



## AI Feedback for LEED Narratives: Exploring LLM Potential

Zhenlin Yang<sup>1</sup>, Laura Cruz-Castro<sup>1</sup>, and Gabriel Castelblanco<sup>1</sup>,  
<sup>1</sup>University of Florida

Effective feedback is crucial in sustainability design education, where students must meet complex criteria such as Leadership in Energy and Environmental Design (LEED) standards. This study explores the potential for an automated feedback system to be powered by Large Language Models (LLMs), designed to deliver real-time, detailed, and rubric-aligned feedback for sustainability-focused projects. Through quantitative and qualitative analysis, the tool's performance was compared to teaching assistant (TA) feedback, focusing on feedback length, alignment with LEED rubrics, clarity, and timeliness. The results show that Llama 8B 3.1 has significantly longer and more structured feedback, offering immediate and iterative responses that facilitated systematic refinement of student work. In contrast, TA feedback, while shorter and often delayed, provided nuanced, context-sensitive insights tailored to individual submissions. The findings demonstrate the effectiveness in delivering standardized and scalable feedback by Llama 8B 3.1, while also highlighting the value of TAs' personalized guidance. This study advances the field of automated feedback systems and underscores the potential of integrating AI-powered tools with human expertise to enhance sustainability design education.

### Introduction

In engineering design education, real-time feedback is instrumental in enhancing student learning and fostering the development of high-quality design solutions (Castelblanco, 2024). Immediate and constructive feedback enables students to iteratively refine their designs, promoting deeper understanding and mastery of design principles (Boud & Molloy, 2013). The iterative nature of design processes benefits significantly from timely insights, as they allow continuous improvement of ideas and approaches (Kolko, 2010).

However, providing individualized real-time feedback presents considerable challenges due to human resource limitations. Instructors often manage large classes and complex curricula, which restrict their ability to offer immediate, detailed feedback to each student (Carless, 2006). Delays in feedback can hinder the learning process, as students may proceed with misconceptions or errors that could have been corrected promptly with timely input.

Advancements in artificial intelligence, particularly in the development of Large Language Models (LLMs), offer promising solutions to these challenges. LLMs, such as Llama 8B 3.1, OpenAI's GPT-3 and GPT-4, have demonstrated remarkable capabilities in natural language understanding and generation, enabling them to provide automated feedback that closely mimics human responses (Floridi & Chiriatti, 2020). These models can process extensive information and generate contextually relevant feedback, making them suitable tools for enhancing educational experiences.

Prior to the emergence of LLMs, automatic feedback systems relied on predefined algorithms and rule-based responses, which often lacked the flexibility to handle the complexity and nuance of open-ended design tasks (Shute, 2008). While useful in certain domains, these systems were limited in their ability to provide personalized and context-specific feedback necessary for supporting creative design processes.

The integration of LLMs into educational feedback systems has expanded the potential for providing dynamic, individualized feedback. Studies have shown that AI-powered feedback can support learning by offering immediate, detailed responses that help students understand and correct their mistakes. For example, Zawacki-Richter et al. (2019) highlighted the ability of AI applications to enhance personalized learning experiences in higher education by adapting to individual needs and providing tailored feedback. In language learning, AI tutors have been effective in improving student outcomes by providing customized feedback on language use and comprehension (Xie et al., 2019).

Despite these advancements, the application of LLMs for feedback in engineering design education remains underexplored. Engineering design projects, especially those focused on sustainability, involve complex decision-making processes that require consideration of environmental, social, and economic factors. Students must integrate standards such as the Leadership in Energy and Environmental Design (LEED) into their projects, which adds layers of complexity to the design process (Kibert, 2016).

Sustainability-focused design education necessitates feedback systems that can comprehend and evaluate multifaceted criteria. Students often lack sufficient guidance in applying sustainability principles to their designs, highlighting the need for tools that can provide specialized feedback in this area (Fenner et al., 2005). The complexity of sustainability decisions requires feedback that is both technically accurate and contextually relevant.

To address these challenges, we explore the potential to create an innovative system designed to provide real-time, automated feedback for sustainability design projects. Leveraging the capabilities of LLMs, and integrating detailed sustainability criteria, focusing on LEED certification standards, to deliver personalized and constructive feedback to students. By emulating the evaluative process of human experts, to enhance the learning experience and support students in developing high-quality, sustainable designs.

