Exploring Generative AI for Higher Education Accreditation: Opportunities and Challenges in ABET Reporting.

Caleb Levi Head, Purdue University – West Lafayette (College of Engineering)

Caleb Levi Head, MS, EIT, is a first-year Ph.D. student in the School of Engineering Education at Purdue University. He obtained his Master of Science in Mechanical Engineering at the University of Arkansas at Little Rock, where his research focused on enhancing technical writing assignments in engineering laboratories through industry collaboration for assignment development. As an instructor for second-year solid modeling (CAD) courses and fourth-year material properties laboratories, he gained valuable insights into the challenges students face in these areas, which have informed his educational approach and teaching philosophy. He also holds a Bachelor of Science in Mechanical Engineering from the University of Arkansas at Little Rock.

Caleb's research interests include industrial collaboration, global competencies in engineering education, and study-abroad experiences in engineering.

Aadithan Anbuvanan, Purdue University – West Lafayette (College of Engineering)

Aadithan Anbuvanan is a first-year Ph.D. student in the School of Engineering Education at Purdue University. He holds a bachelor's degree in Automobile Engineering from Anna University and a master's degree in Industrial Engineering from Clemson University. His research interests include AI-enhanced engineering education, curriculum development, design thinking, and assessment of pedagogical interventions.

Diane Patterson, Purdue University – West Lafayette (College of Engineering)
Dr. Justin L Hess, Purdue University – West Lafayette (College of Engineering)

Dr. Justin L Hess is an assistant professor in the School of Engineering Education at Purdue University. Dr. Hess's research interests include exploring empathy's functional role in engineering and design; advancing the state of the art of engineering.

Dr. Morgan M Hynes, Purdue University – West Lafayette (College of Engineering)

Dr. Morgan Hynes is an Assistant Professor in the School of Engineering Education at Purdue University and Director of the FACE Lab research group at Purdue. In his research, Hynes explores the use of engineering to integrate academic subjects in K-12 cla

Exploring Generative AI for Higher Education Accreditation: Opportunities and Challenges in ABET Reporting.

Abstract

Generative AI is a growing focus of research in higher education. However, there is limited academic literature describing the use of AI in higher education accreditation. Generative AI, with its benefits and challenges, has the potential to improve administrative workflows, thereby saving faculty time and energy that can better be spent on other administrative duties and non-administrative priorities such as teaching and research. However, Generative AI tools require a human-in-the-loop to ensure the validity of the synthesized data before submission. Hence, while Generative AI has the potential to save faculty time, we need a better understanding of its merits and weaknesses in certain areas before this potential can be realized. One area where we theorize that Generative AI can be particularly helpful is for preparing accreditation materials.

This paper focuses on mapping course objectives to ABET learning objectives. Pursuing ABET accreditation is a time-consuming process that requires the creation of a Self-Study report. This report requires student data from multiple courses, instructor feedback and insights, and analysis of program changes. We explore the efficacy of four Generative AI platforms (ChatGPT 5.0, Copilot, Gemini 2.5, Perplexity Pro), including their ability to recreate tables, provide commentary, and map course objectives for ABET continuous improvement documentation. To test each platform, we uploaded sample rubrics, assignment descriptions, syllabi, and other course information into these systems and then prompted each Generative AI to produce analysis tables. We provided each AI with blank tables, including prescribed row and column headers taken from a human-generated report, along with syllabi data and assessment data. We then compared the results generated by each AI platform with the human-generated report by applying a self-developed accuracy scale.

The authors conclude with a discussion of future work and limitations to using generative AI for ABET report writing, including file uploading, and the challenges encountered with prompting across various platforms. In our study, we found that most of the Generative AI models tested produced unreliable outputs, yet ChatGPT 5.0 was found to be the most capable at creating a table from source data. Although generative AI can help reduce the time it takes to create an ABET report by formatting a table and populating it, human review is still necessary since errors were common.

