Exploring NLP-based Methods for Generating Engineering Ethics Assessment Qualitative Codebooks

Ashish Hingle
Information Sciences and Technology
George Mason University
Fairfax, USA
ahingle2@gmu.edu

Andrew Katz
Engineering Education
Virginia Polytechnic Insitute and State University
Blacksburg, USA
akatz4@vt.edu

Aditya Johri
Information Sciences and Technology
George Mason University
Fairfax, USA
johri@gmu.edu

2023 IEEE Frontiers in Education Conference (FIE) | 979-8-3503-3642-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/FIE58773.2023.10342985

Abstract— This Full Research paper presents a comparison of two codebook generation methods using natural language processing (NLP): a human and NLP collaboration method and a fully automated NLP method (referred to as Human-NLP and Auto-NLP, respectively). Codebook generation serves as a preliminary step in most qualitative projects, and using NLP as a tool can help support the analysis and efficiency of the researcher. By utilizing NLP in the early stages of codebook generation, there are opportunities for detailed and productive gains when working with large corpora of textual data. Using NLP at this stage also allows the researcher to make sense of any outputs generated through automated means rather than simply accepting the output as it is. The outcome of both methods tested in this work will be used to evaluate and apply the codes across a large dataset. The Human-NLP method involves generating the initial themes using a large-language model (LLM), and the researcher revises the codebook further. The Auto-NLP method involves generating three rounds of codes, summarizing the codes in each until a saturation level has been reached through the overarching themes. The dataset used for this study comes from an analysis of students' perception and recognition of ethical concepts after participating in a semester-long course focused on ethics, society, and technology. The course introduced students to traditional ethics topics, such as those around engineering disasters, but also explored developing topics, such as facial recognition, dataset bias, and the impact of technology on the global food supply. We collected data between fall 2020 and 2022 from six (6) iterations of a semester-long course. A total of 210 student responses to the question – what did this course teach you about ethics – were analyzed. The results from both Human-NLP and Auto-NLP methods were promising in the level of detail summarized and the similarity of themes across the data. Eight (8) themes were finalized through the Human-NLP method, and twelve (12) were generated through the Auto-NLP method. We present a discussion exploring these themes and the limitations of using these methods.

Keywords— engineering ethics, natural language processing, codebook generation, generative AI

I. INTRODUCTION

Natural language processing (NLP) techniques have continued to be used as promising tools to support students and faculty in education-related tasks [1], [2]. Other related research areas, such as natural language understanding (NLU) and systems built with the advances in these fields, such as chatbots or conversational agents, have also continued to grow in popularity and support [3], [4]. NLP and NLU methods are often used to analyze large text corpora and allow for insights to be generated with trained and untrained data. Generative artificial intelligence (GAI) represents a new subset of artificial intelligence (AI) that relies on NLP and allows for systems to use Large Language Models (LLMs) and are provided with a dataset that the system uses to accomplish other tasks, such as clustering or grouping similar text, or generating human-like text. In recent months, the popularity of *ChatGPT*, *Google's Bard*, and other such generative tools have headlined as productivity assistants [5], [6]. Unsupervised learning, where few labels are given in the test data, is becoming increasingly popular and foundational to LLMs' functions.

The popularity of GAI has also spread to students, faculty, and administrators in the classroom, and the widespread and often hidden use has raised many important questions and challenges about the social implications of working with AI systems [7]. Should interaction with these systems be encouraged, and to what extent? What level of training is needed? How do institutions ensure all students equitably have access to these systems? From a research perspective, how consistent should the results of working with these LLMs be before they are used as learning tools? While the technology is novel, these questions have been a part of ethics discussions across domains, including engineering and computing, for much time. Therefore, there is a continued need to explore how institutions, researchers, and faculty can demonstrate model behaviors and technology implementation.

Ethics education is considered an important part of students' learning experience within engineering and computing majors [1]-[5]. Ethics-related courses are incorporated in most degree programs, especially in the USA, to meet the accreditation requirements. The importance of ethics education in these domains is also mirrored by including professional codes of ethics in professional engineering and computing organizations [8]–[10]. These codes all address components of ethical behavior, including describing how members should behave with integrity and concern for people's safety and privacy while maintaining access to the profession.

