WIP: Just-In-Time AI Assisted Formative Feedback for Written, Oral, Team-Based Assessment Tasks: What Worked, What Didn't and Why

May Lim
School of Chemical Engineering
The University of New South Wales
Sydney NSW 2052 Australia
m.lim@unsw.edu.au

***Abstract*—This innovative practice WIP paper describes the use of three large language model-based pre-trained AI (LLM-based AI), Inflection's Pi, Azure OpenAI GPT-3.5-Turbo and Azure OpenAI GPT-4 models to provide insightful and timely feedback across written, oral, and team-based assessment tasks in a capstone engineering design course. These LLM-based AI could analyze the content of artefacts produced in formative verbal and written tasks, ensuring that the students include relevant information in their report, and the report requirements such as structure, grammar, style, mechanics etc. are met. The models were also used to analyze the meeting transcripts of student teams, thus allowing the teamwork process and contributions from each member of the student team to be monitored closely. The integration of LLM-based pre-trained AI increased the timeliness and effectiveness of formative feedback on students' design and teamwork processes, thereby fostering a more adaptive and personalized learning experience. Any missteps or misunderstandings on the part of the students regarding the task requirements, as well as any issues arising within team interactions can be promptly communicated to the instructor for immediate resolution. While LLM-based pre-trained AI holds significant promise in transforming feedback practice, it is important to acknowledge that there are still limitations and barriers to practical implementation in the classroom. The feedback produced by LLM-based pre-trained AI lacks nuanced contextual understanding. The integration of LLM-based pre-trained AI into feedback practices holds transformative potential for education. On one hand, the models offer the promise of responsive and personalized learning, and the potential to foster critical thinking and problem-solving skills through interactive and adaptive learning platforms. Conversely, the reliability of the models as a feedback tool and students' receptions to their use remains ambiguous. We conclude the paper by proposing some strategies to overcome these limitations and support academics in applying LLM-based pre-trained AI in feedback practice.**

Keywords—Feedback, Computer-based instruction, Design based learning, Written communication, Oral presentations

I. Introduction

The rapid advancement of technologies that can revolutionize every facet of human existence, and the widening gap between classroom instruction and real world practice have prompted many higher education providers to incorporate professional skills and user-centred design principles into their engineering curricular. The integration of technical knowledge with instructions on design thinking, project management, teamwork, as well as effective written and verbal communication is a forward-thinking approach that will equipped the next generation of engineering graduate with the skills needed to thrive in an increasingly complex professional environment. Project-based capstone design courses often serve as the vehicle to deliver these instructions, allowing students to engage with authentic problems and devise innovative solutions that are not only technically sound but also economically sustainable and socially responsible [1].

While authentic, project-based design courses have been reported to engage and motivate students, their implementation still presents several challenges, one of which is the need for course instructors to be less prescriptive in their instruction and provide *just-in-time* feedback to guide student performance [2,3]. This ensures students have the opportunities to independently explore diverse approaches and develop their critical thinking and problem-solving skills *as needed* [4]. Just-in-time feedback, however, requires the course instructor to anticipate the kind of feedback that is needed, when and where the students experience the need, and provide feedback that is focused, personalized, and timely.

Recent advancements in computer-based learning systems present a solution to the challenge of providing just-in-time feedback in project-based capstone courses. For instance, learning analytics and educational data mining technique have been applied to early and accurate identification of struggling students in learning engineering design [5]. The emergence of Large Language Model (LLM) driven artificial intelligence, as exemplified by OpenAI's GPT, Google's PaLM and Gemini, Meta's LLaMA and Anthropic's Claude and Meta's LLaMA family of models further reshaped the landscape. There is one report by Sung et al. [6] to date on the use of LLM-driven AI model to generate personalized feedback and encouragements to students in a makerspace course, but the quality of feedback is inconsistent.

The present study examine how three publicly accessible LLM-driven AI models, Inflection's Inflection-1 and Azure OpenAI gpt-3.5-Turbo and gpt-4, can be employed to evaluate and provide feedback to students on the quality of formative written and oral assessment tasks in a 10-week long, project based design course. This course focuses on the early stages of the design process, and the development of sustainable, user-centric and fit-for-purpose solutions through the application of design thinking methodologies. The students were assigned a project on the valorization and monetization of spent coffee ground (a waste product). The course also aims to develop the students' readiness for professional practice by deepening their understanding and skills in effective project management, teamwork, and communication.

II. AI Assisted Formative Feedback

A. Feedback on Student-Industry Interactions

At the start of the course, students role-play conversation with industry representative at a trade show with Pi, a chatbot based on Inflection's Inflection-1 model, via a web interface at https://pi.ai [8]. The students then attend a trade convention that is related to the course where they were assigned the task of interacting with industry representative at the convention and finding information related to their project. The students shared their experience of the convention, as well as any information they have gained from their interactions with industry representatives at the convention on Padlet, a virtual bulletin board. The students' accounts were then summarised use Azure Open AI's gpt-35-turbo model [9].