This study investigates the effectiveness of GPT4 in providing feedback in a LEED narrative assignment, focusing on two primary research questions:

- RQ: Can an LLM model provide specific and actionable feedback in a LEED narrative assignment to students?

The following sections outline the methodology employed to assess the tool's performance, present the results of our analysis, and discuss the implications for educational practices in sustainability design. By evaluating the ability to replicate human feedback this research contributes to the advancement of automated feedback systems in engineering education.

## Literature Review

### *Theory on Feedback for Design*

Feedback is a critical component in the design process, essential for improving design quality and enhancing learning outcomes. Schön (1983) introduced the concept of the "reflective practitioner," emphasizing the importance of reflection-in-action, where designers continually evaluate and adapt

their work based on feedback. This iterative process allows designers to develop solutions that are more responsive to user needs and contextual factors.

In educational settings, effective feedback fosters self-regulated learning and helps students bridge the gap between current performance and desired outcomes (Nicol & Macfarlane-Dick, 2006). Feedback that is timely, specific, and actionable enables learners to understand their mistakes and make necessary adjustments. Hattie and Timperley (2007) further highlight that feedback is most effective when it addresses three key questions: "Where am I going?", "How am I going?", and "Where to next?" This framework guides learners in setting goals, assessing progress, and planning future actions.

In the context of design education, feedback supports the development of critical thinking and creativity. Adams et al. (2016) discuss how feedback in design fosters an environment where students can explore ideas, receive input, and iterate on their designs, leading to deeper learning and innovation.

### *2.2 Automatic Feedback Before LLMs*

Prior to the advent of LLMs, automatic feedback systems primarily relied on rule-based algorithms and predefined responses. Intelligent Tutoring Systems (ITS) were developed to provide personalized instruction by modeling student knowledge and adapting to their needs (VanLehn, 2006). These systems used decision trees and expert systems to deliver feedback, but they were limited in handling complex, open-ended tasks typical in design.

In writing and language education, Automated Writing Evaluation (AWE) tools like the Criterion system by Educational Testing Service provided feedback on grammar, syntax, and basic composition (Attali & Burstein, 2006). While useful for language mechanics, these tools lacked the ability to provide substantive feedback on content quality or creative aspects of design work.

Similarly, in programming education, automated graders could assess code correctness and style based on test cases and linting tools but struggled with evaluating design patterns or algorithmic efficiency (Ihantola et al., 2010). These limitations highlighted the need for more sophisticated systems capable of understanding context and providing nuanced feedback.

### *LLMs and Feedback*

The emergence of LLMs, such as OpenAI's GPT-3 and GPT-4, has revolutionized the potential for automatic feedback by enabling machines to understand and generate human-like language based on deep learning models trained on vast datasets (Brown et al., 2020). GPT-4 is widely recognized as one of the most prominent LLMs in this field, offering advanced capabilities for language understanding and generation. At the same time, Meta's Llama is another powerful model that demonstrates similar promise for educational applications.

LLMs have been applied in educational contexts to provide dynamic and context-aware feedback. For example, Susnjak (2022) and Cruz Castro (2024) discussed the potential of AI-powered language models like GPT in education, highlighting their capacity to generate meaningful feedback, promote active learning, and adapt to individual student needs, making them effective tools for enhancing educational outcomes.

In the field of design and engineering education, LLMs offer the potential to analyze complex student submissions, understand design rationale, and provide feedback that addresses both technical accuracy and creative aspects. For example, Agrawal et al. (2021) demonstrated how AI could assist in

architectural design education by generating design alternatives and critiquing student work based on specified criteria.

In addition, LLMs' ability to manage temporal dependencies (the need to reference information from previous steps when generating later outputs) has been significantly enhanced by modern architectures and prompting techniques. Early neural models, such as Long Short-Term Memory (LSTM) networks (Hochreiter & Schmidhuber, 1997), addressed vanishing gradients by selectively retaining relevant information over extended sequences. More recently, transformer-based architectures (Vaswani et al., 2017) with self-attention mechanisms have enabled even more powerful context handling across long text passages. Building on these advances, CoT prompting (Wei et al., 2022) guides the LLM to explicitly articulate its intermediate reasoning steps before arriving at a final conclusion. For LEED narrative evaluation, where earlier design decisions must remain consistent with later sustainability criteria, CoT provides a structured, stepwise method of analyzing and explaining the model's reasoning. By doing so, the approach can produce more transparent, context-sensitive, and accurate feedback—especially critical in specialized domains requiring rigorous multi-step logic.