In this work, our goal is two-fold: 1) to use advanced computational methods, specifically LLM-based techniques, to make sense of a large corpus of data collected from students, and

2) to explore and document themes of what students learn as part of a semester-long course focused on ethics within a larger framework of technology and society. Through this work, we hope to contribute to both engineering and computing ethics education literature and methodological advances in engineering education research. Specifically, in what instances and how are LLM-based methods and techniques useful?

Our use of LLM-based methods is driven by both technological advances that allow better analysis of language-based data, in our case, student responses, and also by the opportunity this provides engineering education research generally and ethics education scholars to be able to scale up student assessment. The possible scale of NLP implementation is significant, and this may provide time-saving and resource-efficient techniques in dealing with large corpora of data that are normal to come by for any course in a particular term. Furthermore, we see this as a unique capability for instructors to better understand the outcomes of their teaching over a longer period. At a practical level, this research answers what students learn from an ethics course in their own view, including topical ethics-related issues and what aspects of the course assisted with their learning. It also provides a breakdown of a process that can be used to look at similar data collected from students.

II. Literature Review

A. Using NLP automated methods for analysis

Natural language processing can be used to explore and analyze human textual prompts and generate a suitable response. NLP is a powerful tool when combined with large datasets, and in the case of ChatGPT, datasets related to Internet activity [11]. NLP as a methodology is already widely used in education [12] and engineering and computing education research [13]. The use of NLP with ethics education content is on the rise. One study evaluated ethics-related topics discussed on Stack Exchange and other sites. Activities with ethical dilemmas and questions in them and beyond academia are present, including web scraping, licenses, etc. [14]. In others, researchers have explored syllabi and curricula to understand what is being taught on a specific topic, such as AI Ethics [15].

However, opinions are divided on using and accepting these LLMs across domains. In education, the challenges and concerns of using these systems are academic integrity, lack of regulation, privacy concerns, biases, gender and diversity, accessibility, and equity [11]. Technology students are affected twofold – they must learn to use and build with these tools while also navigating having these tools used on them through their instructors, altering syllabi, and serving as tutors or evaluators. As a result of this dual nature of interaction with these tools, technology students, though this can be extended to all students, must have a deliberate and deep understanding of the ethical landscape of GAI and the data that supports it.

Using automated results raises some questions of reliability, especially due to the obscured nature of the LLM. One area of concern around using automated techniques such as NLP is the reproducibility of results. Even when evaluated under the same conditions using the researcher's own published data, the results from NLP papers cannot be replicated with even moderate success [16].

The use of computational techniques to assist qualitative researchers has expanded in the last decade, especially NLP methods. Many such elements are commonly included in popular qualitative analysis software as well. Some researchers argue that these computation-based methods are not meant to make qualitative research more quantitative or positivist but are applied to help discover information within large amounts of data that might assist with interpretation [17]. The advantage of some of these approaches is that they are "neutral" and applicable regardless of domain size. They reduce the time required for the analysis and can aid the discovery of themes that might not otherwise be detected [18]. Within engineering education, these methods have been applied in conjunction with a theoretical lens to demonstrate their efficacy for theory-driven data analysis [19].

III. Research Method

In this section, we outline our research question, data collection, course description, and technologies/software used. We also describe both Human-NLP and Auto-NLP methods.

A. Research Question

This research presents analyses of a survey administered at the end of the semester in a technology ethics course for engineering and computing students who have largely not been exposed to ethics education. The overall research question for this work is:

RQ: *How effective are NLP-supported methods for generating a qualitative codebook?*

To explore the effectiveness of these methods in generating a qualitative codebook, our research question across the data is: What aspects of technology ethics do students recognize after participating in an ethics-focused course? While the results of this work may not be generalizable due to the course's specifics, the findings highlight technology students' perceptions of ethics and an experimental technique of analyzing participant responses.

This work explores using NLP-assisted hybrid methods to analyze themes from student data in a course on technology ethics. In this section, we will describe the course, data collection, and the analysis procedure.

