

Purdue University

Purdue e-Pubs

Purdue AI in P-12 Education Conference

Nov 11th, 12:00 AM

Evaluating LLM-Generated Personalized Text Content for Middle School Science Students

Michael Vaccaro Jr.

University of Connecticut, michael.t.vaccaro@uconn.edu

Mikayla Friday

University of Connecticut, mikayla.friday@uconn.edu

Arash Zaghi

University of Connecticut, arash.esmaili_zaghi@uconn.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/purdueaip12ed>

Vaccaro, Michael Jr.; Friday, Mikayla; and Zaghi, Arash, "Evaluating LLM-Generated Personalized Text Content for Middle School Science Students" (2024). *Purdue AI in P-12 Education Conference*. 1. <https://docs.lib.purdue.edu/purdueaip12ed/2024/briefs/1>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

EVALUATING LLM-GENERATED PERSONALIZED TEXT CONTENT FOR MIDDLE SCHOOL SCIENCE STUDENTS

Michael Vaccaro Jr., Graduate Student, University of Connecticut, michael.t.vaccaro@uconn.edu

Mikayla Friday, Graduate Student, University of Connecticut, mikayla.friday@uconn.edu

Arash Zaghi, Professor, University of Connecticut, arash.esmaili_zaghi@uconn.edu

Abstract

Generative artificial intelligence and large language models (LLMs) have proven to be disruptors in education due to their ability to produce human-like text, placing these models under heavy scrutiny. However, LLMs embody a diverse knowledge base and have been shown to be few-shot learners (Brown et al., 2020) that can quickly adapt their output in response to user-provided context. Together, these facets situate LLMs as powerful tools capable of developing personalized learning materials for K-12 students without the need for expansive training data. As this potential has yet to be evaluated in literature, this study aims to investigate the ability of LLMs to adapt science texts to middle school students' learning preferences.

Methods and Results

We ran a small-scale experimental ($n = 23$) randomized controlled trial (RCT) to evaluate the potential of GPT-4, an LLM developed by OpenAI, to develop personalized learning materials. To do so, we first divide the personalization task into four stages in which we (1) gather data on a student's learning preferences, (2) build a profile that captures the students' preferences, (3) generate a text that is either aligned or misaligned with the student profile for the experimental and control groups in the RCT, respectively, and (4) test the effectiveness of the LLM in developing the personalized text. In this RCT, GPT-4 was used for stages 2 and 3.

Stage 1, in which student's learning preferences were identified from their selections within four pairs of text, was standardized for all participants to provide a consistent basis for the personalization. We found it helpful to use an existing model to ensure the texts within each pair represented different writing styles. Specifically, we used the four dimensions of the Felder-Silverman model (Felder & Silverman, 1988) to develop pairs of text that targeted the two extremes of each dimension. To test the effectiveness of the LLM, the personalized text generated in stage 3 was tested against a generic text two times. Students in the experimental group were expected to select the personalized text at a higher rate than those assigned to the control group, as the personalized texts were aligned with their choices from stage 1.

From the two tests of personalized vs. generic texts (stages 3 and 4), we found that students were more likely to select the personalized text when it was aligned with their choices from stage 1. Specifically, those in the experimental group ($n = 12$) selected 1.25 personalized texts on average while that of the control group ($n = 11$) was 0.73 on average. From an ordinal logistic regression, these odds are trending significant at the .05 level ($p = .085$). More details regarding the interpretation of these results can be found in Vaccaro et al. (2024).

References

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.

Felder, R. M., & Silverman, L. K. (1988). Learning and teaching styles in engineering education. *Engineering Education*, 78(7), 674–681.

Vaccaro, M., Friday, M., & Zaghi, A. (2024). Evaluating the capability of large language models to personalize science texts for diverse middle-school-age learners. *ArXiv [Preprint]*. <https://doi.org/10.48550/arXiv.2408.05204>