Despite these advancements, challenges remain in applying LLMs for feedback in engineering design. Concerns include ensuring the accuracy and reliability of the feedback, avoiding the propagation of errors or biases present in training data, and integrating AI feedback effectively with human instruction (Holmes et al., 2019).

Furthermore, the application of LLMs in sustainability-focused design education is still emerging. Sustainability design requires not only technical knowledge but also an understanding of environmental impact, social implications, and ethical considerations (Shin et al., 2024). Models like Llama may provide specialized feedback that aligns with sustainability principles and industry standards such as LEED certification, although the field is still growing and demands further research.

## **Methodology**

### *Context and Participants*

This study was conducted during the Spring semester of 2024 within the Construction Capstone Project course at a southeastern university. The course enrolled 43 students and was facilitated by a team comprising two instructors and two graduate TAs. The primary learning objectives were to enable students to design, develop, estimate, schedule, contract, and manage work related to small commercial, residential, civic, or light industrial projects.

We will focus on the LEED narrative component of the capstone project. Obtaining LEED certification requires special considerations during the design process to meet specific requirements related to lighting, energy efficiency, water conservation, and indoor air quality. Students will be tasked with recommending strategies to achieve this goal - specifically, selecting appropriate LEED points - and demonstrating a comprehensive understanding of the LEED certification process.

Upon receiving user-submitted LEED points or uploaded project documents, the backend processes the data to generate customized feedback. It matches user input with the corresponding LEED scoring criteria, calculates total and individual scores, and constructs a detailed response highlighting the compliance level and areas for improvement. The system communicates with an external Llama 8B 3.1 API to obtain feedback, requiring careful design of prompts to ensure accurate and relevant responses. This architecture ensures real-time interaction between users and the system, maintains data integrity, and supports scalability.

### *Prompt Design*

The effectiveness of the feedback depends heavily on how the prompts are designed to elicit accurate and relevant feedback from the AI language model. The prompts are carefully crafted to provide clear instructions and sufficient context, guiding the AI to generate responses that comply with LEED standards and offer actionable insights to the user.

For LEED narratives, the prompt begins by confirming the content type and assessing whether the user's total LEED score meets the minimum requirements for certification. This initial context helps the AI understand the user's current standing regarding LEED certification thresholds. The prompt then includes the LEED narrative assignment rubrics, detailing categories such as "LEED Certification Achievement" and "Reflection of Credit Requirements." Each rubric provides a title, description, total points available, and specific scoring criteria, which guide the AI's evaluation of the user's submission.

To facilitate subsequent extraction of scores and ensure consistency in feedback, the prompt includes clear instructions on how the AI should format its response. The AI is instructed to present the rubric title, assign a score in a specified format (e.g., "Score: X/Y"), and provide concise feedback that explains the reasoning behind the score and suggests improvements. An example is included within the prompt to illustrate the expected format, aiding the AI in understanding how to structure its response effectively. The user's text is then appended to the prompt for the AI to evaluate based on the provided rubrics and instructions.

Moreover, the prompt design incorporates the use of CoT prompting techniques (Wei et al., 2022). By encouraging the AI to reason through the evaluation process step by step, CoT prompting helps the AI produce more coherent and comprehensive feedback. This approach enhances the AI's ability to analyze complex aspects of the user's submission, such as alignment with LEED criteria and the quality of proposed sustainability strategies.