B. Data Collection

The data for the study were collected over four (4) semesters between fall 2020 and 2022 from six (6) with six sections or iterations of the semester-long course. The content, resources, and activities across the course were kept similar, although there were changes in the specific cases introduced. However, the overall topics and themes across the course iterations remained the same. In total, 212 students participated in the data collection (30-40 students per iteration).

During the final week of the course, a survey was conducted among the students in the course. The survey included questions about the student's experiences in the course. The final question on the survey, which was used as the primary question for this work, asked students, after all the activities in the course, "Overall, what did this course teach you about ethics?" The question was left open-ended to allow students to focus on

whatever element they wanted to talk about, and this question served as their final reflection activity.

C. Course Description

The course was a 3-semester credit course offered at a large US public institution. The course was designed to help students better understand and appreciate technology's impact in a global context. This course addressed specific technologies such as artificial intelligence, algorithms, complex information systems, and data-driven procedures. In introducing the overall conversation about technology, the course also addressed specific conversations on fairness, transparency, bias, misinformation, trust, and solidarity in the interaction between humans and technology systems. The course brought in technology perspectives through the resources provided to students worldwide. The course learning outcomes were designed to help students explore complex topics related to placing technology's role in supporting human work, and they were given the tools and space to develop teamwork, critical thinking, and communication skills through the course activities. The specific learning outcomes were to understand the following:

1) *The rapid rate of technological change and its effects on societies.*
2) *The role of IT in globalization and the changing nature of work, governance, communication, and privacy in creating a global civil society and facilitating the work of NGOs.*
3) *Professional codes of ethics, ethical decision-making models, and processes.*
4) *The effects of IT reliance on the environment and global health.*

Students in multiple programs (cybersecurity, information technology, data analytics) were required to take this course as a core requirement covering professional ethics in and beyond the classroom. The course content was developed to cover the breadth of student domain knowledge, and a variety of topics related to technology's role in the global society, the ever-changing nature of work, and the ethical decision-making processes were included. The course was presented in a flexible format to allow for conversations about timely topics such as generative AI, large-language models, and digital twins, which continued gaining ground during the data collection period. To ensure the course material was updated easily, peer-reviewed articles, videos, and other online resources were used.

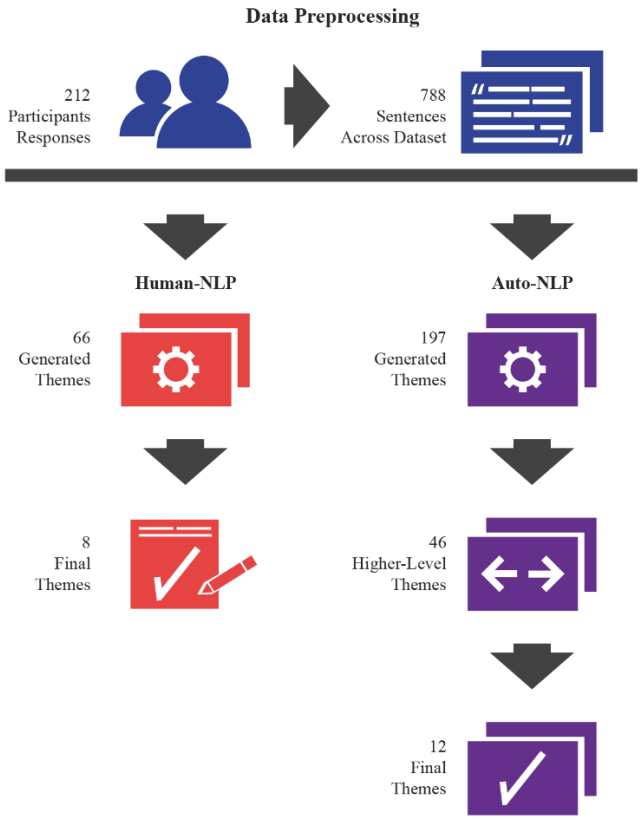
An overarching topic was covered every three weeks to address the breadth of the ethics content. These topics were:

1) *Ethical Responsibility in Professional Engineering*
2) *Implications of Biometric and Facial Recognition*
3) *Algorithms in Everyday Activities*
4) *Environmental Sustainability*

Over 15 weeks, students were provided with resources on each topic, including lectures, videos from professionals and researchers, peer-reviewed articles, and news articles. The focus was bringing together peer-reviewed work for students to engage with and mirroring the knowledge through what they would come across outside the classroom (i.e., news articles and blog posts). Students were also given assignments to reflect on their learning through discussions and writing tasks. Students were given the option to participate in this data collection, and the Institutional Review Board approved the study.