Introduction

ABET (formerly the Accreditation Board for Engineering and Technology) was founded in 1932 to establish guidelines and procedures for postsecondary programs in Applied and Natural Science, Computing, Engineering, and Engineering Technology [1], [2]. These post-secondary programs are evaluated at least every 6 years to assess their compliance with ABET's educational criteria. Many programs use mapping methods to demonstrate how each course in their curriculum falls under the various criteria and meets the educational objectives [1], [3]. Reviewing each course and mapping its outcomes, however, requires significant time and energy for faculty to assess courses and evaluate their effectiveness. Many programs have limited

personnel to handle these reviews, which can lead to a “large flurry of activity just prior to an accreditation visit” [4]. Various tools have developed substantially over the last 20 years to assist faculty in generating reports and mapping outcomes. Some authors utilize cloud computing and project management tools such as Microsoft Access to store files and compile documents for the self-report [4].

Achieving and maintaining ABET accreditation is no simple task, as recurring evaluations prompt each engineering department to proactively gather its materials for review. Departments use the materials to highlight the seven ABET criteria and how they are achieving them through their plans of study. When compiled, self-study reports/binders can be quite lengthy, as shown in **Figure 1** (note, this visual shows the binder of materials that our undergraduate program has provided to ABET reviewers). Creating the ABET report is a time-intensive and multi-year process, as each program must diligently compile and prepare a self-report that accurately depicts the program's current state and its continuous improvement plans.

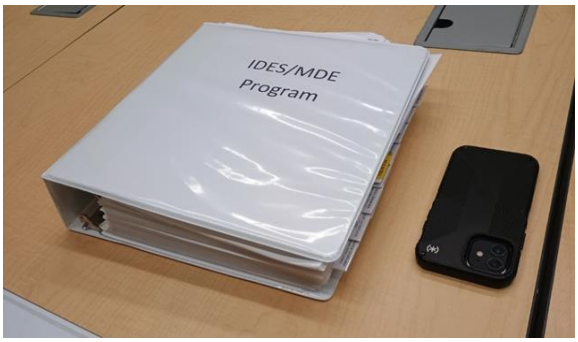

Figure 1: Binder containing a single department’s ABET report and data (iPhone 12 for scale)

Background & Motivation

Many engineering degree programs are accredited through ABET. Programs pursuing ABET accreditation undergo comprehensive reviews at least every six years, where programs and institutions create a Self-Study Report based on their compilation of course syllabi, assignments, sample assignment submissions, design projects, design outcomes, evaluations of student work, among other materials. ABET reviewers meet with faculty, students, and administrators, tour classroom and laboratory facilities, and examine the self-study report that each institution creates. This self-study report outlines how the program meets the ABET criteria, effectively assesses program outcomes, and oversees continuous improvement of the program and its curricula. This self-study report should demonstrate whether the program meets the ABET outcomes required for accreditation. It typically requires many hours of report writing, data collection, organization, and table creation by faculty members and administrative assistants.

Because creating a report requires collecting, summarizing, evaluating, and presenting data, it can be labor-intensive and time-consuming. Thus, researchers have investigated how Generative AI can speed up processes, especially administrative processes. For example, a study found that AI was able to create consistent business reports at a fraction of the time it took human

counterparts to perform [5]. In the field of Engineering Education, one area where we see potential for similar use is ABET accreditation.

Artificial intelligence (AI) has become a new, hot technological trend. More specifically, Generative AI has become popular, especially with the advent of ChatGPT's free model. Generative AI is defined as "computational techniques that are capable of generating seemingly new, meaningful content such as text, images, or audio from training data" [6]. The quality of this generated content has advanced to the extent that it has become difficult to distinguish Generative AI-produced work from genuine human work. Because Generative AI responses have improved, multiple fields have begun investigating how this new technology can be applied [7]. One common target for AI applications is simple, repetitive, or time-consuming tasks, like data synthesis or report writing.

From the perspective of sociotechnical systems, the application of generative AI serves as an augmentative technology that operates within existing institutional processes, rather than making independent decisions. Sociotechnical systems embody the concept that technology alters the way job practices are performed while relying on human control [8]. Within the application of ABET reporting, such systems could decrease administrative workload while maintaining faculty responsibility.