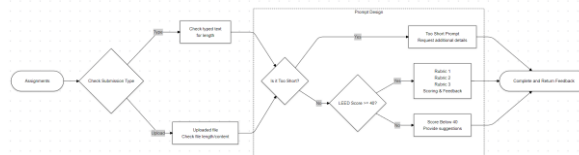
### *Data and Data Analysis*

The data used in this study were derived from two primary sources. The first dataset consisted of feedback provided by TAs on student assignments, collected from the Canvas learning management system. To ensure fairness, any personal or identifying information in the assignments was removed before TAs provided their feedback. The TAs, who were experienced and had been grading across multiple semesters, were instructed to give feedback in the same way they would in a regular class setting. This feedback reflected human insights and was often tailored to address specific aspects of student work. The second dataset was generated by the by prompting the Llama 8B 3.1 model, an automated feedback system designed to evaluate submissions based on predefined LEED rubrics. Unlike traditional student submissions, the data analyzed in this study were generated through experimenters submitting assignments to the tool multiple times to capture iterative feedback. To illustrate the entire process—from receiving an assignment to producing feedback aligned with LEED scoring criteria—Figure 1 presents the system's step-by-step workflow for each submission. This figure shows how the submitted content (typed or uploaded) is evaluated for completeness, scored according to LEED standards, and ultimately returned with tailored feedback. By visualizing these steps, we provide a clear overview of how data were collected, processed, and assessed for this study. The LEED rubrics used by the tool included detailed criteria for evaluating submissions, covering aspects such as certification achievement, reflection on credit requirements, formatting, realistic implementation strategies, and writing clarity.

The theoretical framework guiding this analysis draws on established models of effective feedback. Nicol and Macfarlane-Dick (2006) highlight that feedback should be timely, specific, and actionable

to bridge the gap between current and desired performance. Similarly, Schön's (1983) concept of the "reflective practitioner" emphasizes the role of feedback in fostering reflection-in-action, enabling iterative refinement and improvement. These frameworks provide a basis for evaluating the feedback's quality and its potential to support learning and development.

The analysis methods focused on both quantitative and qualitative aspects of the feedback. Feedback length was analyzed quantitatively, measured in words, to assess the level of detail provided by TAs and the tool. Content analysis was conducted qualitatively to evaluate feedback clarity, specificity, and alignment with learning objectives and LEED criteria. This combined approach allowed for a comprehensive assessment of the feedback's effectiveness.



**Figure 1:** Flowchart for each submission

## Results

### *Prompt*

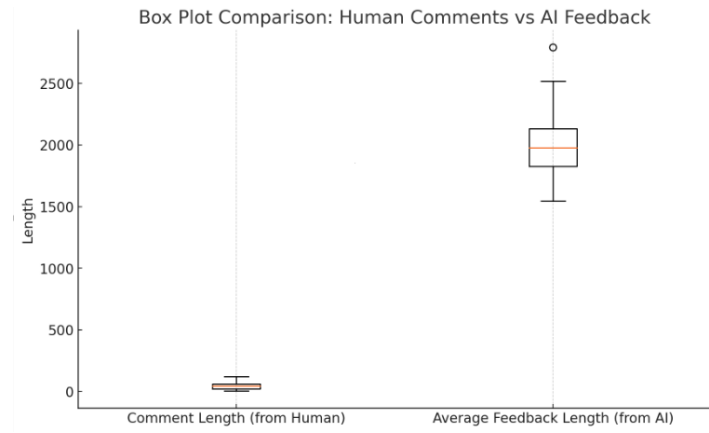
There were 4 iterations to get to the final prompt. One of the principal challenges was balancing the need to guide students clearly while covering all aspects of the LEED certification requirements. In the first iteration, our focus was on removing all student-identifying information to protect privacy; however, after doing so, the model's feedback remained too generic and didn't effectively align with the rubric. In the second iteration, we refined the prompt to explicitly reference each LEED rubric category—such as prerequisite fulfillment, credit point allocation, and realistic implementation strategies. This change helped the model generate more relevant feedback but still lacked detail in terms of actionable suggestions. In the third iteration, we incorporated reflective elements and asked the model to point out possible misconceptions in the submissions while offering step-by-step guidance for improvement. Although this produced more nuanced feedback, users reported that certain aspects needed clearer structure and direct links to rubric criteria. Finally, in the fourth iteration, we focused on clarifying the rubric references further and prompting the model to use consistent formatting and plain language. By showing each phase in our system's flow chart (Figure 1), we illustrate how these iterative refinements enhanced both the thoroughness and usability of the tool's feedback. After all these considerations, the final prompt was:

Demonstrate how the project meets LEED certification by focusing on the selected credits, prerequisites, and related performance metrics. Refer to the rubric's scoring criteria to assess compliance and provide detailed feedback on each element, identifying improvements needed to achieve LEED Certified status.