B. Feedback on Team Dynamics and Interactions

Throughout the course, the students work together in a team of six students. The student teams held regular online meetings on Microsoft Teams outside of class time. The instructor set meetings to record automatically and generate a transcript at the end. The transcripts were then analysed using Azure Open AI's gpt-4 model via the API [9].

C. Feedback on Written and Oral Assessment Tasks

The summative assessment task for the design course consists of a technical report submitted justifying their design solution, as well as a presentation where the student team explain their design process and pitch their design solution. Students are required to write the report 'from scratch' on Microsoft Word documents that were saved to a SharePoint library associated with the course. Mid-way through the term, complete draft of the students' reports were analysed using Azure Open AI's gpt-4 model via the API. The student also practiced describing their solution process and pitching their idea in class throughout the term. These impromptu practice pitches were captured and transcribed using Microsoft Word Dictate, an in-built speech-to-text tool in the word processing app, and the transcribed text was saved as a Microsoft Word document to a SharePoint library associated with the course. The students also have an opportunity to practise delivering their final presentation pitch; this practice pitch was recording and transcribed in Microsoft Teams, in the same manner as the student teams' meetings. The students who were the audience were also required to provide written feedback on the quality and content of the pitch on Padlet. The transcribed text and written feedback were also analysed using Azure Open AI's gpt-4 model via the API.

III. Discussion

A. Utility of the AI Assisted Feedback from the Instructor's Perspective

An autoethnographic account of the instructor's experiences in implementing LLM-based pre-trained AI assisted formative feedback in their course is provided herein. The instructor designed the course and therefore had 'frontline' insight into the support needs of their students. Due to the rapidly evolving nature of LLM-driven AI technologies in 2023, this self-enquiry was used to provide a more nuanced understanding of AI-driven innovation in teaching practice [7].

Student-Industry Interactions • Pi was designed to maintain an interactive text or voice-based dialogue with a human user in a personable manner [8]. For this reason, Pi present an advantage in role-playing over other chatbots such as Open AI's ChatGPT which produces a more 'mechanical' response. Through interacting with Pi, students were able to gain a quick overview of the problem space, discover well-publicized design solutions, brainstorm new and untested solutions, and refine the questions they had prepared for trade conventional. Pi's ability to scaffold learning extends to advising the students on how to approach new and untested solutions. For example, when a student proposed using the spent coffee grounds as pet food, Pi recommended an investigation into the safety of coffee grounds for animal consumption and an examination of whether the grounds possess flavours that may be distasteful to pets. If a student proposed an illogical or nonsensical solution, Pi will direct the students to a more pertinent solution. For example, when a student suggested using the spent coffee ground as a vacuum cleaner, Pi noted this is unprecedented and suggested that the student should instead consider using the spent coffee ground in a household cleaning product.

Interestingly, the students' role-playing with Pi did not reflect real-world interactions with industry representatives at the trade convention. AI generated summary showed the later to be more wide-ranging and less technical due to the fact that industry representative at trade convention are predominantly sales representatives or proprietors of small-to-medium enterprises, who invariably infuse the discourse with local contexts and values. This juxtaposition between AI and human interactions provides a unique opportunity to discuss with students the limitations of AI as an information source.

Team Dynamics and Interactions • Azure Open AI's gpt-4 model was able to summarise meeting discussion, including key decisions, planned actions and follow-ups, as well as extract information about the contributions and question asked by each attendees. The instructor primarily used the information to detect if the student teams had any misconceptions about the project and assessment tasks. An ancillary use is to ascertain that every member of the student team is engaged and working collaboratively with their peers, thus mitigating team dynamic issues such as interpersonal conflict and free-riding.

Preliminary analysis further showed meetings were where student teams collectively interpret the requirements of assessment tasks, and make pivotal decisions regarding the approach and direction of their projects. Moreover, it was found that misconceptions about assessment task requirements predominantly originate within the confines of these meetings. The ability to detect and respond to students' misunderstandings without the need to monitor personally every aspect of the student's interactions with each other is an important development. The instructor can now provide immediate and targeted feedback, ensuring that misconceptions are addressed promptly, without disrupting the flow of collaborative work and the critical dialogues between members of a student team. This maintains the authenticity of the assessment task.

