The Future of Engineering Education
2024 Annual Conference & Exposition
Oregon Convention Center
Portland, OR . June 23 - 26, 2024
ASEE
Paper ID #41721

From Manual Coding to Machine Understanding: Students' Feedback Analysis

Mr. Abdulrahman Alsharif, Virginia Polytechnic Institute and State University

Abdulrahman M. Alsharif is a research assistant for the Engineering Education Department and a PhD candidate at Virginia Tech.

Dr. Andrew Katz, Virginia Polytechnic Institute and State University

Andrew Katz is an assistant professor in the Department of Engineering Education at Virginia Tech. He leads the Improving Decisions in Engineering Education Agents and Systems (IDEEAS) Lab.

From Manual Qualitative Thematic Coding to Generative AI Understanding: Engineering Students Feedback Analysis

Abstract

In this study, we evaluate the use of generative AI (GAI) models for qualitative coding of open-ended student responses, compared to traditional natural language processing (NLP) methods. The main objective was to explore an in-house GAI method to develop themes from students' feedback responses. A systematic four-step process of text extraction, embedding, clustering, and code generation was employed on responses from a large engineering course regarding the transition to online learning during COVID-19. A locally deployed GAI model (Dolphin-Mistral 2.6) was used for privacy-preserving text extraction, with the UAE-Angle embedding model enabling the clustering of similar responses. GAI was then leveraged to generate qualitative codes and themes from the clusters. Human evaluation (i.e., human in the loop process) found the GAI-generated codes displayed high similarity to human-generated codes, with minor terminology distinctions. Key themes emphasized the importance of instructor feedback, communication strategies, engagement approaches, and resource accessibility for effective online learning experiences. Treemap visualizations aided the interpretation of the hierarchical code structure. While human input was still required for consolidating overlapping sub-codes, the study demonstrates GAI's potential to semi-automate qualitative coding tasks traditionally performed manually, while ensuring data privacy through local deployments. Future work could explore more advanced GAI models to further streamline the clustering and code generation workflow.

Introduction

With the rise of generative AI (GAI) tools such as GPT-4, Claude, Llama, and others, interpreting textual data and generating coherent responses is now possible with powerful capabilities. Traditionally, analyzing large qualitative datasets like student survey responses requires extensive time. Natural language processing (NLP) can help decrease the time needed and provide a solution to this obstacle. However, NLP cannot perform equally well on every task; some tasks see better performance than others. Because each model is trained on different datasets, choosing a model that best fits the data on hand is very important before starting the work. Not only does knowing what data a model was trained on give insight into its strengths and limitations for different tasks but also understanding the training data of a model provides information about the contexts and patterns the model will recognize. This knowledge allows for an assessment of which types of tasks the model will execute effectively versus those at which it may struggle. Selecting an appropriate model and having knowledge of its training data helps ensure optimal results. For example, NLP techniques like sentiment analysis on short responses and word clustering perform relatively well [1]. But, when applied to large text formats, the accuracy of NLP can drop.

Applying NLP to qualitative research requires understanding four key elements: the analytical goals, the textual data available, model capabilities, and appropriate data cleaning for good model performance. Multiple steps are needed including assessing what questions the analysis aims to answer, determining the format and content of the texts at hand, selecting an NLP model with fitting strengths, and cleaning data. Hence, the patterns match what the model was trained on. Carefully considering each of these areas allows for successful application of NLP to qualitative research tasks. Although NLP could do qualitative research tasks relatively well (e.g., sentiment analysis and clustering themes), the challenge of working with more complex textual data remains. Another challenge is generating coherent results to provide users. Model outputs sometimes require human input to fine-tune. For example, a user may need to manually name and organize text clusters in a way that makes sense for their purposes. Some level of human involvement can be needed to ensure models deliver optimal outputs for end applications and users' needs are met.

On the other hand, GAI can automatically cluster text if used in ways to respect the context window limits and identify themes while also providing suitable names and reasoning for its outputs. Unlike some NLP models, GAI tools build in coherent naming and justification to accompany analysis results. With this process, GAI provides transparent and integrated reasoning that helps users interpret AI-generated text groupings and themes. It can produce outputs (e.g., themes or clusters) coupled with clear explanations, GAI can provide understandable qualitative analyses with less need for human input in comparison to NLP [2]. Nonetheless, it is remarkable to see the rapid technological advances happening now. With the emergence of GAI, this study aims to compare the performance of GAI to more traditional NLP methods in developing and creating themes. This study builds on our previous work [3] and will provide a comparative analysis of the capabilities of GAI in qualitative coding versus NLP.