### *Length Analysis*

The comparison of feedback length between TAs and the tool reveals a substantial difference, as illustrated in Figure 2. The box plot demonstrates that tool-generated feedback is consistently longer and more detailed than TA feedback. The median length of tool feedback significantly exceeds that of TA feedback, with the interquartile range (IQR) for tool feedback spanning approximately 1,800 to 2,200 words, compared to an IQR of 20 to 60 words for TA feedback. Outliers in the TA feedback

distribution further underscore the variability in human-provided responses, which ranged from brief comments of fewer than 10 words to occasional detailed responses approaching 120 words.



**Figure 2:** length of comments of human grader and tool

The greater length of tool feedback can be attributed to its standardized and systematic alignment with LEED rubrics. In contrast, TA feedback often consists of concise and targeted comments, likely reflecting the time constraints and specific focus of human reviewers. While the extensive nature of tool feedback ensures comprehensive coverage of evaluation criteria, the brevity of TA feedback may be indicative of its more personalized and context-aware nature.

When considering the effectiveness of feedback, the greater length of tool-generated responses provides students with detailed guidance on how to meet specific criteria, making it particularly beneficial for iterative improvement in assignments structured around standardized frameworks like LEED.

### *Qualitative Analysis*

To assess the quality and effectiveness of feedback, a detailed qualitative analysis was conducted comparing the feedback provided by TAs and generative AI model. The analysis focused on five key dimensions derived from the LEED rubrics: LEED Certification Achievement, Reflection of Credit Requirements, Formatting, Realistic Implementation of Credits, and Grammar, Structure, and Clarity. Table 1 provides illustrative examples of feedback from both sources. The first column collects all the comments from the TA regarding the LEED narrative part of the current student assignment. These comments were written by the TA while reviewing the student's work, providing evaluations or asking questions about specific items or sentences. The second column contains comprehensive feedback generated by the tool after the student uploaded their assignment to it.

Table 1. TA Feedback vs Tool Feedback	
TA Feedback	Feedback From Tool
Should be broken down by credit with points claimed - see Assignment three presentation pdf	<p><b>**LEED Certification Achievement**</b></p> <ul style="list-style-type: none"> <li>- The narrative generally demonstrates the project's LEED certification potential, with efforts to achieve various credits, but lacks clear explanations for some prerequisites, which may raise concerns about the project's chances of achieving LEED certification.</li> <li>- More details on prerequisites, such as the definition of "onsite" and "simultaneous" for the Fundamental Commissioning and Verification credit (EA Prerequisite 2), are needed.</li> <li>- The narrative fails to address some LEED prerequisites, such as the Minimum Insulation Requirements of Exterior Envelopes (SS Prerequisite 1), and misaligned credits, such as the Enhanced Commissioning credit (EBOM Credit 3) and the Optimize Energy Performance credit (EA Credit 1).</li> </ul>
What strategies will your project employ to reach the credit requirements?	<p><b>**Reflection of Credit Requirements**</b></p> <ul style="list-style-type: none"> <li>- Most credit requirements for the SSc2, SSc5.1, and SSc5.2, for example, are correctly identified, but lack some detail, such as the implementation plan and responsible parties.</li> <li>- Several credit requirements, including the Fundamental Refrigerant Management credit (EA Prerequisite 2) and the Construction and Demolition Waste Management credit (MR Credit 1), are not adequately reflected, with missing implementation details.</li> </ul>
Non-roof measures are also required to gain full points for this credit	<ul style="list-style-type: none"> <li>- Mentioning credits without sufficient explanation is avoided in this narrative, but the explanations given for some credits could be more comprehensive.</li> </ul>
How will your project achieve this reduction in energy usage?	<p><b>**Formatting: Credit Names and Points Claimed**</b></p> <ul style="list-style-type: none"> <li>- Most credits are named and points claimed in this narrative, but minor inconsistencies are present, such as missing or misplaced credits, e.g., SS Credit 1.</li> <li>- Credit names or points claimed are generally consistent, but minor errors, e.g., incorrect points value for the Enhanced Indoor Air Quality Strategies credit (EQ Credit 3), are present.</li> </ul>
Which renewable energy systems does your project plan to employ?	<p><b>**Realistic and Detailed Implementation of Credits**</b></p> <ul style="list-style-type: none"> <li>- Vague or unrealistic explanations for some credits, such as the SSc2, SSc4.1, and EQ Credit 1, with minimal LEED compliance details.</li> <li>- The Implementation Plan lacks sufficient information on responsible parties, timelines, and budget, making it difficult to assess the feasibility of credit achievement.</li> <li>- No realistic strategies for achieving credit, such as the Construction and Demolition Waste Management credit (MR Credit 1) and the Enhanced Commissioning credit (EBOM Credit 3), are provided.</li> </ul>
Explanation for Building Life-Cycle Impact Reduction missing	<p><b>**Grammar, Structure, and Clarity**</b></p> <ul style="list-style-type: none"> <li>- Minor grammatical errors, such as missing articles and incorrect verb tenses, are present throughout the narrative, but mostly do not impact readability.</li> </ul>
Which regional priority credits did your project achieve?	<ul style="list-style-type: none"> <li>- Well-written and clear explanations for some credits are evident, but these are offset by more obscure passages that make comprehension challenging.</li> <li>- Still, some readability issues remain, such as the reiteration of repetitive statements in multiple sections, and minor sentence-level errors that intrude upon the narrative flow.</li> </ul>



