Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025
Paper ID #49218

From Reflection to Insight: Using LLM to Improve Learning Analytics in Higher Education

Dr. Nasrin Dehbozorgi, Kennesaw State University

I'm an Assistant Professor of Software Engineering and the director of the AIET lab in the College of Computing and Software Engineering at Kennesaw State University. With a Ph.D. in Computer Science and prior experience as a software engineer in the industry, my interest in both academic and research activities has laid the foundation to work on advancing educational technologies and pedagogical interventions.

Mourya Teja Kunuku, Kennesaw State University

Ph.D. student at Kennesaw State university. Research Interest include Deep learning, Generative AI, LLMs

From Reflection to Insight: Using LLM to Improve Learning Analytics in Higher Education

Abstract

The integration of Artificial Intelligence (AI) into educational tools has revolutionized modern education by enhancing pedagogical practices and learning analytics. The emergence of Large Language Models (LLMs) has further accelerated this transformation by enabling complex analysis of textual data that would otherwise be labor-intensive for instructors. Reflective writing is a key component in educational practices which foster deeper cognitive and metacognitive skills among students. Typically, reflective techniques require students to articulate their learning processes in natural language. However, the effectiveness of these practices is maximized when students receive feedback on their reflective writings. Due to the time-consuming nature of analyzing these writings, the implementation of reflective practices has been limited. In this study, we introduce 'Cogni-Reflect', an LLM-powered tool designed for the automated analysis of student reflections. Cogni-Reflect extracts students' learning outcomes and challenges from their reflective submissions and visualizes the frequency distribution of these topics through a dynamic dashboard. This visualization enables instructors to apply timely interventions after each class session based on students' learning trajectories. The analysis of the model's performance is promising, demonstrating over 95% accuracy in extracting meaningful topics for analyzing students' understanding of the subject matter.

Introduction

Computer Science education has undergone significant changes as a result of the rapid advancement of AI. Students are becoming more dependent on these technologies due to the accessibility of AI-based content. Concerns have been raised about student engagement, learning outcomes, and retention rates in higher education due to the widespread use of AI resources. Student success depends heavily on engagement, which can be attained by implementing formative assessment, critical thinking, and reflective thinking techniques. Reflection is particularly important among these because it encourages the growth of metacognitive and critical thinking abilities. Quizzes, exams, and surveys which are conventional methods for collecting feedback and evaluating students' learning don't have the ability to reveal students' learning progress in real time. The manual review of students' reflections takes a lot of time and could be biased in some ways. Comparatively, quantitative methodologies might be unable to adequately capture the complex nuance and depth of reflective thought. Traditional ML-based automated text analysis methods are promising, however they require a large amount of data for training and may have trouble correctly interpreting context-specific nuances. This emphasizes the need for more research on developing AI-powered systems that perform in-depth analysis of students' reflections. Such a system should give educators insightful information about their students' Learning Outcomes (LO) and challenges, enabling them to modify their courses and plan for future improvements. In our previous works [4, 22], we developed a reflection analysis tool that uses NLP and LLM methods to extract students' learning outcomes from their class reflections. The main goal was to provide insights to instructors for improving course content, instructional strategies, and evaluation methods through an interactive dashboard that dynamically displays students' learning outcomes and challenges. To address the shortcomings of the prior work, in this study we proposed a more advanced tool 'Cogni -Reflect' powered by the capabilities of the LLMs for in-depth analysis of students' reflections. In the following section, we provide an overview of the related literature, followed by our proposed methodology. We then conclude with discussions on the findings and outline plans for future work.