Students were also given a case study to further situate their knowledge from reading to practice. The case studies would serve as the basis for the peer activity in the course – role-playing the different participants and their perspectives on the case. As role-playing in the classroom can be new to many students, they were provided with examples of exemplary role-play scenarios that they could watch and understand the activity. Following their engagement with the materials, students participated in the role-play activity in synchronous online breakout rooms through the campus-mandated learning management system. The authors monitored the rooms to help with any concerns and facilitate a smooth session in case of any unforeseen scenarios. After the activity, students debriefed and answered post-activity questions, which encouraged them to summarize their experience from the role of their perspective but also on behalf of their own perspective. The role-play activity participation and assignments together comprised a significant portion of their final grade in the course. Overall, the role-play activities were well received by participants, and analysis of their participation has been shown to increase awareness of ethical dilemmas and expand student learning regarding applied technology ethics [20]–[22].

Fig. 1. Overview of Codebook Generation Procedures



D. Technologies and Software Used

For both Human-NLP and Auto-NLP methods, the technical and software specifications were kept identical. Python was used as the scripting language to conduct data preprocessing and interact with the LLM.*Anthropic's Claude v-1 model* [23] was used as the LLM for the summarization tasks. For text embeddings, we used the *all-mpnet-base-v2* model from the sentence transformers Python package [24] and agglomerative clustering with ward linkage as implemented in the *scikit-learn* Python package.

In the following sections, we describe the process used for each method.

E. Human-NLP Analysis Method

The Human-NLP method uses a human-in-the-loop approach to generate an initial set of themes, which the researchers revise to create the final qualitative codebook. This method includes a built-in human check to ensure that the themes being generated by the models make sense from the first set of results.

TABLE I. HUMAN-NLP METHOD FINAL CODEBOOK

Overarching Theme	*Theme and Description*	*Count*	*Example Quotes*
Learning about Ethics, Reasoning, and Considerations	**Learning about ethics concepts** Highlighting different ethical theories being explored through the resources.	80	*"Ethics was an undefined box in my mind that only included, "Be good and don't take bribes." This class taught me different definitions of ethics." – Student 1*
	Connecting experiences in the course to real-life Highlighting the transfer of learning from their course experiences to their personal experiences.	165	*"Companies I might work for could be involved in something unethical, or even I could get caught up in something. Knowing how these case studies play out develops a sense of awareness." – Student 1*
	The overall course learning experience Highlighting aspects of the course that assisted in learning.	51	*"The core foundations of this class made the experience of learning ethics engaging, such as watching the videos and answering free-text formed questions after to really help us remember the main point of the video." – Student 2*
	Using case studies and role-plays Describing how the use of case studies and role-playing activities affected their learning.	212	*"The scenarios which were given to us were all clear and gave me a firm understanding of how I needed to play my role, given the particular scenario." – Student 3*
Ethics Connection to Technology and Society	**Tech ethics considerations** Discussions on bias, discrimination, privacy, safety, transparency, standards, and lack of regulations.	98	*"Lastly, in a world that is beginning to be even more interconnected with technology than we could have ever imagined, we should consider our safety and our privacy as it is something that we have a right to control. – Student 4*
	Connection between technology and ethics Describing how technology is affected by ethics across different domains.	106	*"Predictably, ethics are complex, but how technology complicates them was not something I put enough thought into at first. Technology solves many material problems, but seems to create many more ethical problems via the drawbacks." – Student 5*
Organizations' Responsibility to Society	**Organizational values** The need for organizational ethics, including topics like accountability, social responsibility, and trust.	38	*"There has to be a balancing act between the company leadership and the government for companies that provide services to the public and audits to make sure that their practices and intentions are ethical with the safety of the public in mind." – Student 6*
	Consequences (or lack) of actions Often in association with the readings, videos, and resources provided.	101	*"There are many factors that you have to consider when solving an issue, and even the most perfect answer may have its flaws and repercussions that follow." – Student 7*