Literature Review

Natural Language Processing

The field of computational linguistics and NLP has a long history, with roots dating back to the 1950s [4]. Early NLP systems were limited in their capabilities and largely relied on rule-based approaches, but the development of machine learning algorithms in the 1980s and 1990s led to significant advances in the field [5]. Nowadays, NLP is a rapidly growing field that has the potential to revolutionize the way we teach and learn [6]. By enabling computers to understand and process human language, NLP can help educators identify patterns and trends in student learning, facilitate more personalized and effective instruction, and provide students with new ways to interact with educational materials [6], [7]. NLP has a wide range of applications, including language translation, text summarization, and sentiment analysis Its value comes from analyzing large amounts of text data [2]. For example, its applications have been used to analyze social media posts to track public opinion and identify trends (e.g., O'Connor [8]). In the field of education, it has been applied to the analysis of student essays to provide feedback, teamwork review analysis, and students' feedback loop [1], [3], [9]. Another application is in the generation of natural language text (e.g., machine translation systems use NLP to translate text from one language to another) [10].

In addition, it has been used to generate feedback on student writing [11] and to create personalized study materials [12]. It also can facilitate more personalized and effective instruction [13]. By analyzing large amounts of language data, NLP can help identify areas where students may be struggling and provide customized support and feedback [13]. This can help teachers tailor their instruction to the specific needs of each student, leading to more effective and efficient learning [13]. It could aid professors in grading and delivering timely feedback, allowing them to focus on other responsibilities and offer more extensive feedback to students [14]. This can help students receive the support they need to succeed and can improve the overall efficiency of the educational process [15]. Giving computers the ability to understand and process human language in a way that is intuitive and natural, has the potential to significantly increase the efficiency and effectiveness of education [15].

Along with the numerous potential applications of NLP, there are also challenges and limitations to its use. One issue is the need for large amounts of high-quality annotated data for training and evaluating its models [16]. Another one is data imbalance occurs when there are far more examples of some classes than others in a dataset used to train AI models. This issue is common in NLP tasks, including in the education domain. Manual annotation by experts can provide labeled data to train models, but this process is slow [16]. Even with some labeled data, model performance suffers from biased and uneven data distributions. Having significantly more data for some text classes makes it hard for the model to learn concepts from minority classes adequately. More balanced datasets and better sampling techniques are needed to improve model training with imbalanced education data. [16]

Artificial Intelligence and Generative AI

Artificial intelligence (AI) refers to a complex array of interdisciplinary concepts and techniques, including machine learning, natural language processing, robotics, and computer vision, that collectively enable the development of intelligent machines and systems. In his work, Grewal [17] conducted a critical conceptual analysis of the definitions of artificial intelligence, ultimately recommending the following definition: "the mechanical simulation system of collecting knowledge and information and processing intelligence of universe: (collating and interpreting) and disseminating it to the eligible in the form of actionable intelligence." [17, p. 13]. While GAI evolved from AI, its technical definition differs from standard AI. There are three fundamental categories to define GAI. Specifically, GAI is characterized as a technological approach that (i) utilizes deep learning models to (ii) produce human-like content, such as images and words, in response to (iii) multifaceted and diverse prompts, including languages, instructions, and questions [18]. In other words, AI-generated content that is indistinguishable from human content whether the human who produced the content is a novice or an expert [19].

Multiple applications of GAI have been used such as Katz et al. [20] combined text embedding models and generative text models to analyze over 1,000 career interest essays from undergraduate engineering students. They found that their model could self-evaluate the accuracy of its cluster labeling, with 86-93% agreement with human raters. Their results show NLP and LLM methods can automatically analyze unstructured text to gain insights into student experiences [20]. Another application that applied GAI in clustering labels after coupling it with NLP. The approach followed an NLP traditional method which was applied to make the clustering process of students' responses and then GAI model (GPT-3.5) labeled these clusters [21]. This approach resulted in more concise cluster labeling in comparison to other traditional NLP methods alone [21].

Additionally, as Large Language Models (LLMs) increase and rapidly develop, many organizations and researchers compete to create more powerful and advanced GAI models. These new models aim to outperform older versions [22]. GAI models come as applications or tools like ChatGPT, GitHub Copilot, and Bard to name a few. One key example is the GPT model, which has gone through versions 3, 3.5, and now 4, each with different capabilities [22]. When new GPT versions are released, they often gain new features, capabilities, and parameters compared to previous versions [22]. Also, OpenAI and other research groups constantly work to improve LLMs and other AI models. This could impact the accuracy of the information in this paragraph over time. Hence, it's important to stay updated on advances in natural language processing and AI to understand the current strengths and weaknesses of GAI and related models. This study investigates the potential application of a locally run GAI model for textual analysis. Specifically, we examine open-ended engineering students' responses to a general feedback survey asking about experiences in the rapid transition to online learning prompted by the COVID-19 pandemic [3]. The main objective is to develop themes from students' feedback responses. The overarching research question guiding this study is as follows:

RQ: How can a locally run generative AI (GAI) system be leveraged to effectively develop qualitative themes from students' feedback?