The analysis reveals distinct patterns in the feedback styles of TAs and the tool. TA feedback often highlights specific and actionable questions tailored to individual student submissions. For instance, TAs directly ask for clarification, such as “What strategies will your project employ to reach the credit requirements?” or “Which renewable energy systems does your project plan to employ?” These questions encourage critical thinking and focus on gaps in student understanding. However, TA feedback frequently lacks the comprehensive detail needed to address broader structural issues, such as providing detailed explanations of LEED prerequisites or alignment with rubrics.

In contrast, the tool’s responses are more structured and standardized, closely aligned with LEED rubrics. For example, in the dimension of LEED Certification Achievement, the tool evaluates the narrative holistically, identifying both strengths and weaknesses, such as efforts made toward certification and missing prerequisites. The tool also provides detailed suggestions, including specific LEED prerequisites like “Minimum Insulation Requirements of Exterior Envelopes (SS Prerequisite 1)” or alignment issues with credits like “Enhanced Commissioning credit (EBOM Credit 3).” These detailed observations help students align their projects with the LEED criteria more effectively.

The tool further excels in addressing formatting and implementation details. Feedback on credit names and points claimed, for example, identifies inconsistencies or errors, such as incorrect point values for “Enhanced Indoor Air Quality Strategies credit (EQ Credit 3).” Additionally, the tool provides specific critiques on the lack of feasibility in implementation plans, including insufficient information on responsible parties, timelines, and budgets. These insights emphasize practical improvements necessary for meeting credit requirements.

One key advantage of the tool lies in its ability to deliver feedback promptly and iteratively. Students can submit their assignments multiple times and receive immediate, detailed feedback after each submission. This iterative process not only facilitates real-time learning but also enables students to refine their work systematically, addressing specific deficiencies highlighted by the tool. In contrast, TA feedback is inherently constrained by time, often delivered after a significant delay due to the need to review multiple student submissions. The lack of timeliness in TA feedback can hinder a student’s ability to make immediate improvements, potentially limiting their learning opportunities within the iterative design process.

In the dimension of Grammar, Structure, and Clarity, the tool identifies errors such as missing articles or incorrect verb tenses but offers a clear narrative structure overall. Conversely, TA feedback often avoids grammatical critiques, focusing instead on substantive content issues. Despite this, TA feedback may leave more room for interpretation, potentially leading to miscommunication.

However, the tool has limitations in providing nuanced, context-sensitive advice. While TAs offer personalized guidance that may resonate better with individual students, the tool’s feedback sometimes lacks the adaptability needed to address unique project contexts. For instance, vague or unrealistic explanations, such as “minimal LEED compliance details,” could leave students uncertain about how to refine their work.

### **Conclusion and Discussion**

This study investigated the viability of an LLM-powered feedback system in a sustainability-focused design context, comparing the generative AI model’s outputs against TA feedback. The results illustrate both the model’s potential for delivering detailed, rubric-aligned guidance and the enduring value of human insight. On one hand, the tool’s extensive coverage of LEED criteria can serve as a beneficial roadmap for students seeking comprehensive direction in sustainability design. On the other, TA feedback remains indispensable, providing context-sensitive remarks that are concise, immediately actionable, and tailored to individual students’ needs.

