Towards Scalable AI Feedback Systems: Preparing A Turing-Test-Inspired Experiment

Peter A. M. Ruijten-Dodoiu[1,†], Manual Oliveira[1,*,†] and Esther Ventura-Medina[1,†]

[1] *Eindhoven University of Technology, Eindhoven, Netherlands*

Abstract

Generative AI (GenAI) is increasingly positioned as a transformative tool in education. This work explores how GenAI can enhance reflective practices to foster growth mindsets—an essential trait for resilience, motivation, and lifelong learning. Drawing on frameworks for process-oriented education and ongoing experimentation, we propose an approach leveraging prompt engineering and fine-tuning to create human-like, iterative feedback on student reflections. To evaluate this, a Turing-test-style experiment is designed to assess whether students can differentiate between AI- and human-generated feedback. By addressing challenges such as feedback scalability, ethical transparency, and student trust, we hope to find ways to design more adaptive, inclusive, and growth-focused educational ecosystems.

Keywords

Growth Mindsets, Reflective Feedback, Generative AI, Turing Test

1. Introduction

As GenAI becomes embedded in educational practice, it brings opportunities to transform how we support student growth. Growth mindsets, defined as the belief in the malleability of one's abilities, are vital for fostering resilience, motivation, and lifelong learning [1]. However, embedding these traits into student development through self-reflection remains resource-intensive. As [1] note, fostering a belief in effort-driven improvement requires consistent reinforcement through feedback, making it essential to design scalable systems that align with this principle. However, providing individualized feedback, critical for meaningful reflection, is often infeasible in large student cohorts due to limited educator capacity [2].

Building on process-oriented approaches, this discussion investigates how GenAI can close the gap. Through techniques like prompt engineering [3] and fine-tuning [4], we aim to develop feedback mechanisms that replicate the nuance and depth of human feedback, enabling scalable yet personalized support. This innovation aligns with educational goals of fostering resilience and reflective learning while addressing practical constraints faced by educators [5].

1.1. Integrating GenAI with Reflective Practices

Reflective practices are integral to fostering growth mindsets, encouraging students to evaluate their learning processes and identify areas for improvement. However, many struggle to translate reflections into actionable outcomes. This gap is where GenAI may prove to be useful. By analyzing reflection patterns, GenAI tools can offer tailored feedback that highlights themes and provides

Second International Workshop on Generative AI for Learning Analytics, 2025.

* Corresponding author.
† These authors contributed equally.
p.a.m.ruijten@tue.nl (P.A.M. Ruijten-Dodoiu); m.j.barbosa.de.oliveira@tue.nl (M. Oliveira); e.ventura.medina@tue.nl (E. Ventura-Medina)
0000-0003-1900-3415 (P.A.M. Ruijten-Dodoiu); 0000-0002-6220-0695 (M. Oliveira); 0000-0002-1041-945X (E. Ventura-Medina)

actionable suggestions (see also [6]). For example, a student reflecting on what they have learned in the past few weeks, and what they can do to improve, might receive tailored feedback suggesting practical strategies, while also recognizing their efforts to improve. GenAI-generated feedback can be framed to emphasize the process of learning, reinforcing the value of effort, persistence, and incremental progress.

Fostering a growth mindset in students hinges on the quality and framing of feedback they receive. Effective feedback highlights effort, improvement, and actionable steps while avoiding fixed-ability language [1]. When students perceive feedback as constructive and tailored to their needs, they are more likely to internalize a belief in their capacity to grow and improve. This is particularly important in reflective practices, where the goal is not just to evaluate past performance but also to inspire iterative learning and self-improvement. Feedback that reinforces effort and improvement can help students view challenges as opportunities for growth, aligning directly with the principles of a growth mindset.

However, as we have seen, providing such feedback at scale is a persistent challenge in education. GenAI tools offer a potential solution by generating feedback that mirrors these growth-oriented principles. For example, AI feedback can be explicitly framed to recognize effort: "Your reflection shows strong engagement with the topic. Keep building on this effort.", emphasize process: "Your challenge in learning new applications of the theory demonstrates a valuable learning opportunity to refine your approach.", and provide actionable steps: "Consider setting smaller, achievable goals to improve your learning in future courses."

We are preparing an experiment that tests how well AI can replicate these growth-oriented qualities in feedback. The experiment will assess whether GenAI feedback resonates with students as deeply as human feedback in promoting growth mindsets. Our methodology integrates prompt engineering and fine-tuning, allowing GenAI to adapt to specific educational contexts. Prompt engineering involves strategically crafting inputs to guide AI outputs [3], while fine-tuning uses targeted training data to optimize AI for specific tasks [5]. By applying both methods, we aim to balance adaptability and precision in generating feedback that aligns with educational objectives. In our process-driven learning environments, such as challenge-based learning (CBL), this approach holds particular promise. Regular self-reflections tied to Intended Learning Outcomes (ILOs) already form part of the curriculum, creating an opportunity for GenAI to contextualize feedback within a student's broader learning journey.