Methodology

In this study, we employed a systematic four-step strategy to derive labels/topics from the responses provided by students. The process involved Text Extraction, Text Embedding, Clustering, and Code Generation. To begin with, we utilized a locally deployed Dolphin-Mistral version 2.6 model from the *Ollama* platform for Text Extraction. This model allowed us to extract information from each student's response while ensuring data privacy and security by executing the process on our own computer. Following the extraction phase, we proceeded to the Text Embedding step. Here, we employed the UAE-Angle text embedding model [23] sourced from the HuggingFace Transformers repository. This model enabled us to embed each extracted idea, facilitating further analysis. After embedding the responses, we moved on to Clustering. Through clustering analysis, we grouped together similar ideas, which enhanced the organization and categorization of the extracted information. Finally, we engaged in Code Generation. By utilizing the clusters established in the previous step, we prompted a generative model to create codes representative of the grouped ideas. This process allowed us to generate straightforward and meaningful representations of the extracted information, completing our systematic approach to label/topic generation from students' responses.

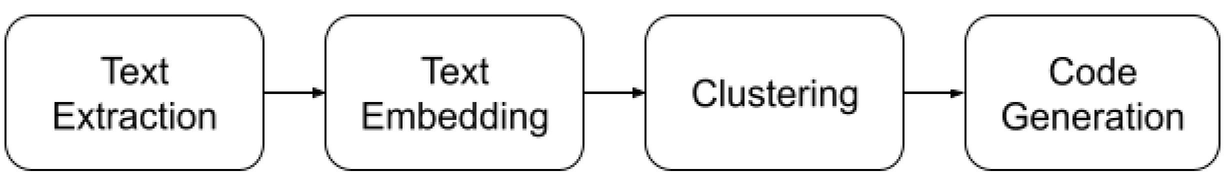

Figure 1. Study Process Overview

Data Collection

In the Spring of 2020, we gathered information through open-ended questions on a course feedback survey for a large first-year engineering class at a public research university. The survey was administered mid-semester after all classes had suddenly transitioned online due to the pandemic. The survey went beyond typical course feedback to understand how students managed the sharp change to remote learning. Participation in responding to individual questions was voluntary. For our analysis, we focused on students' narrative responses to 2 open-ended prompts about their experiences as we did in our previous work [3]. As seen in Table 1, the number of responses varied per question given the optional nature of responding.

Table 1. Questions from end-of-semester survey used in the analysis

Questions	Number of Responses
1. What could your instructor have done differently in the online transition to help you learn?	1,066
2. What did your instructor do during the transition to online learning that helped you to stay engaged in the course?	1,197

Analysis and Results

In the analysis, following the processes of clustering and code generation, A human evaluator was tasked with assessing the results to ensure alignment, conducting a terminology comparative analysis between the terminology used by the AI and that used by humans to determine similarity or dissimilarity of topics. It was observed that the codes exhibited a high degree of similarity, with minor distinctions noted. For instance, in response to question 2, "Assignments explanations" emerged as a recurrent theme in the human-generated codes, though in a slightly varied form in the generated codes. The GAI method consistently reflected a thematic focus on "Feedback" with several sub-topics identified under this umbrella, including frequent and timely feedback. Upon reviewing the labels generated by the generative model, frequent labels were assigned to a main topic. Subsequently, the process of developing these main topics entailed utilizing GPT-3.5, with humans reviewing the main labels to ensure their accuracy and alignment with the original labels. The main topics from questions 1 and 2 are presented in Table 2.

Table 2. Main topics for Q1 and Q2

Q1 Main Topics (n=8)	**Q2 Main topics (n=9)**
Online Learning Experience	Instructor Support
Interactivity and Engagement	Communication
Communication	Engagement

Instructor Support	Course Flexibility
Feedback	Teaching Methods
Instructions and Resources	Assignments
Flexibility	Course Resources
Teaching Methods	Course Structure
	Student Support

Visualizing results from models can be achieved through the use of treemaps, which help organize and display textual data efficiently [24]. Treemaps present hierarchical structures in a visually intuitive manner, making them suitable for analyzing text-based information [24]. Through the use of treemaps, we can easily show the relationships and proportions between different categories or themes present in the textual data. For our results, Figures 1 and 2 show the main topics from the generated codes (e.g., topics).