Written and Oral Assessment Tasks • Azure Open AI's gpt-4 model was able to provide feedback on the grammar and mechanics of the students' report, however, the utility of this feedback was low because students already have access to other writing assistant tools and academic writing services. To a certain degree, the gpt-4 model can also identify if the students' reports and oral presentations had integrated elements stipulated in the assessment rubrics. The model can also summarize the content of reports, presentations and peer feedback, as well as generate questions related to the content. However, the gpt-4 model cannot reliably evaluate the quality of the report. For instance, it cannot ascertain if information contained within the report is relevant or correct. The gpt-4 model's response was also variable and occasionally, the model would 'hallucinate' and include extraneous information. This is expected as LLM-based pre-trained AI models lack evaluation capability [10].

B. Students' Perceptions of AI Assisted Feedback

A survey was developed to understand students' experiences and perceptions of AI-assisted feedback. A total of 17 students' responses were received (32% of cohort). This is typical in an academic setting, and is due to the fact that the students were only invited twice to complete the survey in order to reduce the number of inauthentic responses.

Figure 1 shows students' responses, gauged on a Likert scale, to questions about their experience with AI-assisted feedback (top), and their perception of the quality of AI-generated feedback relative to feedback provided by the course instructor and their peers for different assessment tasks (bottom). An open question "Can you describe your experiences with AI generated feedback?" provide insights into the student's varied experiences involving AI-generated feedback. Some students found AI feedback valuable, helping them enhance their work through recommendations. Others, however, reported negative experiences where the AI failed to accurately assess their work. There appeared to be a recurring theme in the response regarding how the AI had failed to recognise some sections of the reports, or failed to capture certain aspects of the student's team contribution. The students' affective response to feedback produced by AI were generally positive. However, the students did not perceive the feedback to be better than feedback provided by their instructor or peers. This response can be explained by the fact that most of the feedback generated by the AI lacked nuanced contextual understanding and therefore is of low utility to the students. Another possible explanation is students make lack the evaluative judgement and feedback literacy needed to make full use of the AI generated feedback [11,12].

IV. Conclusions and Future Investigations

The integration of large language models (LLMs) based AI into feedback practices holds transformative potential for education. On one hand, LLM-based pre-trained AI offers the promise of responsive and personalized learning, and the potential to foster critical thinking and problem-solving skills through interactive and adaptive learning platforms. Conversely, the reliability of LLM-based pre-trained AI as a feedback tool and students' receptions to their use remains ambiguous. We posit that gpt-4 model can serve as a supplementary tool for instructors in the provision of feedback but it is unreliable as a standalone feedback mechanism. The instructor's discipline expertise and intuition on what will or will not work in the classroom is still critical.

While not addressed directly in this paper, there are also concerns related to ethical use of AI, data privacy, and the digital divide between institutions with and without access, as well as capacity and capability to wield these tools. Specifically, the toolchain needed to deploy AI tools and apps in a manner that met the institution's data, privacy, cybersecurity policies are not always unavailable. The cost of building and implementing custom build tools is also prohibitive. The bricolage and ad hoc innovation approach used herein, where the instructor used Python script and Microsoft application such as Word and Teams to 'stitch' together a feedback tool is unlikely to be scalable [13,14]. We hereby proposed that for education innovation around LLM-based pre-trained AI to thrive, the institution will have to develop: (1) short and long term institutional AI strategy and roadmap; (2) internal or external partnerships that bridges the institution's capability and capacity gaps; (3) robust guiderail and policies around the use of AI in the classroom; (4) program to identify and upskill teaching staff wanting to upskill; and (5) process for scaling and translating successful implementations. Some of these points can form the grounds for future investigation.

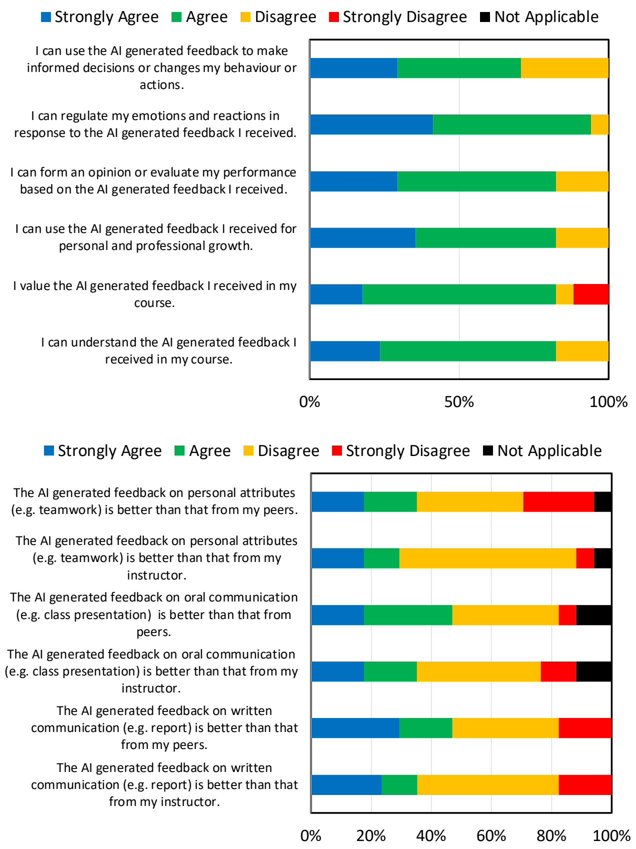

Fig. 1. Students' responses to questions about their affective experience with AI-assisted feedback (top) and their perception of the quality of AI-generated feedback relative to feedback provided by the course instructor and their peers for different assessment tasks (bottom).