Related Work

Students engage in reflection by critically evaluating their educational experiences, identifying strengths and areas for development, and formulating growth-oriented strategies. This process enhances self-awareness, critical thinking, and metacognitive abilities. Research indicates that reflection boosts student engagement [1] and improves learning outcomes by helping identify knowledge gaps [2]. It fosters active learning, increases comprehension, and facilitates knowledge application in new contexts, leading to better academic performance [3,4]. Additionally, reflective practices enhance students' motivation, perseverance, and self-efficacy [5]. Reflection is supported by formative assessment and enhances teacher performance as well [6]. Educators use it to evaluate their methods, adjust their approaches, and promote professional development which leads to effective instruction and improved pedagogy. In the educational sector, instructors incorporate various reflective learning techniques into their curriculum to enhance students' metacognitive abilities, which provides them with opportunities to understand their own learning experiences [7]. In the CS education domain, specific competencies such as problem-solving, algorithmic thinking, synthesis, and evaluation necessitate the incorporation of the reflection process [8]. The adoption of diverse reflective learning methodologies prepares students to think effectively when designing and developing systems [9]. According to Fekete [10], the reflection process has an indirect yet significant impact on cultivating a diverse range of technical literacies, which is the ultimate goal of CS education. Therefore, it is worth encouraging more reflection practices in educational institutes. Educators utilize a variety of reflection tools, including quantitative techniques such as surveys and questionnaires to evaluate reflective thinking, as well as reflective journals and instructor-led discussions or prompts [11], which require manual analysis of the reflections. Each strategy has its drawbacks. For example, quantitative methods like surveys may fail to capture the depth and nuances of reflective thinking, while manually analyzing students' reflections can be time-consuming and subjective, particularly in large classes [12]. Regardless of the specific method or aspect of reflection being utilized, the manual analysis of reflections remains a resource-intensive task for educators, which hinders its widespread application despite its benefits [9]. Conversely, this challenge has spurred recent research focused on leveraging advanced AI and NLP techniques to automate the analysis of reflections [9]. Significant advancements have been made in applying AI for learning analytics and text analysis in educational data [12]. However, there remains a gap in utilizing cutting-edge models for real-time, automated, and in-depth analysis of students' reflections.

To automate the analysis of reflections, educational researchers have employed a wide range of technologies and NLP/ML models. These techniques include topic modeling, which identifies the primary ideas and themes in reflective texts [13],[14]; text classification models that categorize reflections based on established standards [15]; Latent Dirichlet Allocation (LDA), commonly used for topic modeling in student reflections [13], and LLM-based approaches to analyze narratives of student reflections[22] .Additionally, various ML methods such as Support Vector Machines (SVM) [16] and Naive Bayes classifiers [17], [18] have been employed, along with the more recent transformer-based language models such as BERT (Bidirectional Encoder Representations from Transformers) which is a pre-trained model introduced by Google to optimize various NLP tasks and achieving state-of-the-art performance [19]. The review of existing research on automated models for reflection analysis classifies them into three main approaches: 1) dictionary-based approaches, which calculate word frequency using a predefined dictionary; 2) rule-based approaches, which require experts to define rules and extract patterns for text interpretation; and 3) ML approaches, which primarily utilize ML and NLP algorithms for text analysis. Among the classification methods, SVM and Naïve Bayes are identified as key classifiers applied in reflection analysis models [20],[12] while Random Forests and Neural Networks have been shown to outperform these models on certain datasets [21]. In a recent study, researchers used DistilBERT, a transformer-based bidirectional deep contextual language model, for automatic feature generation and applied a logistic regression classifier on the generated features for scoring purposes, fine-tuning the model with context-specific data [2]. However, gaps remain in understanding the scalability and generalizability of these advanced models across diverse educational contexts. Further research is needed to examine the robustness of these models in real-time educational settings and their integration with existing educational technologies to effectively enhance reflective learning practices. While significant advancements have been made in utilizing AI for text analysis on educational data, the emergence of LLMs has opened new horizons in applying cutting-edge generative AI models for in-depth analysis of students' reflections. In the next section, we present our proposed LLM-based model, Cogni-Reflect, for analyzing students' narrative reflections.