Figure 2. Topics and Sub-topics For Question 1 (Main topic N=8)

For Figure 2. we looked into the generated codes from the model for question 1 which included main topics, Instructor Support, Communication, Engagement, Course Flexibility, Teaching Methods, Assignments, Course Resources, Course Structure, and Student support. These topics reflect students' responses regarding the elements that the instructor could have implemented differently during the transition to online learning to assist them in their learning process. Moreover, the squares' sizes and spacing are based on the number of sub-topics within each main topic - the more sub-topics, the larger the square.

Figure 3. Topics and Sub-topics For Question 2 (Main topic N=9)

For Figure 3, students' topics for Question 2 about the online learning experience are visualized through sectors like interactivity and engagement, communication, instructor support, feedback, instructions and resources, flexibility, and teaching methods. These segments allow for a quick assessment of learner engagement, communication effectiveness, support availability, feedback quality, resource accessibility, flexibility options, and teaching methodologies. Representing these elements shows students recommendations to improve the online learning experience, Tree maps are an effective method for visualizing grouped topics. Categorization aids in displaying various sub-topics, highlighting differences in the number of sub-topics within each category. Furthermore, an additional step aimed at reducing the number of subtopics, as they sometimes overlap with each other. For instance, in question 1, the main topic is "Communication" with the following subtopics: Online Interaction, Clear Communication, Consistent Communication, Encouragement of Open Expression, Effective Communication, Maintaining Effective Communication, Regular Communication, Prompt Response, Clear Instructions, Email Communication, Quick Email Communication, and Timely Communication. These can be consolidated under communication strategies. Please view the appendix for the rest of the merged sub-topics To understand how the GAI main topics relate to the human-generated qualitative themes, we can look at Figure 4. The GAI main topics essentially serve as overarching categories that encompass multiple related sub-topics (Figures 3 and 2). These main topics can then be mapped to the corresponding human qualitative themes. For example, the "Online Learning Experience" main topic generated by the GAI incorporates sub-topics such as online learning difficulties and online learning challenges, which align with the human themes around creating virtual classes and lecture recordings to cope with the challenges and difficulties of transitioning online. For more details on the qualitative developed human topics please see our previous work [3].