Notably, while lengthier feedback can create a broader perspective on project strengths and weaknesses, excessive detail may dilute clarity. Students might benefit from a mixed strategy that combines the AI's thoroughness with human-led focus on the most critical issues. Moreover, given that LEED-based assignments often involve nuanced considerations—such as diverse environmental, social, and economic factors—the personal guidance from TAs can bridge contextual gaps that an AI might overlook. Consequently, an ideal approach likely integrates automated tools to provide structural, iterative feedback alongside targeted human feedback for deeper or specialized questions.

### Future Work

Moving forward, an essential next step involves systematically evaluating the usability of the feedback tool. Future research will focus on conducting formal usability tests and analyzing user feedback to refine the tool's interface and improve the clarity and impact of its feedback. Such efforts will help align the tool more closely with educational objectives, ensuring that it not only supplies comprehensive rubric-based evaluation but also adapts to learners' varying needs and diverse project contexts.

### References

- Adams, R. S., Daly, S. R., Mann, L., & Dall'Alba, G. (2016). Being a Professional: Three Lenses into Design Thinking, Acting, and Being. *Design Studies*, 47, 123–153.
- Attali, Y., & Burstein, J. (2006). Automated Essay Scoring With e-rater® V.2. *Journal of Technology, Learning, and Assessment*, 4(3).
- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698–712. <https://doi.org/10.1080/02602938.2012.691462>
- Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31(2), 219–233. <https://doi.org/10.1080/03075070600572132>
- Castelblanco, G., Cruz-Castro, L. and Yang, Z., 2024. Performance of a Large-Language Model in scoring construction management capstone design projects. *Computer Applications in Engineering Education*, 32(6), p.e22796.
- Cruz Castro, L., G. Castelblanco, and P. Antonenko. 2024. LLM-based System for Technical Writing Real-time Review in Urban Construction and Technology.
- Fenner, R. A., Ainger, C. M., Cruickshank, H. J., & Guthrie, P. M. (2005). Embedding sustainable development at Cambridge University Engineering Department. *International Journal of Sustainability in Higher Education*, 6(3), 229–241. <https://doi.org/10.1108/14676370510607205>
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681–694. <https://doi.org/10.1007/s11023-020-09548-1>
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112.
- Hochreiter, S. (1997). Long Short-term Memory. *Neural Computation MIT-Press*.

- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign.
- Ihantola, P., Ahoniemi, T., Karavirta, V., & Seppälä, O. (2010). Review of Recent Systems for Automatic Assessment of Programming Assignments. *Proceedings of the 10th Koli Calling International Conference on Computing Education Research*, 86–93.
- Nicol, D., & Macfarlane-Dick, D. (2006). Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice. *Studies in Higher Education*, 31(2), 199–218. [Link](#)
- Kibert, C. J. (2016). *Sustainable Construction: Green Building Design and Delivery* (4th ed.). Wiley. <https://doi.org/10.1002/9781119174752>
- Kolko, J. (2010). Abductive thinking and sensemaking: The drivers of design synthesis. *Design Issues*, 26(1), 15–28. [https://doi.org/10.1162/DESI\\_a\\_00064](https://doi.org/10.1162/DESI_a_00064)
- Shin, J., Cruz-Castro, L., Yang, Z., Castelblanco, G., Aggarwal, A., Leite, W., Carroll, B., 2024. Understanding optimal interactions between students and a chatbot during a programming task, In: *Proceedings of the 2024 Winter Simulation Conference (WSC)*, pp. 21–30.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. Basic Books. [Link](#)
- Susnjak, T., & McIntosh, T. R. (2024). ChatGPT: The end of online exam integrity?. *Education Sciences*, 14(6), 656.
- VanLehn, K. (2006). The Behavior of Tutoring Systems. *International Journal of Artificial Intelligence in Education*, 16(3), 227–265.
- Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35, 24824-24837.
- Xie, H., Chu, H.-C., Hwang, G.-J., & Wang, C.-C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers & Education*, 140, 103599. <https://doi.org/10.1016/j.compedu.2019.103599>
- Zawacki-Richter, O., Marín, V., Bond, M., & Gouverneur, F. (2019). Systematic Review of Research on Artificial Intelligence Applications in Higher Education—Where Are the Educators? *International Journal of Educational Technology in Higher Education*, 16(1), 39.