Methodology

In this study, we present Cogni-Reflect, a novel LLM-powered tool designed to analyze students' narrative reflections and provide actionable insights to educators. Cogni-Reflect leverages the advanced capabilities of LLMs to extract topics related to students' challenges and areas of interest during lectures. These insights are then visualized on a dynamic dashboard, offering two levels of analysis: class-level insights, which summarize weekly trends, and student-level analytics, which track individual learning progress over a semester. This dual-layered approach empowers instructors to implement timely and targeted interventions, enhancing overall learning outcomes. The development of Cogni-Reflect involved three main stages: iterative fine-tuning of the LLaMA 3.1 model, systematic construction and evaluation of datasets, and seamless integration of the fine-tuned model into a user-friendly dashboard.

Data Collection

Data for the tool was collected using the "minute paper" method, where students provided brief feedback after each class session on two aspects: the concepts they learned and the challenges they faced. This method was implemented over four years within a college-level Software Engineering program, yielding a comprehensive dataset that formed the foundation for fine-tuning the model. The dataset used in this study consists of 20 sessions, with each session containing 10 student reflections, leading to a total of 200 reflections per iteration. The model was not trained on a specific course but rather on how to understand student reflections and identify patterns, making it generalizable across different subject areas. The fine-tuning process was conducted iteratively over 10 iterations, with each iteration incorporating a new set of 20 sessions (200 reflections). Throughout this process, the dataset was split into 70% training, 15% validation, and 15% testing to ensure that model evaluation was performed on unseen data. This dataset served as the foundation for fine-tuning the LLM, allowing the model to identify recurring patterns and topics of challenge or interest at both the course and individual student levels. We acknowledge the importance of addressing potential biases in LLMs. In our approach, the use of real-world data for both training and testing inherently reduces bias, as it reflects diverse student demographics and writing styles. Additionally, the training and test splits were carefully designed to ensure a balanced representation, further mitigating bias. By fine-tuning the model on this real-world data, we enhance its ability to adapt to authentic student reflections, thereby reducing the risk of misinterpretations during topic extraction. These steps collectively help minimize potential biases in the model's responses.

Iterative Fine-Tuning of LLM (Memory Tuning)

The fine-tuning of the LLaMA 3.1 8B Instruct model began with prompt engineering to establish a baseline performance. The process was designed as an iterative cycle to progressively enhance the model's accuracy and robustness in extracting meaningful topics from student reflections. The initial stage involved the creation of an evaluation dataset comprising 20 manually curated samples that presented diverse and challenging scenarios for the model. This dataset served as

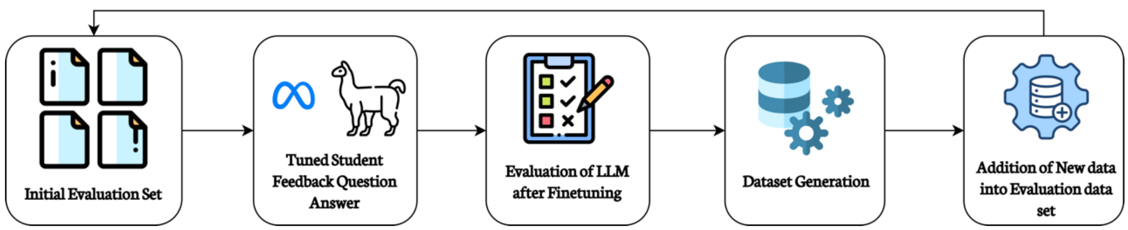

Figure 1: Iterative Fine-tuning Process

the foundation for assessing the model's initial performance. Subsequent iterations expanded the dataset by adding 20 additional samples in each round. After processing these new samples, the results were appended to the growing dataset, increasing its size and diversity. This expansion allowed the model to progressively adapt to more complex and nuanced feedback.