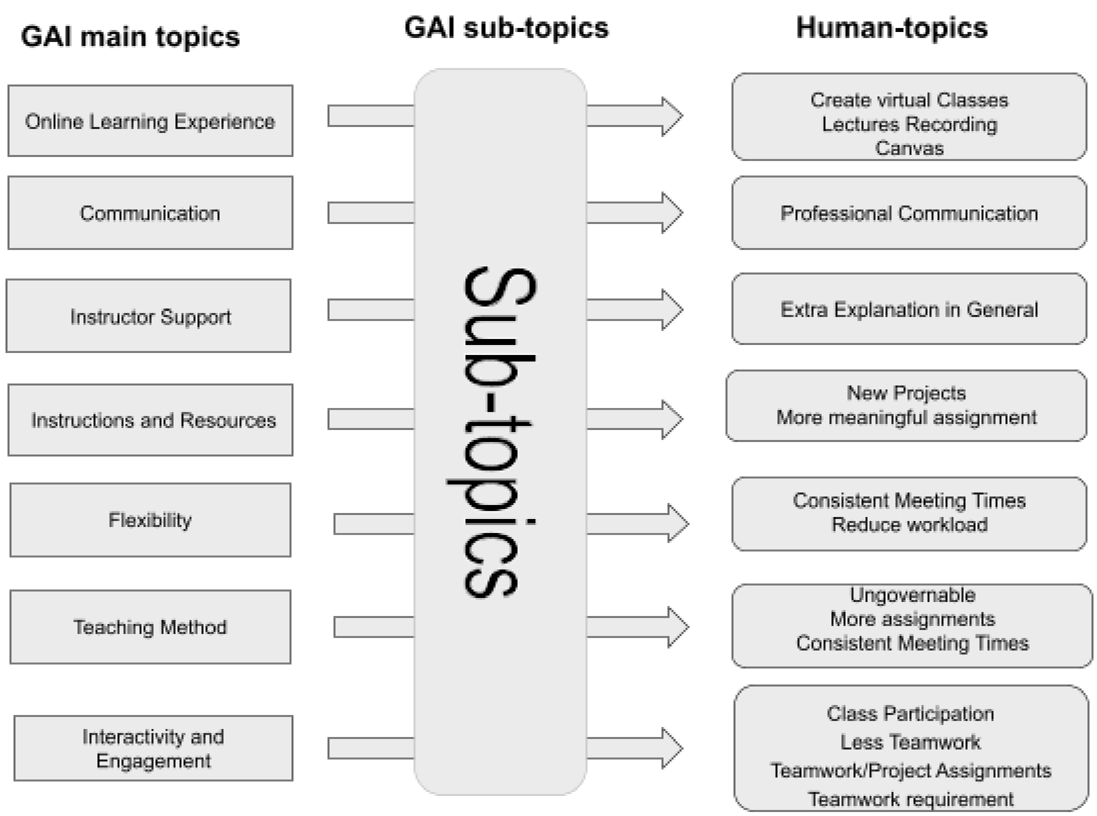

Figure 4. Q1 GAI-generated topics mapped to human-generated topics (themes)

The GAI and the manual qualitative coding approach identified several main topics related to the online learning experience during the pandemic, suggesting alignment in capturing core student concerns. Both highlighted topics around interactivity/engagement, communication, instructor support, feedback, instructions/resources, flexibility, and teaching methods. For example, the "Interactivity and Engagement" topic from the GAI aligned with human codes around class participation, teamwork, and project assignments - all factors impacting how engaged students felt. The "Feedback" topic also directly matched between GAI and humans, underscoring its importance. Some human topics mapped clearly onto the more broadly defined GAI topics, such as "Professor communication" rolling into the general "Communication" category. The other topic"New Project Instructions and Resources" human topic fits within the broader "Instructions and Resources" GAI topic. In other cases, the GAI topical categorization differed slightly in framing from the human versions while still capturing related ideas or the underlying meaning of the topic. For instance, the GAI's "Flexibility" topic appears to encompass human topics around reducing workload, consistent meeting times, and general flexibility accommodations. While the terminology and categorization structure differed somewhat, there was strong overlap and agreement in the core topics and concerns raised by students in their feedback on the online learning transition. Question two followed the same process please see Figure 5 and the Appendix for themes and sub-topics.