TABLE II. AUTO-NLP METHOD FINAL CODEBOOK

Overarching Theme	*Theme and Description*	*Count*	*Example Quotes*
Learning about Ethics, Reasoning, and Considerations	**The importance and application of ethics** Ethics should be carefully considered and applied in all areas of business and society.	190	*"Companies I might work for could be involved in something unethical, or even I could get caught up in something. Knowing how these case studies play out develops a sense of awareness." – Student 1*
	Student learning and growth Codes indicating students gained knowledge, skills, or experienced personal development through the ethics course.	20	*"Overall, this course has taught me more about ethics in such a short amount of time as compared to any other course or curriculum." – Student 8*
	Learning through iteration Gaining knowledge through repeated cycles of progress, setbacks, and renewed progress.	1	"We can progress, recede a little, and progress again." – Student 9
	Developing ethical understanding through exposure and application Gained a nuanced comprehension of ethical issues by examining diverse perspectives and concepts and applying them to real-world situations.	37	*"The scenarios which were given to us were all clear and gave me a firm understanding of how I needed to play my role, given the particular scenario." – Student 3*
	Developing an understanding of ethics through experience and reflection Expressions indicate insight into ethics by discussing real-world examples and reflecting on how their views have evolved.	197	*"Ethics was an undefined box in my mind that only included, "Be good and don't take bribes." This class taught me different definitions of ethics." – Student 1*
Ethics Connection to Technology and Society	**Ethical considerations in research and tech** Recognition of the need to address issues of bias, privacy, transparency, and unintended harm in research and the development of algorithms and AI systems.	32	*"Lastly, in a world that is beginning to be even more interconnected with technology than we could have ever imagined, we should consider our safety and our privacy as it is something that we have a right to control. – Student 4*
	Societal implications and ethics of technology Technology development and use have wide-ranging effects on society that necessitate consideration of ethical principles to guide responsible innovation.	216	*"Predictably, ethics are complex, but how technology complicates them was not something I put enough thought into at first. Technology solves many material problems but seems to create many more ethical problems via the drawbacks." – Student 5*
Organizations' Responsibility to Society	**Organizational and personal responsibility** The duty of individuals and groups to establish, enforce, and uphold ethical standards through oversight, communication, and accountability.	44	*"There has to be a balancing act between the company leadership and the government for companies that provide services to the public and audits to make sure that their practices and intentions are ethical with the safety of the public in mind." – Student 6*
	Holistic ethical reasoning Approaching ethical issues by considering the consequences of one's actions and decisions on society in a comprehensive, equitable way.	36	*"There are many factors that you have to consider when solving an issue, and even the most perfect answer may have its flaws and repercussions that follow." – Student 7*
Attitudes Toward the Future of Technology	**Skepticism towards knowledge and authority** Expressions of distrust, anxiety, or cynicism related to accepting knowledge and wisdom from those in positions of power or influence.	9	*"The idea that we may not know we are being watched or used as guinea pigs, but absolute could be without out our knowledge in certain situations." – Student 10*
	Work-life balance The importance of maintaining a balance between one's professional and personal life.	3	*"Work and happiness need to be balanced as opposed to the old model of work." – Student 11*
	Exaggerated positivity Statements that convey an inflated level of support, enthusiasm, or praise for an idea, concept, or individual.	2	*"In Glenn Greenwald's TED Talk, literally everything that this man said was true." – Student 12*

The steps for the Human-NLP process can be summarized as follows:

1) *Data Preprocessing: Anonymize and remove any identifiers from the participant dataset.*
2) *Data Preprocessing: Tokenize the corpus into sentences.*
3) *LLM Interaction: Using Python, request summarization, and generation of themes.*
4) *Human Interaction: Explore and revise the generated output of themes/codes.*
5) *Human Interaction: Create a revised list of themes.*

As using LLMs in exploring qualitative data can introduce biases at various stages of the process (training, labeling design, or policy choices) [25], for this method, we built purposeful interaction between researchers and the output of any generated content from the LLM. For example, step 3 above was introduced as a means of sensemaking and checking the results. Additionally, because the LLM is a semantic tool, some summaries generated may have similar meanings although slightly different descriptions. For example, two generated summaries from the responses were: "The course taught them about the importance of ethics in technology" and "The course taught them a lot about ethics and technology." On exploring these codes in more detail, the differences in responses were negligible, so they were combined in the revisions of the codes.

