student engagement, coping, and everyday resilience. In *Handbook of research on student engagement*, pages 21–44. Springer, 2012.

[18] Q. Xia, T. K. Chiu, C. S. Chai, and K. Xie. The mediating effects of needs satisfaction on the relationships between prior knowledge and self-regulated learning through artificial intelligence chatbot. *British Journal of Educational Technology*, 54(4):967–986, 2023.