Early iterations focused primarily on reducing errors such as hallucinations, where the model extracted incorrect or irrelevant topics. As the fine-tuning progressed, later iterations refined the model's ability to handle nuanced feedback and address edge cases effectively. Each iteration involved three key steps: evaluating the model on the expanded dataset, conducting detailed error analysis, and retraining the model using the augmented dataset. This iterative loop ensured continuous learning and improvement, as shown in Figures 1 & 2. The process culminated in a model that achieved over 95% accuracy in topic extraction, demonstrating its capability to analyze student reflections with a high degree of precision and reliability.

Evaluation Dataset and Scoring

The evaluation process began with the construction of a small yet high-quality dataset of 20 manually curated examples designed to challenge the model's topic extraction capabilities. This initial dataset allowed for a focused and controlled approach to identifying weaknesses and implementing improvements. The iterative nature of the process ensured that each subsequent round of evaluation built upon the insights gained in the previous iteration, delivering tangible and measurable improvements at every stage. As the dataset expanded with each iteration adding 20 new examples per cycle it became progressively more representative of the diverse and nuanced challenges the model was likely to encounter in real-world applications. A similarity-based scoring approach was employed to evaluate the model's performance, allowing for semantic variations while preserving the core meaning of the extracted topics. For example, phrases such as "struggling with requirement validation" and "requirement validation challenges" were treated as equivalent due to their semantic similarity. Two key elements of the evaluation process were structured output parsing and performance metrics. The structured output parser ensured that the model produced consistent responses in the correct format, and it was used to evaluate empty responses for validity. Empty responses were deemed correct if no challenges were identifiable in the student feedback; otherwise, they were marked as errors.

Performance metrics tracked the accuracy of valid topic extraction and the correct handling of empty responses, starting with a baseline of 55% for valid topic extraction and 30% for correct empty responses. Over 10 iterations, the model's performance improved significantly, reaching over 95% accuracy in valid topic extraction and 87% accuracy in handling empty responses. These results demonstrate the effectiveness of the iterative approach in refining the model's ability to extract meaningful insights from student reflections while handling edge cases and nuanced feedback more effectively.