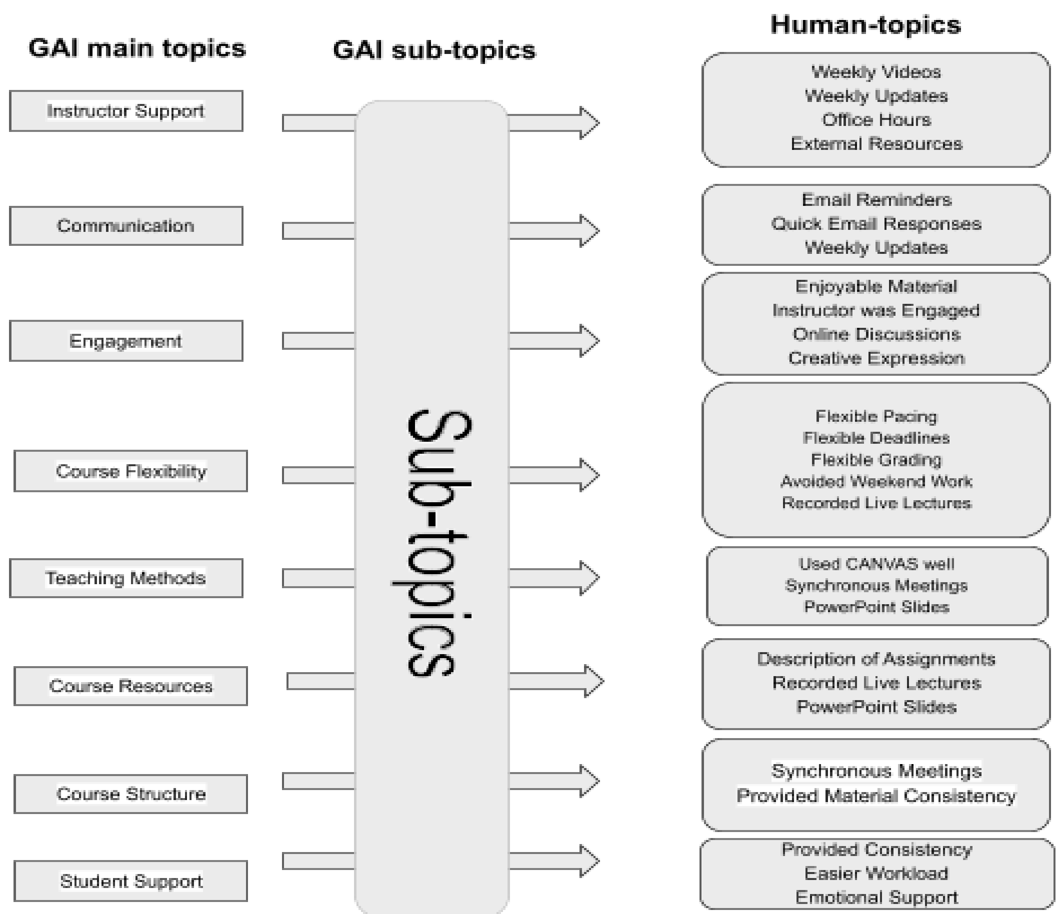

Figure 5. Q2 GAI-generated topics mapped to human-generated topics (themes)

Discussion

The systematic approach employed in this study aimed to derive meaningful insights from students' responses to open-ended prompts regarding their experiences with online learning during the COVID-19 pandemic [3]. To answer the research question, utilizing a combination of text extraction, embedding, clustering, and code generation techniques, we were able to filter a large amount of qualitative data into manageable themes and topics. The use of a locally deployed Dolphin-Mistral version 2.6 model from the Ollama platform for text extraction provided a secure and privacy-conscious means of processing sensitive student data. This approach ensured compliance with data protection regulations while allowing for comprehensive analysis. Additionally, employing the UAE-Angle text embedding model facilitated the transformation of extracted ideas into a format conducive to clustering and further analysis [23]. The evaluation of the generated codes against human-generated counterparts [3] revealed a high degree of similarity, indicating the effectiveness of the automated clustering and code generation processes. While minor distinctions were noted, particularly in the formulation of themes, the overall thematic focus remained consistent. For instance, the prominence of "Feedback" as a recurrent theme underscored the significance of timely and constructive feedback in facilitating student engagement and learning during the transition to online instruction. The use of treemaps to visualize the main topics and sub-topics derived from students' responses provided an all-around overview of the key themes emerging from the data. Also, treemaps for textual data can be used to organize and hierarchically display textual data. They facilitate the identification of relationships and proportions between different categories or themes. This visual representation enhanced the interpretability of the results, allowing for an understanding of students' comments on the online learning experience.

Furthermore, to fully answer our research question, the specific differences noted between the GAI-generated codes and the human-generated codes highlight some of the potential strengths and limitations of using generative AI for qualitative coding tasks. While the overall thematic focuses were highly aligned, the GAI system at times used slightly different terminology or framed themes in a more concise manner compared to the human analyst or researcher. These differences have implications for the qualitative analysis process. On the other hand, the GAI's ability to quickly identify and summarize core themes from large amounts of text data demonstrates its potential for significantly accelerating and scaling up the coding process. The models can extract high-level insights that may be difficult for human analysts to see when dealing with thousands of individual responses. However, the terminology distinctions emphasize the importance of incorporating human oversight and validation into the analysis process. The GAI outputs are viewed as an initial pass to rapidly acquire potential themes, but to properly refine how the themes are framed, and named, and how to hierarchically organize the themes to best fit the research context and goals. In this process feedback and revision from an expert is still required. With this human-in-the-loop step, there is a chance that important nuances could be recovered or themes could be over-generalized [19]. To leverage GAI most effectively for robust qualitative analysis, a best practice would be to incorporate the model's outputs into an iterative coding process that combines the AI's speed and ability to synthesize large datasets with human validation, context-specific refinement of the qualitative coding scheme, and sense-making of the results. The GAI codes can serve as a rich starting point for analysis, but human analysts should thoroughly review, reorganize, and clarify the coding outputs to ensure precise alignment with the research aims. Additionally, documentation and communication of the specific AI model employed, its training data, known strengths/limitations, and the human refinement process used are critical to ensure methodological transparency and replicability of the qualitative coding approach. We would like to point out that leveraging an in-house, locally deployed generative AI model like Dolphin-Mistral version 2.6 enabled us to gain granular insights into students' feedback on their online learning experiences during the pandemic. This approach avoided potential costs associated with external GAI services, while also ensuring data privacy and addressing concerns around exposing sensitive student information to third-party platforms or cloud-based AI systems which is a major concern in the community. The ability to conduct such nuanced qualitative analysis in a secure, self-hosted environment, without compromising the depth of insight, shows the promising potential of localized AI models for institutions dealing with sensitive data.