In a study performed by Sun and Yao [9], two authors who have significant experience reviewing ABET program materials, they evaluated how AI can be applied to all steps of the ABET certification timeline, including the Self-Study report, 7-day Response report, and 30-day Response report. For simple, 1-prompt questions, they found that ChatGPT provided enough guidance and executable details for the majority of categories they studied [9]. However, their study was limited to written advice and analysis and does not capture the practical problem of generating other aspects of the report, such as tables figures. In this study, we build from this prior work, exploring the effectiveness of Generative AI for producing such materials.

Theoretical Framework

This study applies the Qualifying Qualitative Research Quality (Q3) framework [10] in a new context: accreditation. Models of research rigor and overall research quality, as they pertain to interpretive educational research, are relevant to consider as a lens through which to view the application of Generative AI for creating accreditation materials. There are aspects of structuring results and interpretations, professional exercises, as well as open documentation practices when preparing accreditation materials that the Q3 framework can support. The Q3 model operationalizes quality as an informant of intentional and transparent processes, such as the processes informing the development and interpretation of data, rather than just the outcomes of research [11]. In the case of program accreditation, the focus on process remains applicable, since there is a need for assessment data to be collected and interpreted in high-quality ways to assess desired student outcomes and to improve programmatic outcomes and programmatic review processes.

In engineering education literature, the Q3 framework has been studied in relation to large language models (LLMs), contending that the impact of LLMs is limited to the processing and

representation of data, rather than the construction of meaning [10]. Such nuances take on particular importance in the ABET accreditation process, where humans must be able discern the trustworthiness of Generative AI outputs and, in many cases, explain the process for generating such outputs. Specifically, the ABET process involves collecting, summarizing, sorting, organizing, and interpretating data. Generative AI may be an effective tool for supporting data collection and interpretation, but users need to be able to discern how these systems work and the extent to which outputs of Generative AI are trustworthy.

Several dimensions of quality in Q3 are particularly relevant to discussions of AI-augmented accreditation procedures and give rise to potential concerns when using Generative AI in accreditation. Procedural validation emphasizes the importance of transparent and rule-governed procedures which can be compromised by the opaque and probabilistic nature of Generative AI outcomes [12]. Process reliability highlights issues of stability and consistency, which have been consistently noted in generative AI systems [10]. Theoretical validation once again highlights the alignment between generated representations and educational reality, underscoring the importance of disciplinary knowledge in assessing AI-generated outputs. Lastly, ethical validation centers on issues of responsibility and accountability, particularly in high-stakes formal applications such as accreditation [11].

We surmise that Generative AI can be a valuable tool for pursuing program accreditation. As these examples show, applying the validation aspects from Q3 framework can inform new ways of using Generative AI in accreditation processes while surfacing and providing potential responses (e.g., prompts, human responsive) to validation concerns.

Methods

One way to use AI in ABET report creation involves creating tables of the data. This task is theoretically suitable for Generative AI as it is a task that is straightforward and easily verified, and other studies have found that generative AI is capable of creating trustworthy tables and written analysis [5], [9]. We began this study by testing Generative AI's written analysis, but we removed this focus from our analysis in this project to limit the scope of this project to table development alone.

We tested four Generative AI: ChatGPT 5.0, Copilot, Gemini 2.5, and Perplexity Pro. We chose these programs because they had generative capabilities and were easily accessible to university faculty and students. The generative component was the most important since we are trying to generate tables and written analysis to use toward developing an ABET self-study report from source data and materials.

Our source material that we tested was deidentified graded rubrics for a final project from two courses. We prompted each Generative AI to use the source material to create a table and/or write analysis on the table in the style of an ABET report. To verify the Generative AI outputs, we compared the AI response to an existing ABET report completed in 2025. **Figure 2** depicts our process, including examples of the tables we shared with each Generative AI and an example output provided by a Generative AI system.

