International Society of the Learning Sciences

Can AI Help Teachers Write Higher Quality Feedback? Lessons Learned From Using the GPT-3 Engine in a Makerspace Course

Gahyun Sung, Harvard University, gsung@g.harvard.edu
Léonore Guillain, Taskbase, leonore.guillain@gmail.com
Bertrand Schneider, Harvard University, bertrand_schneider@gse.harvard.edu

Abstract: We explore how a cutting-edge language model, GPT-3, can be used to augment and assist periodic feedback writing in a makerspace course. Personalized messages were generated using student data then edited and combined with human instructor feedback. We discuss the lessons learned: namely, AI did well in summarizing work and positive encouragements, yet could write off-target feedback for struggling students. An initial interview with an instructor revealed that future iterations must consider ways to formalize and manage human expert roles.

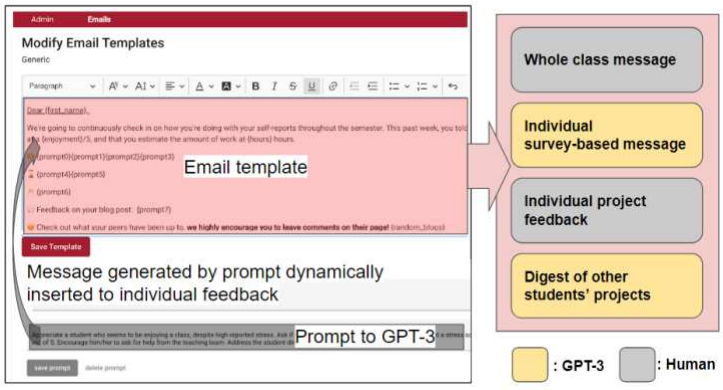
Introduction and setting

Could AI technologies augment periodic feedback, a standard practice in many learning environments? We study how this might be done in one context less explored but ripe for automated data-based feedback systems: a makerspace, a physical project-based learning environment where students monitoring can be a challenging task. Our context is a semester-long course on digital fabrication at a graduate school of education in the U.S., where students (n = 19) applied standard makerspace tools for a personal project. For many, the course is not only a chance to learn a new skillset, but to revisit their identities and perspectives on learning. Yet, we also saw students who quietly struggle to keep up. While instructors reaching out could help these students, this requires keeping tabs on multiple data sources and writing time-consuming feedback. Our study tests whether AI could assist this instructor responsibility, with the current paper focusing on results of our explorations on what tasks AI could and could not do well in an educational context, as well as lessons learned on supporting AI-human partnership.

A platform for developing a human-AI partnership: Preliminary findings

Our platform used GPT-3, a cutting-edge generative AI model that can generate humanlike speech with a high level of sophistication and naturalness. During an exploratory phase, we fed various blog posts written by students to GPT-3 to identify when and how the system might be used in a project-based environment. Our explorations led us to believe GPT-3 may provide the starting point of a helpful feedback: it was able to 1) use supportive language to appreciate projects (e.g., imaginative, impressed); 2) recognize the work put in (e.g., I appreciate the effort, I'm impressed); and 3) value students' strategies (e.g., you did a great job of staying organized throughout the process). Based on our explorations, we developed a writing platform seen in Figure 1. The system is shown on left, while on right is the structure of the periodic feedback sent out.

Figure 1
Interface for generating feedback through human-AI collaboration & feedback structure



How can GPT-3 be used in project-based learning contexts?

Our platform used GPT-3 for two tasks: generating encouraging messages for students and summarizing student blog posts on their weekly projects. Again, GPT-3 showed clear value in summarizing student reflections of their own work. The machine-generated summaries were concise and accurate, giving instructors overviews on how student work was progressing. They were also close enough to natural speech to directly add to emails, making it