Conclusion

We conclude that the GAI model effectively summarized and organized the key themes in a similar manner to manual human coding. The four-step strategy used in this study proved effective in extracting, organizing, and synthesizing qualitative data from students' responses to open-ended prompts regarding their experiences with online learning. It leveraged both automated and human-driven processes, we were able to gain valuable insights into the challenges and opportunities associated with remote instruction during unprecedented times. The findings of this study are consistent with our prior research efforts. However, the primary objective of this study is to evaluate the efficacy of GAI in generating themes as compared to a natural language processing approach. In this context, Ollama's Dolphin-Mistral version 2.6 demonstrated proficiency in categorizing the codes while ensuring the privacy and security of student information, without exposure to cloud-based AI or third-party online platforms.

Limation and Future Work

One of the notable challenges encountered in the analysis process was the presence of overlapping sub-topics within the main themes. While the use of GAI facilitated the merging of related sub-topics, human supervision was necessary to ensure the accuracy and coherence of the merged categories. Moving forward, we will be exploring effective methods for quantifying qualitative codes associated with the identified themes in this study. Future research could explore alternative approaches (e.g., GPT-4) to streamline the clustering and code generation processes, potentially leveraging advanced natural language processing techniques to automate the identification and consolidation of overlapping themes.

References

[1] A. Alsharif, A. Katz, D. Knight, and S. Alatwah, "Using Sentiment Analysis to Evaluate First-year Engineering Students Teamwork Textual Feedback," in *2022 ASEE Annual Conference & Exposition*, 2022. Accessed: Nov. 28, 2023. [Online]. Available: https://peer.asee.org/41460.pdf

[2] R. S. Baker and P. S. Inventado, "Educational Data Mining and Learning Analytics," in *Learning Analytics: From Research to Practice*, J. A. Larusson and B. White, Eds., New York, NY: Springer, 2014, pp. 61–75. doi: 10.1007/978-1-4614-3305-7_4.

[3] A. Katz, M. Norris, A. M. Alsharif, M. D. Klopfer, D. B. Knight, and J. R. Grohs, "Using Natural Language Processing to Facilitate Student Feedback Analysis," presented at the 2021 ASEE Virtual Annual Conference Content Access, Jul. 2021. Accessed: Dec. 07, 2023. [Online]. Available: https://peer.asee.org/using-natural-language-processing-to-facilitate-student-feedback-analysis

[4] D. Jurafsky and J. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, vol. 2. 2008.

[5] C. Manning and H. Schutze, *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.

[6] M. C. Chen *et al.*, "Deep Learning to Classify Radiology Free-Text Reports," *Radiology*, vol. 286, no. 3, pp. 845–852, Mar. 2018, doi: 10.1148/radiol.2017171115.

[7] K. Kukich, "Techniques for automatically correcting words in text," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 377–439, Dec. 1992, doi: 10.1145/146370.146380.

[8] B. O'Connor, R. Balasubramanyan, B. Routledge, and N. Smith, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series," *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 4, no. 1, Art. no. 1, May 2010, doi: 10.1609/icwsm.v4i1.14031.

[9] T. Prinsloo and J. P. Van Deventer, "Using the Gartner Hype Cycle to Evaluate the Adoption of Emerging Technology Trends in Higher Education – 2013 to 2016," in *Emerging Technologies for Education*, T.-C. Huang, R. Lau, Y.-M. Huang, M. Spaniol, and C.-H. Yuen, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017, pp. 49–57. doi: 10.1007/978-3-319-71084-6_7.

[10] D. Patil, S. B. Chaudhari, and S. Shinde, "Novel Technique for Script Translation using NLP: Performance Evaluation," in *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Mar. 2021, pp. 728–732. doi: 10.1109/ESCI50559.2021.9396969.

[11] R. Delmonte, "Linguistic Knowledge and Reasoning for Error Diagnosis and Feedback Generation," *CALICO J.*, vol. 20, no. 3, pp. 513–532, 2013, doi: 10.1558/cj.v20i3.513-532.

[12] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education – where are the educators?," *Int. J. Educ. Technol. High. Educ.*, vol. 16, no. 1, p. 39, Oct. 2019, doi: 10.1186/s41239-019-0171-0.

[13] D. McHugh, S. Shaw, T. R. Moore, L. Z. Ye, P. Romero-Masters, and R. Halverson, "Uncovering themes in personalized learning: Using natural language processing to analyze school interviews," *J. Res. Technol. Educ.*, vol. 52, no. 3, pp. 391–402, Jul. 2020, doi: 10.1080/15391523.2020.1752337.

[14] S. Knight, S. Buckingham Shum, P. Ryan, Á. Sándor, and X. Wang, "Designing Academic

Writing Analytics for Civil Law Student Self-Assessment," *Int. J. Artif. Intell. Educ.*, vol. 28, no. 1, pp. 1–28, Mar. 2018, doi: 10.1007/s40593-016-0121-0.