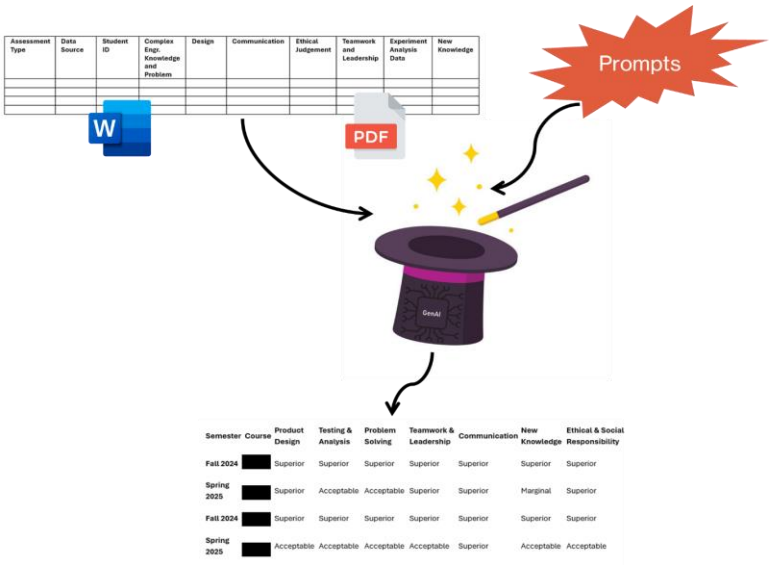

Figure 2: Flowchart of methods used in study.

We used rubrics for two projects from different courses as source data for this investigation. Two researchers independently prompted the AI programs using two distinct prompting styles. The first prompt sequence was verbose, wherein the prompter provided detailed instructions and written explanations of tables and analysis. The first prompt sequence proceeded as follows:

1. Can you extract the scores for each outcome in this document if E=4, G=3, A=2, and U=1? (*The prompter attaches a source document*)
2. Similarly, can you extract the scores for each outcome in this document? (*The prompter attaches the other source document*)
3. Can you combine the data into a single table where row 1 is the data from the second document and the data in row 2 is the data from the first document? Each column in the new table is a learning outcome from this list: Complex Engineering Knowledge/Problems, Design, Communication, Ethics/Global, Teamwork/Leadership/Project Management, Experiments/Analyze Data, and Acquiring New Learning. Match the outcomes from the documents to the learning outcomes in the list as best as possible with no double counting.
4. Can you average the score in each entry of the table so that there is only 1 score per each cell?
5. Can you replace the scores such that S=4, S-=3.7, Ac+=3.3, Ac=3, Ac-=2.7, M+=2.3, M=2, M-=1.7, U+=1.3, and U=1? Pick the closest letter criteria.
6. Can you analyze the table in a similar style to the following paragraph? Assume S = superior, Ac = acceptable, M = marginal, and U = unsatisfactory.

The second prompt sequence was more concise and used images or screenshots to describe instructions. Examples for this second prompt sequence included the following:

1. How can I draft this table with the assessment files I have uploaded? *(The prompter attaches an image of the table template and the source documents.)*

2. For MDE use Excellent=4, Good=3, Acceptable=2, and Unacceptable=1 for table creation
3. Summarize the box in the table to a singular value like given in the table
4. Use the same metrics from the template table
5. Relate 4 - Superior, 3 – Acceptable likewise

Once a prompt sequence achieved a desired table in one of the four AI programs, we repeated the exact same prompts for the other AI programs. After reviewing the prompt sequences and results, we selected prompt sequence 2 since its style was more efficient, requiring less prompts and less source documents.

We then graded the outputs by using a rubric on a scale from 1-3 (which correspond to poor, acceptable, and good respectively). The rubric was proposed by one researcher, and the rubric and criteria were discussed and approved by the other team members. As a result, the final rubric contains key metrics determined through consensus. We scored all Generative AI based on six criteria, which we defined as follows:

1. Correct Values – Does the final table calculate outcomes correctly?
2. Readability – Is the table legible and interpretable?
3. Mapping Outcomes – Are the course outcomes in the source data mapped to the ABET outcomes?
4. Consistency – Do repeated prompts result in the same results?
5. Conciseness – Can a program create a table with minimal number of prompts and require no clarifying prompts?
6. Explanations – Does it provide explanations on why it mapped certain outcomes?