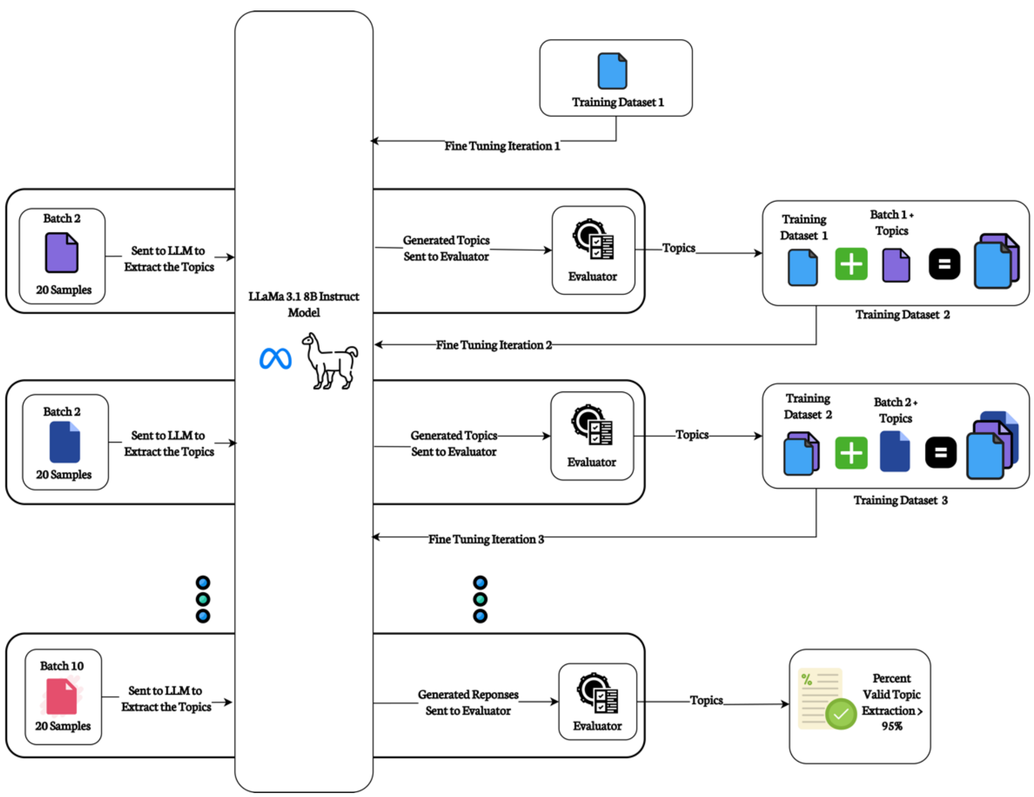

Figure 2: Fine-tuning Breakdown

Cogni-Reflect System Architecture

The architecture of Cogni-Reflect is designed to streamline the analysis of student reflections and provide actionable insights through a simple and efficient workflow as shown in Figure 3.

Instructors upload reflections in Excel format via the user interface (UI), where the fine-tuned LLaMA 3.1 model processes the data to extract key topics related to challenges and learning outcomes. The system supports two modes of analysis: Weekly Analytics, which aggregates class-wide reflections to identify top challenges and interests, and Student Analytics, which provides individualized feedback by tracking each students' learning progress over the semester. Processed results are stored in a MongoDB database and displayed on a dynamic dashboard as shown in Figure 4, allowing instructors to visualize trends and make data-driven teaching decisions. This architecture, by providing real-time feedback, empowers educators to enhance both instructional strategies and student outcomes effectively.