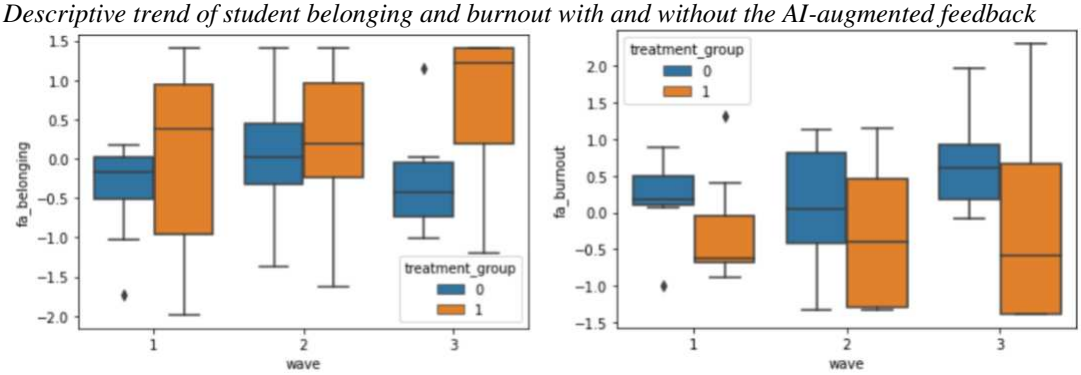
easier to encourage peer engagement. GPT-3 also performed acceptably in writing congratulatory messages. However, we noticed that feedback was often insensitive for struggling students. One prompt ("Write a supportive message to a student with {*stress* (*high)} and {*enjoyment* (*low)}.") may result in messages of varying quality:

> **Message 1**: We're sorry to hear that you're not enjoying the class and that you're feeling so stressed. Can you tell us more about what's making you feel this way? We want to make sure we're doing everything we can to support you. Are there any specific things we can do to help you feel more comfortable and successful in the class?
> **Message 2**: We think that you're not enjoying the class and that you're feeling slightly stressed. I encourage you to reach out to the teaching team for help.

While the first message is not perfect, the second message is too short to reasonably expect any effect, and the term "slightly" depreciates the level of reported stress. While further tinkering with the prompt could result in better feedback, this may defeat the purpose of making the process easier for human instructors.

Below, we show a descriptive graph of self-reported levels of belonging and burnout for students who did and did not receive the AI-augmented areas of the feedback (i.e., yellow boxes in figure 1). Students who received augmented feedback (orange) seem to report higher belonging and lower burnout, but further analyses on the final data will test these differences statistically and explore the effects with student interviews.

Figure 2
Descriptive trend of student belonging and burnout with and without the AI-augmented feedback



There were also two lessons learned from involving instructors with the platform. One, we quickly realized that human input was an indispensable, limited resource. The makerspace facilitator we interviewed noted that this seemed to add a layer of complexity to writing feedback. While the learning curve may be justifiable for a larger class - for instance, this may make feedback possible in a course with 100+ students -, we acknowledge that these concerns hit a pain point for future iterations. Second, the facilitator noted that for instructors to put trust in these messages, a basic understanding of how AI worked and where it failed was necessary. Conversely, our interviewee also pointed out that other facilitators might be too trusting of the output, and neglect to thoughtfully edit AI feedback. This is a central tension: how do we prompt users to think deeply about the quality of their feedback, but at the same time lessen workload? These questions will be guiding our next iteration efforts.

Conclusion

To sum up, the design and implementation of an AI-augmented feedback system revealed that GPT-3 is helpful in summarizing written products and writing words of encouragement, yet can write off-target feedback for struggling learners. An instructor interview also revealed that we need to consider resource constraints and ways to build trust in AI. In future work, we hope to build system-side quality safeguards, e.g., automatically 'flag' sub-par or repetitive messages. We also hope to include existing richer data sources in the input, such as open-ended student feedback, and in particular the camera-based location data from the makerspace, an objective measure of work patterns inside the space. These additional inputs are hoped to increase the accuracy and relevance of AI-generated feedback. On the other end of the pipeline, we also hope to seek advice from students and instructors about effective feedback in different scenarios, to create better prompts for GPT-3. Throughout these improvements, we aim to continuously engage instructors and learners in an iterative co-design process.

References

Dale, R. (2021). GPT-3: What's it good for?. *Natural Language Engineering, 27*(1), 113-118.