The need for discipline-specific use cases: Computing teachers' views of LLM programming error messages explanations in the classroom

We examine the importance of discipline-specific research on large language models (LLMs) in education, using a recent empirical study on LLM programming error messages (PEMs) explanations in secondary computer science education as a case study. Using LLMs to enhance and explain PEMs, which are often difficult to decipher by novices, has been identified as a promising use case for using LLMs to support learning. The study investigated educators' views on using LLM explanations in the classroom, revealing insights that highlight the nuanced implications of these technologies in particular educational contexts.

RQ1: What are secondary educators' views on the potential classroom use of LLM program error message explanations?

RQ2: In what ways can a feedback literacy perspective support the analysis of educators' views of potential classroom use of LLM program error message explanations?

Eight expert secondary computer science educators from England were interviewed on their experiences exploring an augmented IDE that used OpenAI's GPT-3.5 to generate explanations for Python program error messages. We correlated teachers' views with both enhanced PEM guidelines (programming-specific research on how to improve PEMs) and feedback literacy theory (general education research). This dual analysis revealed that while many educator concerns aligned with existing PEM guidelines, feedback literacy theory provided a more holistic framework for understanding the implications of LLM use in this context.

The discipline-specific approach allowed us to identify key themes relevant to computer science education, such as the debate over including code solutions in explanations, the importance of consistent use of key concept words, and concerns about potential student over-reliance on LLM feedback. The study's findings suggest that LLM content should be encouraging, detailed, jargon-free, use keywords consistently, align with learning objectives, and avoid or delay providing code solutions. IDE design should allow educators to monitor student use of LLM explanations and help manage invalid or unrelated feedback. Focusing on this specific use case, the research revealed insights that may not have emerged from broader, non-contextual studies of LLM applications in education.

The findings also have direct implications for the design of LLM-augmented programming environments, lesson and curriculum design, and teacher professional development in computer science education. For example, the study suggests that LLM explanations should aim to guide understanding rather than simply provide solutions, and that both students and teachers may need support in developing AI and feedback literacy to effectively use and interpret LLM-generated feedback.

More broadly, this case underscores the value of discipline-specific LLM research in education. Such studies can:

1. Uncover unique challenges and opportunities within particular subject areas
2. Inform the development of specialized LLM applications tailored to specific educational contexts
3. Provide empirical evidence to guide policy and practice in integrating LLMs into specific curricula
4. Contribute to the refinement of general theories about LLM use in education by providing concrete, contextual examples

As the educational community grapples with the implications of AI technologies, such targeted investigations will play a crucial role in shaping effective, ethical, and pedagogically sound implementations of LLMs across various educational domains.