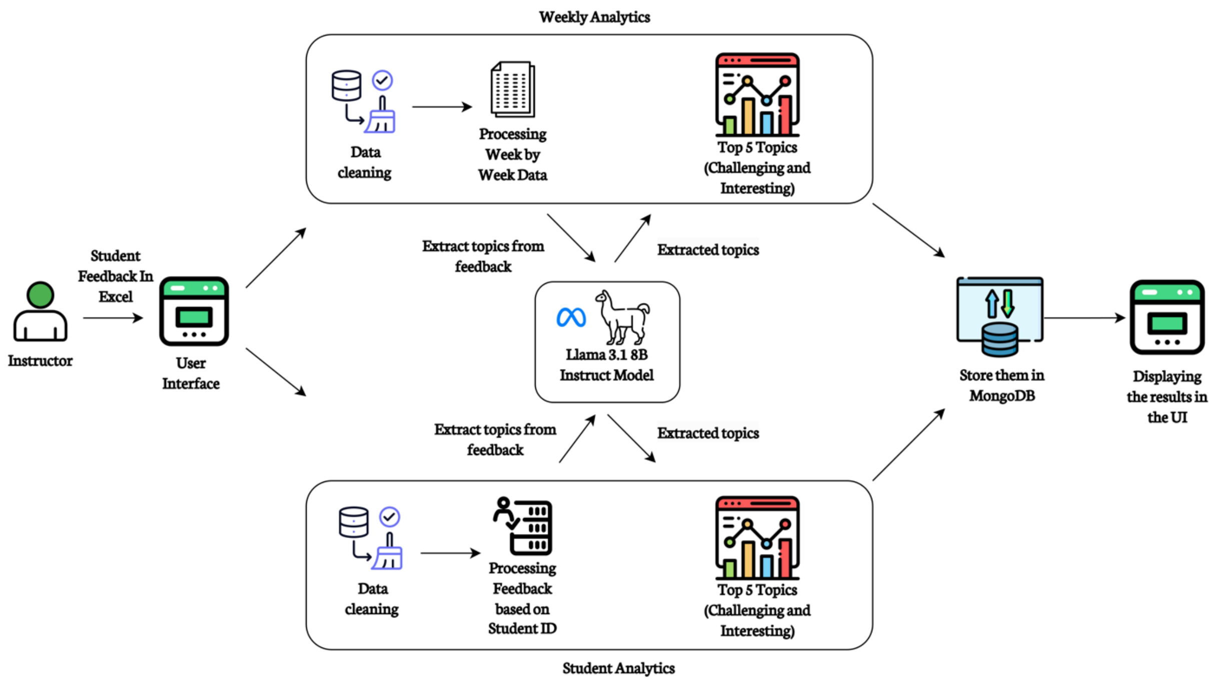

Figure 3: Cogni-Reflect Application Workflow

Results

The performance of the Cogni-Reflect model was evaluated over 10 iterations of fine-tuning, demonstrating substantial improvements in its ability to extract meaningful topics from student reflections. Two key metrics were tracked: valid topic extraction accuracy and correct handling of empty responses.

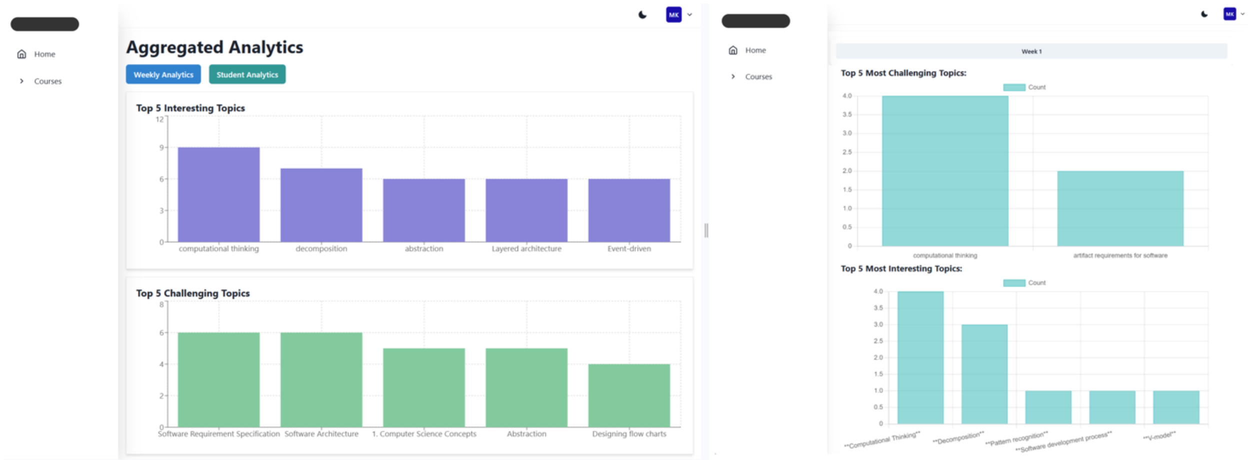

Figure 4: Sample UI - Aggregated Analytics (left), Weekly Analytics (right)

The baseline performance began at 55% for valid topic extraction and 30% for empty responses. Through iterative fine-tuning, the model achieved over 95% accuracy in valid topic extraction and 87% accuracy in handling empty responses, as shown in Figure 5.

Early iterations focused on reducing errors such as hallucinations, where the model extracted irrelevant or incorrect topics. Later iterations refined the model's ability to handle nuanced feedback and edge cases more effectively. This iterative approach ensured continuous improvement, with each round of fine-tuning building on the previous one. The results demonstrate that the model is well-equipped to analyze student reflections with high precision, providing reliable and actionable insights for educators.