We provide the rubric items associated with each criterion in Appendix A. From the scores achieved, we observed outputs, variants in scores per rubric item, and then discerned the Generative AI that was the most useful for each criterion.

Results

All Generative AI programs we tested could generate a table from prompts and source data we provided. However, some Generative AI were better than others at understanding prompts and creating an effective table. As Figure 3 shows, each coder provided each Generative AI output with a score where a score of 3 is good, 2 is acceptable, and 1 is poor. Overall, we found that ChatGPT outperformed Perplexity Pro, Gemini, and Copilot on nearly all rubric criteria. In some cases, we chose not to score criteria 4 to prevent longitudinal factors affecting our results. This criterion could not be sufficiently captured in this study since we had a single prompting period. We expand on the results for each Generative AI in the following sections.

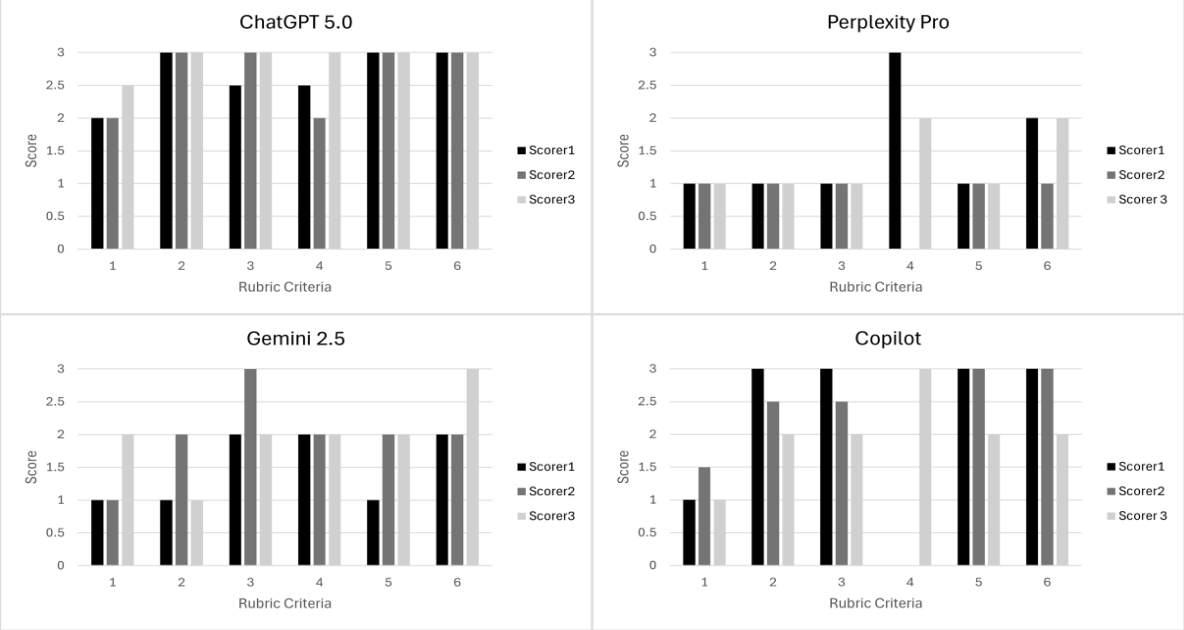

Figure 3: Inter-rater Scoring of Six Rubric Criteria: (1) correct values, (2) readability, (3) mapping outcomes, (4) consistency, (5) conciseness, and (6) explanations.
Note: Each criterion was scored on a scale of 1-3, scoring from poor (1) to good (3).

ChatGPT 5.0

ChatGPT 5.0 scored the best overall at creating a table from the source data. For this style of prompting, ChatGPT was able to generate a legible table populated with correctly calculated values. Although ChatGPT needed clarification prompts for the generated table to match the baseline table, the prompts did not have to be reworded or expounded in order for ChatGPT to understand and make corrections. The explanations ChatGPT generated described which course outcomes matched the ABET outcomes, and these descriptions were reasonable and logical. As a result, verification was much easier to perform.