[15] Dr. Khaled, "Natural Language Processing and its Use in Education," *Int. J. Adv. Comput. Sci. Appl.*, vol. 5, no. 12, 2014, doi: 10.14569/IJACSA.2014.051210.

[16] T. Shaik *et al.*, "A Review of the Trends and Challenges in Adopting Natural Language Processing Methods for Education Feedback Analysis," *IEEE Access*, vol. 10, pp. 56720–56739, 2022, doi: 10.1109/ACCESS.2022.3177752.

[17] P. Grewal, "A Critical Conceptual Analysis of Definitions of Artificial Intelligence as Applicable to Computer Engineering," *IOSR J. Comput. Eng.*, vol. 16, pp. 09–13, Jan. 2014, doi: 10.9790/0661-16210913.

[18] W. M. Lim, A. Gunasekara, J. L. Pallant, J. I. Pallant, and E. Pechenkina, "Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators," *Int. J. Manag. Educ.*, vol. 21, no. 2, p. 100790, Jul. 2023, doi: 10.1016/j.ijme.2023.100790.

[19] A. Alsharif, D. Knight, and A. Katz, "The Evolution of Technology in Education and the Emerging Role of Generative A," in *In X. Lin, R. Y. Chan, & S. Sharma (Eds.), ChatGPT and global higher education: Using artificial intelligence in teaching and learning*, STAR Scholars Press, 2024 (in Publication).

[20] A. Katz, U. Shakir, and B. Chambers, "The Utility of Large Language Models and Generative AI for Education Research." arXiv, May 29, 2023. Accessed: Jun. 05, 2023. [Online]. Available: http://arxiv.org/abs/2305.18125

[21] Y. Gamieldien, J. M. Case, and A. Katz, "Advancing Qualitative Analysis: An Exploration of the Potential of Generative AI and NLP in Thematic Coding." Rochester, NY, Jun. 21, 2023. doi: 10.2139/ssrn.4487768.

[22] OpenAI, "GPT-4 Technical Report." arXiv, Mar. 27, 2023. Accessed: Jun. 05, 2023. [Online]. Available: http://arxiv.org/abs/2303.08774

[23] X. Li and J. Li, "AnglE-optimized Text Embeddings." arXiv, Nov. 08, 2023. doi: 10.48550/arXiv.2309.12871.

[24] Y. Tu and H.-W. Shen, "Visualizing Changes of Hierarchical Data using Treemaps," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 6, pp. 1286–1293, Nov. 2007, doi: 10.1109/TVCG.2007.70529.

Appendix

Table 3. Q1 Main topics and Mereged Sub-topics

Main Label	Merged Sub-Topics
Online Learning Experience	Bettering Online Learning Experience
	Positive Online Learning Transition
	Challenges with Online Learning
	Pandemic-Related Challenges in Online Learning
	Time management and issues
Interactivity and Engagement	Active/Interactive Learning
	Addressing Lack of Engagement
Communication	Clear/Improved Communication
	Enhancing Online Interaction and Communication
	Addressing Communication and Interaction Gaps
Instructor Support	Instructor Accessibility and Support
	Instructor Improvements in Online Learning
Feedback	Frequent and Timely Feedback
	Insufficient Quality Feedback
	Low Feedback Quality
Instructions and Resources	Clear/Explicit Instructions
	Improving Course Resources
Flexibility	Flexible Course Structure
Teaching Methods	Hands-On Learning and Synchronous Opportunities
	Instructional Tools and Aids

Table 4. Q2 Main topics and Mereged Sub-topics

Main Topic	Merged Sub-Topics
Instructor Support	Instructor Support and Engagement:
	Adpative Supportive Teaching Methods

Communication	Communication Strategies
	Consistency and Updates
	Media Integration and Feedback
Engagement	Engagement Techniques
	Interactive Activities
	Active Learning
	Improve Content Engagement
	Skill Application
	Virtual Engagement
	Peer Review Engagement
Course Flexibility	Flexible Deadlines
	Flexible Scheduling
	Teaching Method Flexibility
	Teaching Methods (Online)
Teaching Methods	Better Video Usage
	Technology Training
	Visual Aids
Assignments	Assignment Types
	Group Work
	Assignment Difficulty
Course Resources	Online Resources
	Accessible Resources
	Digital Resources
Course Structure	Weekly Structure
	Consistent Pacing and Schedule
	Communication and Updates
Student Support	General Support
	Mental Health and Emotional Support
	Resource and Notification Support