F. Auto-NLP Analysis Method

The Auto-NLP method uses the LLM to generate the codebook in iterative phases. An initial set of child themes is generated, then summarized into a smaller set of parent themes, ultimately creating the final codebook. The process uses iterative rounds of text embedding in a high dimensional space using a pre-trained embedding model [20], clustering using agglomerative clustering, and summarizing by prompting a GAI model to create a code for the text in the cluster. The steps for the Auto-NLP process can be summarized as follows:

1) *Data Preprocessing: Anonymize and remove any identifiers from the participant dataset.*
2) *Data Preprocessing: Tokenize the corpus into sentences.*
3) *LLM Interaction: Using Python, request summarization, and generation of initial themes.*
4) *LLM Interaction: Using Python, request summarization, and generation of secondary themes from initial themes.*
5) *LLM Interaction: Using Python, request summarization, and generation of final themes.*

IV. Results and Discussion

In this section, we first discuss observations from implementing the Human-NLP and Auto-NLP methods. We then explore the themes in the context of the course. Tables I and II show the final themes in the qualitative codebooks for both the Human-NLP and Auto-NLP methods.

Regarding our first goal of exploring the data with automated assistance, both methods were based on the 788 sentences tokenized from the original dataset. The Human-NLP method yielded eight final themes, which were iteratively grouped into three overarching themes. These three themes represent the learning about ethics in the context of the course, the connection between ethics and technology, and the responsibility of organizations/developers of technology. These three themes were overall representative of the structure of the course, and the resources and assignments that the students conducted mirrored these themes. Similarly, the Auto-NLP method yielded 12 final themes, grouped into four overarching themes. The first three were overall very similar to the Human-NLP method. However, a final overarching theme across the work was students' attitudes toward the future of technology.

Overall, both methods provided a structured qualitative codebook that could be applied across the data. Additionally, comparing the final themes in each method, seven themes are mirrored across the two methods, although they are described in different vocabulary. Some examples of the alternate vocabulary include:

- "*Connecting experiences in the course to real life*" and "*The importance and application of ethics*."
- "*Using case studies and role-plays*" and "*Developing ethical understanding through exposure and application*."
- "*Consequences (or lack) of Actions*" mirrors "*Holistic ethical reasoning."*

All example quotes underlined in Tables 1 and 2 were shared across methods. The clustering algorithm was used multiple times in the Auto-NLP method. Still, the finding is striking as it highlights the potential for reaching similar themes between the assisted and fully automated methods.

However, some differences between how these codes were applied to the sentences may be due to context not present in the LLM analysis. We know that case studies and role-plays were used as foundational tools throughout the course to explore the topics in this course. In the Human-NLP analysis, this became the focus of a theme. However, in the LLM, the case study and role-play aspects were distributed into several other themes, predominantly "Developing an understanding of ethics through experience and reflection" and "Developing ethical understanding through exposure and application." The emotions and feelings around the case studies and role-plays were focused on more through the Auto-LLM method, which may result from its training data. This can also be seen in the frequency of some of the themes generated. The Auto-NLP method generated rather top-heavy themes, as 4 of the 12 themes have less than ten occurrences. However, this result may differ with modified parameters that account for cluster size in the narrow themes.

Regarding the second goal of exploring the generated themes, both methods highlight and demonstrate that students expressed an understanding of the relationship between technology, society, and the role of human values. Each learning outcome from the course can be tied back to the themes elicited through the analysis.

Across both methods, discussions about the role of ethics in technology, especially identifying concepts surrounding fairness, transparency, bias, and trust, were highlighted by students. They did so in an applied manner, consistent with how the course activities encourage students to engage in discussions. Looking through student responses associated with

the "Learning about Ethics, Reasoning, and Considerations" overall theme, many students mentioned not knowing or having a limited understanding of what ethical decision-making entails, other than ideology that is commonplace (i.e., do not harm others). Exploring the concepts in the course exposed them to some ethical theories but largely focused on helping them realize that ethics is not a single phenomenon that can only be engaged with a positivist viewpoint; rather, ethics is a constantly changing construct as societal norms change.