One downside of ChatGPT is it has a limit to the number of documents that can be uploaded with one prompt. For the free version of ChatGPT, only five documents can be uploaded and there is a file size limit as well. Thus, ChatGPT would be a poor choice for future applications with large datasets.

Perplexity Pro

We found that Perplexity Pro was not effective at creating a table from source data. For prompt sequence 1, Perplexity did not create a table but rather described how someone would create a table with the given source data. As a result, it scored mainly 1's across all criteria except Explanations criteria. Perplexity is capable of creating a table, but with the simple instructions of prompt sequence 1, it did not understand our goal.

It is important to note that the results shown in Figure 3 are for the Pro version of Perplexity which is unlocked using an academically affiliated email address. Because Perplexity did poorly even with an upgraded version, we do not currently advise using this program for table generation.

Gemini 2.5

Gemini 2.5 was found to be acceptable at creating a summary table from source data. Because of an error in outcome mapping, there was an error in the final table values, so this dropped the score of criteria 1 and 3. Despite generating two separate tables for each set of source data, the tables were legible. In addition, Gemini did provide some explanations to the table generation process, but these explanations did not delve into how it mapped outcomes or averaged the source data scores. Gemini scored a 2 for conciseness because it required an additional prompt where the user had to select the ideal option between two generated results.

Copilot

We found that Copilot generates an acceptable summary table from source data. The table Copilot generated was legible and similar format to the baseline table, and it was populated by a single value. However, it ignored one of the source datasets, so it scored low on the correct value criterion. Copilot did provide some explanation to how it mapped course outcomes to ABET outcomes, so it achieved good to acceptable rankings for criteria 3 and 6.

Although Copilot scored well, we do not recommend using this AI for table generation due to file limitations and formatting problems. Since Copilot has a limit of 3 attachments, it is not ideal datasets with multiple source documents. In addition, the generated table is poorly formatted since it cannot be copied directly from Copilot's interface. The generated table lacks cells, so the user must instead create the table cells in an external program before copying the values from Copilot. Because of this tedious formatting, we will not be using Copilot for table generation in the future.

Discussion

This paper provides a use case for applying the Q3 framework to explore the use of Generative AI in program accreditation. As we noted in the theoretical framework section, the Q3 framework can guide faculty to account for students' and other stakeholders' social realities in the "making" and "handling" of data for accreditation purposes. While our study focused on the reliability of producing tables for a self-study report, we surmise that Generative AI can also be used to support the reporting of assessments on student outcomes. Applying the Q3 framework to other aspects of the accreditation process can help programs think more creatively about their continuous improvement processes.

Another takeaway from our study is that Generative AI use goes beyond accreditation. Humans should still be guiding, reviewing, and sometimes rejecting the generated work, as there are limitations with utilizing Generative AIs [10], [13], [14]. In our study, most of the Generative AI produced unreliable outputs. Among the readily accessible generative AI platforms, ChatGPT 5.0

was found to be the most capable of creating a table from source data. In combination with human review, we believe that generative AI can help reduce the time it takes to create an ABET report by formatting a table and populating it with values from source data. However, it is still imperative that the results be human reviewed, as errors were seemingly commonplace. To better optimize table generation for accreditation purposes, specifically, future research should focus on how prompt wording and sequencing affects Generative AI outputs.

This work has several limitations. First, Generative AI systems have advanced significantly in skill and will continue to improve rapidly. Thus, while this work serves as a reference point in time, research on Generative AI use can become obsolete as development and enhancements of Generative AI accelerate. Second, we focused on open-access prompt-and-generation models, so we were not using the most advanced Generative AI systems. We chose this method to explore how reliable the most accessible models were for the public. Third, we recognize that each Generative AI is different in purpose and specialty. Some Generative AIs may work better at synthesizing information, others at generating files, and others at analysis. Finally, while we offer lessons from prompt sequencing based on our iterations, this study does not apply principles of prompt engineering. Because of this, our prompts cannot be seen as a one size fits all solution. Rather, we hope these lessons can inform future efforts at prompt engineering, specifically as it pertains to the development of accreditation materials.