Acknowledgment

The author thanks [Names] for their assistance in the design of the survey question and collection of the survey response.

References

[1] G. Charosky, L. Hassi, K. Papageorgiou and R. Bragós, "Developing Innovation Competences in Engineering Students: A Comparison of Two Approaches," Eur. J. Eng. Educ., vol.47, no. 2, pp. 353-372, 2022. https://doi.org/10.1080/03043797.2021.1968347

[2] D. M. Gilbuena, B. U. Sherrett, E. S. Gummer, A. B. Champagne and M. D. Koretsky, "Feedback on Professional Skills as Enculturation into Communities of Practice," J. Eng. Educ., vol. 104, pp. 7-34, 2015. https://doi.org/10.1002/jee.20061

[3] G. N. Novak, E. T. Patterson, A. Gavrin and W. Christian, Just-in-Time Teaching: Blending Active Learning and Web Technology, Saddle River, NJ: Prentice Hall, 1999.

[4] M. J. Prince and R. M. Felder, "Inductive Teaching and Learning Methods: Definitions, Comparisons, and Research Bases," J. Eng. Educ., vol. 95, no. 2, pp. 123–138, 2013. https://doi.org/10.1002/j.2168-9830.2006.tb00884.x

[5] W. Xing, B. Pei, S. Li, G. Chen and C. Xie. "Using Learning Analytics to Support Students' Engineering Design: The Angle of Prediction," Interact. Learn. Environ., vol. 31, no. 5, pp. 2594-2611, 2023. https://doi.org/10.1080/10494820.2019.1680391

[6] G. Sung, L. Guillain, and B. Schneider, "Can AI Help Teachers Write Higher Quality Feedback? Lessons Learned from Using the GPT-3 Engine in a Makerspace Course," in P. Blikstein, J. Van Aalst, R. Kizito and K. Brennan (Eds.), Proceedings of the 17th International Conference of the Learning Sciences (ICLS 2023), pp. 2093-2094. https://doi.org/10.22318/icls2023.904961

[7] J. Mao, E. Romero-Hall and T. C. Reeves. "Autoethnography as a Research Method for Educational Technology: A Reflective Discourse," Educ. Technol. Res. Dev., 8 September 2023. https://doi.org/10.1007/s11423-023-10281-6

[8] "Inflection-1," Inflection AI, Palo Alto, CA, USA, Tech. Memo. Accessed: May. 13, 2024. [Online]. Available: https://inflection.ai/assets/Inflection-1_0622.pdf

[9] "Azure OpenAI Service Models", Microsoft, Redmond, Washington, U.S, webpage. Accessed: May. 13, 2024. [Online]. Available: https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/models

[10] J. Oh, E. Kim, I. Cha and A. Oh, "The Generative AI Paradox on Evaluation: What It Can Solve, It May Not Evaluate", Accessed: May. 13, 2024. [Online]. Available: arXiv:2402.06204 [cs.CL].

[11] Darvishi, A., H. Khosravi, S. Sadiq, D. Gašević and G. Siemens. "Impact of AI Assistance on Student Agency," Comput. Educ., vol. 210, pp. 104967, 2024. https://doi.org/10.1016/j.compedu.2023.104967

[12] Tai, J., R. Ajjawi, D. Boud, P. Dawson and E. Panadero. "Developing Evaluative Judgement: Enabling Students to Make Decisions About the Quality of Work," High Educ., vol. 76, pp. 467-481, 2018. https://doi.org/10.1007/s10734-017-0220-3

[13] P. Bouvier-Patron, "Bricolage – From Improvisation to Innovation: The Key Role of 'Bricolage'," in Innovation Economics, Engineering and Management Handbook: Vol. 2 Special Themes, edited by D. Uzunidis, F. Kasmi and L. Adatto, 67-73. John Wiley & Sons, New York, 2021. https://doi.org/10.1002/9781119832522.ch6

[14] S. McHugh, N. Carroll and C. Connolly. "Low-Code and No-Code in Secondary Education - Empowering Teachers to Embed Citizen Development in Schools," Computers in the Schools, pp. 1-26, 2023. https://doi.org/10.1080/07380569.2023.2256729