In talking about "Ethics Connection to Technology and Society," students highlighted the complexity of the systems at play, intention vs. reality, and weighed the pros and cons of building technology to serve vs. replace. Discussions on this theme were nuanced and provided students with a more holistic view of the forces they will likely face as they transition to the workforce.

Through both methods, students highlight the idea of responsibility and accountability. In most of the responses, they did so as the organizational responsibility – what should we as members of society expect of the organizations that are developing and implementing new technologies? This theme also highlighted the lack of regulation governing development and concerns about what this means going forward.

Overall, the themes addressed through the student's responses match well with the overall learning outcomes of the course, and students expressed satisfaction with using an interactive learning approach to deliver the course content.

V. STUDY CONSIDERATIONS AND LIMITATIONS

Overall, this methodology highlights an experimental use of NLP methods to assist with what can be a time-consuming process at the start of analyzing a larger dataset. However, some limitations should be taken into consideration with the use of these methods. First, it is important to note that any outputs produced as a result of the GAI were explored in more detail by the authors. This is because the GAI tools themselves are still experimental products, and the output from any such model should only be accepted by thoroughly reviewing them. This is an important step for future work on NLP techniques in this space. The purpose is to assist, not replace, the researcher.

A second limitation and point of consideration is the selection of specific algorithms and methods, which can contribute to the randomness of using NLP tools. The replicability of the analysis for the initial themes can be a challenge, especially if a deterministic algorithm is not used. If a non-deterministic clustering algorithm is used (e.g., K-means since the initial means are randomized), the output generated would be different, and whether these would be subtle differences or significant is difficult to tell beforehand. Likewise, the GAI text model could also be non-deterministic depending on the hyperparameters one uses. To address this issue, best practices suggest setting the temperature parameter for the model to zero. Introducing or maintaining the human in the loop through the Human-NLP method can help to verify the results by functioning as a sensemaking component. However, the researcher may influence the groupings and themes as a result of their own understanding of the concepts. Other biases will need to be explored in the way these methods evolve.

A third limitation is the speed at which the tools and models are being developed. At the time of preparing the work for publication, the model used has already received a significant update, and results may already be different.

ACKNOWLEDGMENTS

This work was partially supported by US National Science Foundation Awards #1954556, 1937950, 1939105; and USDA/NIFA Award #2021-67021-35329. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

REFERENCES

[1] P. M. Davies, R. J. Passonneau, S. Muresan, and Y. Gao, "Analytical Techniques for Developing Argumentative Writing in STEM: A Pilot Study," *IEEE Trans. Educ.*, vol. 65, no. 3, pp. 373–383, Aug. 2022, doi: 10.1109/TE.2021.3116202.

[2] C. G. P. Berdanier, E. Baker, W. Wang, and C. McComb, "Opportunities for Natural Language Processing in Qualitative Engineering Education Research: Two Examples," in *2018 IEEE Frontiers in Education Conference (FIE)*, Oct. 2018, pp. 1–6. doi: 10.1109/FIE.2018.8658747.

[3] A. Kamath and R. Das, "A Survey on Semantic Parsing." arXiv, May 29, 2019. Accessed: May 18, 2023. [Online]. Available: http://arxiv.org/abs/1812.00978

[4] A. Bhattacharjya, A. Punyamurthi, S. Atota Lakshmi, A. Iragala, M. C. Thallapureddy, and S. K. Kannaiah, "An Exploratory Study on Chatbots," in *2022 2nd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, Nov. 2022, pp. 340–344. doi: 10.1109/ICE3IS56585.2022.10010154.

[5] "Introducing ChatGPT." https://openai.com/blog/chatgpt (accessed May 10, 2023).

[6] "Bard." https://bard.google.com (accessed May 21, 2023).

[7] A. Johri, A. S. Katz, J. Qadir, and A. Hingle, "Generative artificial intelligence and engineering education," *Journal of Engineering Education*, vol. 112, no. 3, pp. 572–577, 2023, doi: 10.1002/jee.20537.