2. Turing-test experiment: Evaluating feedback authenticity

Paradigms like CBL put emphasis on self-reflection which in turn leads to a need for timely processing of student input. This is where applications of GenAI could be introduced. One of the key challenges in integrating GenAI into education is ensuring that the feedback generated by AI is both perceived as valuable and trusted by students. For GenAI tools to support reflective practices effectively, they must emulate the depth, nuance, and context-awareness of human feedback. However, it is equally important to validate whether students can discern differences between AI- and human-generated feedback and whether they view AI feedback as credible and actionable. To address this, we propose the experiment will use a Turing-test-style design. This experiment is grounded in two objectives:

- To evaluate the quality of AI-generated feedback against human standards.
- To assess how closely AI feedback aligns with students' expectations for meaningful and contextually relevant feedback.

The ability of AI tools to analyze patterns in reflective entries, identify gaps, and provide targeted suggestions makes them invaluable in settings where timely, personalized feedback is traditionally constrained by resources. However, scaling such tools while maintaining quality introduces a unique challenge: balancing the need for efficiency with the expectation for human-like responsiveness and insight. The Turing-test paradigm is particularly well-suited for this challenge. As an example, modified variants of the Turing test have recently been employed in psychological research to understand which type of social information people find most likely to be AI-generated [7, 8]. Such modified variants of the Turing test allow to evaluate not just the technical capabilities of GenAI but also its ability to meet human expectations in an educational context. This experiment places students in the role of evaluators, asking them to compare feedback generated by AI and human educators for their reflective entries. By doing so, it aims to explore two critical questions:

1. How do students perceive the quality of AI-generated feedback compared to that of their instructors?
2. Does the feedback from AI meet the pedagogical and emotional needs students associate with human input?

To answer these questions, participants will evaluate two feedback samples for each reflection—one generated by a teacher and the other by a GenAI model trained for this specific task with real feedback data. Participants will identify which feedback they believe is human-authored and which is AI-generated. This study provides insights into the perceived authenticity, quality, and relevance of AI feedback. The experiment thus examines whether GenAI feedback can mirror the motivational and process-oriented framing essential for fostering growth mindsets. The data collection for the experiment is planned in February 2025.

2.1. Addressing Challenges and Opportunities

The integration of GenAI into reflective practices presents immense opportunities for fostering growth mindsets, but it also requires addressing specific challenges to ensure its ethical, effective, and sustainable use. Insights from the Turing-test experiment provide a foundation for navigating these challenges, enabling thoughtful deployment of GenAI tools in education. This section explores three critical areas that should be addressed to maximize the potential of AI-driven feedback.

2.1.1. Ethical Transparency

For GenAI tools to foster trust among students, it is essential to maintain transparency about their role in generating feedback. Students must understand how AI operates, what its limitations are, and how it complements (rather than replaces) human educators. Lack of transparency can lead to distrust or misconceptions about the reliability of AI-generated feedback, potentially undermining its effectiveness.

For reflective feedback that is written by GenAI to be successfully adopted by students, it is important to clearly communicate to students when and how AI-generated feedback is used, ensuring they view it as a tool to enhance their learning, rather than one that replaces instructors. Students should also be involved in discussions about the role of AI in education, fostering a sense of ownership and engagement. A challenges is to balance simplicity and depth in explanations of the AI's role in education. Overly technical descriptions might confuse students, while oversimplifications could misrepresent its capabilities.

2.1.2. Bias Mitigation

AI models are susceptible to biases embedded in their training data, which can lead to unfair or unbalanced feedback. For instance, a model trained on predominantly positive and optimistic reflections might struggle to offer constructive criticism to students. Ensuring fair and suitable feedback is crucial for broad adoption.

We are currently collecting data from multiple courses in which students actively reflect on their own growth and development, and as such are building a dataset that represents real student struggles and real instructor feedback. Although the dataset includes reflections from diverse courses, further diversification will be necessary to address biases in language patterns, cultural context, and feedback framing. Also, whether these data will be sufficient to train a model in giving appropriate feedback is yet unknown.

2.1.3. Balancing Automation and Human Oversight

While GenAI excels at automating routine feedback tasks, human oversight remains indispensable for nuanced, complex reflections that require deeper contextual understanding. A hybrid approach, combining AI and human educators, can balance scalability with personalized attention. We are explicitly stating that the goal is not to replace feedback from human instructors, but rather augment it.