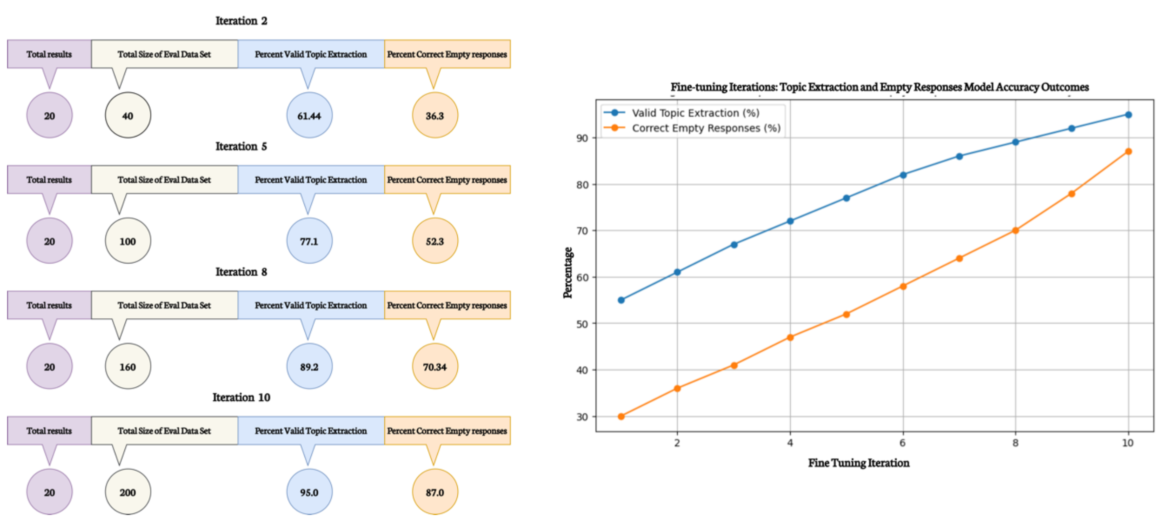

Figure 5: Fine-tuning Iteration Results

Conclusion and Future Work

In our previous iterations of the work, we utilized **BERT** and other open-source, non-fine-tuned models for analyzing student reflections. However, through our experiments, we observed that these models struggled with accurately extracting meaningful topics, especially in handling nuanced student feedback and identifying empty responses. The iterative fine-tuning approach applied to Cogni-Reflect significantly improved performance, leading to better topic extraction accuracy and handling of diverse writing styles. These improvements motivated us to further develop and publish our refined approach. While we acknowledge the importance of comprehensive benchmarking against other state-of-the-art models, our current focus has been on demonstrating the iterative fine-tuning process's impact on enhancing reflection analysis. This study highlights the effectiveness of iterative fine-tuning in optimizing the LLaMA 3.1 8B Instruct model for analyzing student reflections. With over 95% accuracy in valid topic extraction and 87% accuracy in handling empty responses, the model has proven its ability to provide reliable insights that can enhance teaching strategies and improve student outcomes. These advancements demonstrate the value of leveraging LLMs for automating reflective learning analysis. Future work will focus on further improving the model's generalizability to diverse feedback types and evaluating its scalability in real-time classroom environments. These enhancements aim to solidify Cogni-Reflect as a reliable and practical solution for automating reflective learning analysis and supporting data-driven decision-making in education. We will integrate additional data resources into the model to facilitate continuous learning. Additionally, we will conduct instructor usability studies to evaluate its effectiveness in their teaching practices and student learning. By providing students with resources tailored to topics they find challenging and offering the entire class relevant materials for collective challenges, these improvements will further empower educators to create dynamic and effective learning environments.

References

[1] Muhsin Menekse, Saira Anwar, and Zeynep Gonca Akdemir, "How do different reflection prompts affect engineering students' academic performance and engagement?" *The Journal of Experimental Education*, vol. 90, no. 2, pp. 261–279, 2022, Taylor & Francis.

[2] Ahmed Ashraf Butt, Saira Anwar, Ahmed Magooda, and Muhsin Menekse, "Comparative Analysis of the Rule-Based and Machine Learning Approach for Assessing Student Reflections," *Proceedings of the 16th International Conference of the Learning Sciences-ICLS 2022*, pp. 1577–1580, 2022, International Society of the Learning Sciences.

[3] Jennifer A. Moon, "Reflection in learning and professional development: Theory and practice," Routledge, 2013.