References

[1] K. Shryock and H. Reed, "ABET accreditation: Best practices for assessment," presented at the 2009 Annual Conference & Exposition, 2009, pp. 14–148.

[2] ABET Engineering Accreditation Commission, "ABET criteria for accrediting engineering programs," Baltimore, MD, 2025.

[3] C. L. McCullough, "Can ABET assessment really be this simple?," presented at the 2020 ASEE Virtual Annual Conference Content Access, 2020.

[4] R. Cliver, W. M. Leonard, E. Dell, and R. A. Merrill, "ABET Report Generation," presented at the 2011 ASEE Annual Conference & Exposition, 2011, pp. 22–129.

[5] Rahul Modak, "Generative AI for automated business report generation and analysis," *World J. Adv. Eng. Technol. Sci.*, vol. 15, no. 2, pp. 894–901, May 2025, doi: 10.30574/wjaets.2025.15.2.0610.

[6] S. Feuerriegel, J. Hartmann, C. Janiesch, and P. Zschech, "Generative AI," *Bus. Inf. Syst. Eng.*, vol. 66, no. 1, pp. 111–126, Feb. 2024, doi: 10.1007/s12599-023-00834-7.

[7] K. Yelamarthi, R. Dandu, M. Rao, V. P. Yanambaka, and S. Mahajan, "Exploring the potential of generative AI in shaping engineering education: Opportunities and challenges," *J. Eng. Educ. Transform.*, pp. 439–445, 2024.

[8] G. Baxter and I. Sommerville, "Socio-technical systems: From design methods to systems engineering," *Interact. Comput.*, vol. 23, no. 1, pp. 4–17, 2011.

[9] W. Sun and J. Yao, "Exploring the Potential Application of ChatGPT in Preparing for ABET Accreditation," Jun. 09, 2023. doi: 10.36227/techrxiv.23392412.v1.

[10] D. Reeping, C. Hampton, and D. Özkan, "Interrogating the Use of Large Language Models in Qualitative Research Using the Qualifying Qualitative Research Quality Framework," *Stud. Eng. Educ.*, vol. 6, no. 2, 2025.

[11] J. Walther and N. Sochacka, “Qualifying qualitative research quality (The Q 3 project): An interactive discourse around research quality in interpretive approaches to engineering education research,” presented at the 2014 IEEE Frontiers in Education Conference (FIE) Proceedings, IEEE, 2014, pp. 1–4.
[12] J. Walther, N. W. Sochacka, and N. N. Kellam, “Quality in interpretive engineering education research: Reflections on an example study,” *J. Eng. Educ.*, vol. 102, no. 4, pp. 626–659, 2013.
[13] P. A. Christou, “Thematic analysis through artificial intelligence (AI),” *Qual. Rep.*, vol. 29, no. 2, 2024.
[14] T. Jowsey, V. Braun, V. Clarke, D. Lupton, and M. Fine, “We reject the use of generative artificial intelligence for reflexive qualitative research,” *Qual. Inq.*, p. 10778004251401851, 2025.

Appendix A: Rubric for AI Table Generation

Criterion	Poor	Acceptable	Good
Correct values	• Calculations yield incorrect or no values • Fails to answer prompt • Table is not populated • AI generated data not in sample set	• Calculations yield values close to expected • Answers prompt such that the main idea is represented • Table is populated	• Calculations yield expected values • Answers prompt exactly • Table is populated
Readability	• Table is not legible/ interpretable by the reader • Data is not summarized to a single value • Table format makes no sense	• Majority of table entries are legible/ interpretable • Data is summarized • Table format makes some sense	• All table entries are legible/ interpretable • Data is summarized • Table format is perfectly logical
Mapping Outcomes	• No mapping of multiple course outcomes to a single ABET outcome • There is no logic behind the mapping of outcomes (the topics are too different) • AI generated outcomes not in sample set	• Some course outcomes are mapped to ABET outcomes • It could be argued that there is similarity between mapped outcomes	• All course outcomes are mapped to ABET outcomes • The mapped outcomes makes sense