Feedback written by GenAI could be checked and edited by instructors to safeguard its quality and appropriateness. This allows educators to focus on higher-order coaching tasks while delegating routine feedback generation to AI. As such, we could use AI feedback as a starting point, enabling educators to refine it for specific cases. Challenges include ensuring that educators have the skills and tools to effectively interpret and build upon AI-generated feedback, and that we should avoid over-reliance on AI, as this would reduce the positive influence of human interactions in learning. In other words, human oversight remains crucial for addressing reflections that delve into personal challenges, ensuring that the feedback considers the emotional and situational context.

3. Insights gained from the experiment

The Turing-test experiment offers a unique opportunity to address these challenges by evaluating how students perceive and engage with AI feedback. Insights from the experiment can guide improvements in transparency, reduce biases, and optimize the balance between automation and human input. Through deliberate attention to these challenges, GenAI can move beyond simply scaling feedback to becoming a transformative partner in fostering growth mindsets and reflective learning. This experiment serves as a critical validation step for the integration of GenAI tools into reflective practices. If successful, it will demonstrate not only the capability of GenAI to emulate human feedback but also its potential to enhance the scalability and personalization of reflective education. Importantly, it will shed light on the specific conditions under which students perceive AI-generated feedback as valuable, thus informing the design of future tools and frameworks.

By situating students as active evaluators, the Turing-test experiment also fosters transparency and trust in AI systems. It creates a collaborative space where students can critically engage with the evolving role of technology in their learning journey, laying the groundwork for more inclusive and reflective educational ecosystems.

4. Future directions and Conclusion

The potential applications of GenAI extend beyond individual courses, offering opportunities to transform broader educational frameworks. GenAI could augment and catalyze the process of extracting insights from student reflections across multiple courses, providing educators with insights into student development trends. This data could inform interventions and curriculum design. Such longitudinal insights could enable educators to tailor curriculum design to address recurring student challenges, fostering a culture of iterative improvement across courses.

From formative assessments to personalized learning paths, the scalability of GenAI positions it as a cornerstone of future educational innovations. The integration of Generative AI into reflective practices offers transformative potential for fostering growth mindsets at scale. By providing nuanced, personalized feedback through prompt engineering and fine-tuning, GenAI addresses critical challenges of scalability and quality. The proposed inspired Turing-test experiment provides a rigorous evaluation framework, ensuring that these tools meet the high standards expected in education.

As we continue to explore these opportunities, GenAI emerges as a partner in creating more reflective, resilient, and growth-focused educational ecosystems. Through this integration, we aim to empower students to achieve not only academic success but also the lifelong learning skills necessary to thrive in an ever-evolving world.

Declaration on Generative AI

The authors declare that Generative AI tools have not been used to prepare the manuscript.

References

[1] Dweck, C. S., & Yeager, D. S. (2019). Mindsets: A View From Two Eras. Perspectives on Psychological Science, 14(3), 481-496. https://doi.org/10.1177/1745691618804166

[2] Joyner, D. A. (2017, April). Scaling expert feedback: Two case studies. In Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale. https://doi.org/10.1145/3051457.305145

[3] Giray, L. (2023). Prompt engineering with ChatGPT: a guide for academic writers. Annals of biomedical engineering, 51(12), 2629-2633. https://doi.org/10.1007/s10439-023-03272-4

[4] Parthasarathy, V. B., Zafar, A., Khan, A., & Shahid, A. (2024). The ultimate guide to fine-tuning llms from basics to breakthroughs: An exhaustive review of technologies, research, best practices, applied research challenges and opportunities. https://doi.org/10.48550/arXiv.2408.13296

[5] Nicol, D. J., & 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. https://doi.org/10.1080/03075070600572090

[6] Lin, C. J., Lee, H. Y., Wang, W. S., Huang, Y. M., & Wu, T. T. (2024). Enhancing reflective thinking in STEM education through experiential learning: The role of generative AI as a learning aid. Education and Information Technologies, 1-23. https://doi.org/10.1007/s10639-024-13072-5

[7] Oliveira, M. J. B., Brands, J., Mashudi, J., Liefooghe, B., & Hortensius, R. (2023). Perceptions of Artificial Intelligence system's aptitude to judge morality and competence amidst the rise of chatbots. https://doi.org/10.31234/osf.io/vsxr4

[8] Tucciarelli, R., Vehar, N., Chandaria, S., & Tsakiris, M. (2022). On the realness of people who do not exist: The social processing of artificial faces. iScience, 105441. https://doi.org/10.1016/j.isci.2022.105441