[4] Aditya Vadapally, Nasrin Dehbozorgi, and Dinesh Chowdary Attota, "Minute-paper dashboard: Identification of learner's misconceptions using topic modeling on formative reflections," *2022 IEEE Frontiers in Education Conference (FIE)*, pp. 1–5, 2022, IEEE.

[5] Tuğrul Yürük, "Kitap, Tez, Sempozyum Değerlendirmeleri/John Dewey, How We Think, A Restatement of The Relation of Reflective Thinking to The Educative Process," DC Heath and Company, 1933, 10+ 301 pages.

[6] David J. Nicol and Debra Macfarlane-Dick, "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice," *Studies in Higher Education*, vol. 31, no. 2, pp. 199–218, 2006, Taylor & Francis.

[7] Donald A. Schon, "Educating the reflective practitioner: Toward a new design for teaching and learning in the professions," *Australian Journal of Adult Learning*, vol. 50, no. 2, pp. 448–451, 2010.

[8] Sue Inn Chng, "Incorporating reflection into computing classes: models and challenges," *Reflective Practice*, vol. 19, no. 3, pp. 358–375, 2018, Taylor & Francis.

[9] Huda Alrashidi, Nouf Almujally, Methaq Kadhum, Thomas Daniel Ullmann, and Mike Joy, *"Evaluating an automated analysis using machine learning and natural language processing approaches to classify computer science students' reflective writing,"* in *Pervasive Computing and Social Networking: Proceedings of ICPCSN 2022*, pp. 463–477, 2022, Springer.
[10] Alan Fekete, Judy Kay, Jeff Kingston, and Kapila Wimalaratne, "Supporting reflection in introductory computer science," *ACM SIGCSE Bulletin*, vol. 32, no. 1, pp. 144–148, 2000, ACM New York, NY, USA.
[11] David Kember, Jan McKay, Kit Sinclair, and Frances Kam Yuet Wong, "A four-category scheme for coding and assessing the level of reflection in written work," *Assessment & Evaluation in Higher Education*, vol. 33, no. 4, pp. 369–379, 2008, Taylor & Francis.
[12] Thomas Daniel Ullmann, "Automated analysis of reflection in writing: Validating machine learning approaches," *International Journal of Artificial Intelligence in Education*, vol. 29, no. 2, pp. 217–257, 2019, Springer.
[13] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, no. Jan, pp. 993–1022, 2003.
[14] Nasrin Dehbozorgi and Stephen MacNeil, "Semi-automated analysis of reflections as a continuous course," in *2019 IEEE Frontiers in Education Conference (FIE)*, pp. 1–5, 2019, IEEE.
[15] Fabrizio Sebastiani, "Machine learning in automated text categorization," *ACM Computing Surveys (CSUR)*, vol. 34, no. 1, pp. 1–47, 2002, ACM New York, NY, USA.
[16] Thorsten Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *European Conference on Machine Learning*, pp. 137–142, 1998, Springer.
[17] Andrew McCallum and Kamal Nigam, "A comparison of event models for naive Bayes text classification," in *AAAI-98 Workshop on Learning for Text Categorization*, vol. 752, no. 1, pp. 41–48, 1998, Madison, WI.
[18] Rafael A. Calvo and Sidney D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 18–37, 2010, IEEE.
[19] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
[20] Charu C. Aggarwal and ChengXiang Zhai, "A survey of text classification algorithms," in *Mining Text Data*, pp. 163–222, 2012, Springer.
[21] Manuel Fernández-Delgado, Eva Cernadas, Senén Barro, and Dinani Amorim, "Do we need hundreds of classifiers to solve real world classification problems?" *The Journal of Machine* Learning Research, vol. 15, no. 1, pp. 3133–3181, 2014, JMLR.org.
[22] Dehbozorgi, N., & Kunuku, M. T. (2024, March). An LLM-based Reflection Analysis Tool for Identifying and Addressing Challenging Topics. In Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2 (pp. 1618-1619).