[8] "Code of Ethics | National Society of Professional Engineers." https://www.nspe.org/resources/ethics/code-ethics (accessed May 09, 2023).

[9] "ACM Code of Ethics and Professional Conduct." https://www.acm.org/code-of-ethics (accessed May 09, 2023).

[10] "Web Machine Learning Ethics." https://www.w3.org/TR/webmachinelearning-ethics/ (accessed May 09, 2023).

[11] "ChatGPT, artificial intelligence and higher education: What do higher education institutions need to know? – UNESCO-IESALC." https://www.iesalc.unesco.org/en/2023/04/14/chatgpt-and-artificial-intelligence-in-higher-education-quick-start-guide-and-interactive-seminar/ (accessed May 10, 2023).

[12] 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.

[13] A. Katz, C. Brozina, A. Johri, and L. Shuman, "Analysis of Advances in Engineering Education Publications (2007-2020) to Examine Impact and Coverage of Topics," in *2022 IEEE Frontiers in Education Conference (FIE)*, Oct. 2022, pp. 1–9. doi: 10.1109/FIE56618.2022.9962725.

[14] R. Alfayez, Y. Ding, R. Winn, and G. Alfayez, "What is Discussed About Software Engineering Ethics on Stack Exchange (Q&A) Websites? A Case Study," in *2022 IEEE/ACIS 20th International Conference on Software Engineering Research, Management and Applications (SERA)*, May 2022, pp. 106–111. doi: 10.1109/SERA54885.2022.9806760.

[15] R. T. Javed *et al.*, "Get out of the BAG! Silos in AI Ethics Education: Unsupervised Topic Modeling Analysis of Global AI Curricula," *The Journal of Artificial Intelligence Research*, vol. 73, pp. 933–965, 2022, doi: 10.1613/jair.1.13550.
[16] A. Belz, S. Agarwal, A. Shimorina, and E. Reiter, "A Systematic Review of Reproducibility Research in Natural Language Processing," *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pp. 381–393, 2021, doi: 10.18653/v1/2021.eacl-main.29.
[17] P. J. Tierney, "A Qualitative Analysis Framework Using Natural Language Processing and Graph Theory," *International review of research in open and distance learning*, vol. 13, no. 5, pp. 173–189, 2012, doi: 10.19173/irrodl.v13i5.1240.
[18] K. Crowston, E. E. Allen, and R. Heckman, "Using natural language processing technology for qualitative data analysis," *International Journal of Social Research Methodology*, vol. 15, no. 6, pp. 523–543, Nov. 2012, doi: 10.1080/13645579.2011.625764.
[19] C. G. P. Berdanier, C. M. McComb, and W. Zhu, "Natural Language Processing for Theoretical Framework Selection in Engineering Education Research," in *2020 IEEE Frontiers in Education Conference (FIE)*, Oct. 2020, pp. 1–7. doi: 10.1109/FIE44824.2020.9274115.
[20] A. Johri and A. Hingle, "Learning to Link Micro, Meso, and Macro Ethical Concerns Through Role-Play Discussions," in *2022 IEEE Frontiers in Education Conference (FIE)*, Uppsala, Sweden: IEEE, Oct. 2022, pp. 1–8. doi: 10.1109/FIE56618.2022.9962560.
[21] A. Hingle, H. Rangwala, A. Johri, and A. Monea, "Using Role-Plays to Improve Ethical Understanding of Algorithms Among Computing Students," in *2021 IEEE Frontiers in Education Conference (FIE)*, Oct. 2021. doi: 10.1109/FIE49875.2021.9637418.
[22] A. Hingle and A. Johri, "Assessing Engineering Student's Representation and Identification of Ethical Dilemmas through Concept Maps and Role-Plays," in *2022 ASEE Annual Conference & Exposition*, 2022.
[23] "Introducing Claude," *Anthropic*. https://www.anthropic.com/index/introducing-claude (accessed May 21, 2023).
[24] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks." arXiv, Aug. 27, 2019. doi: 10.48550/arXiv.1908.10084.
[25] E. Ferrara, "Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models." arXiv, Apr. 18, 2023. Accessed: May 15, 2023. [Online]. Available: http://arxiv.org/abs/2304.03738