

Student Use and Evaluation of AI For Clinical or Educational Tasks

Jae Thompson, Kaitlin DeMoss, Lynn Gilbertson PhD, Cody Marie Busch EdD, & Jon Spike
University of Wisconsin – Whitewater

Introduction

As generative artificial intelligence (AI) becomes more ubiquitous and accessible, there is a call to promote awareness, implementation, and evaluation skills among AI users. Free AI platforms offer unprecedented opportunities to transform how education and healthcare tasks are approached. However, a standardized process for evaluating AI output has yet to be reported.

Project Aims

Support students' technology fluency through a structured process that guides the use and evaluation of AI technologies.

Participants

Thirteen graduate students in a speech language pathology master's program. Eleven undergraduate students in a video games and learning course.

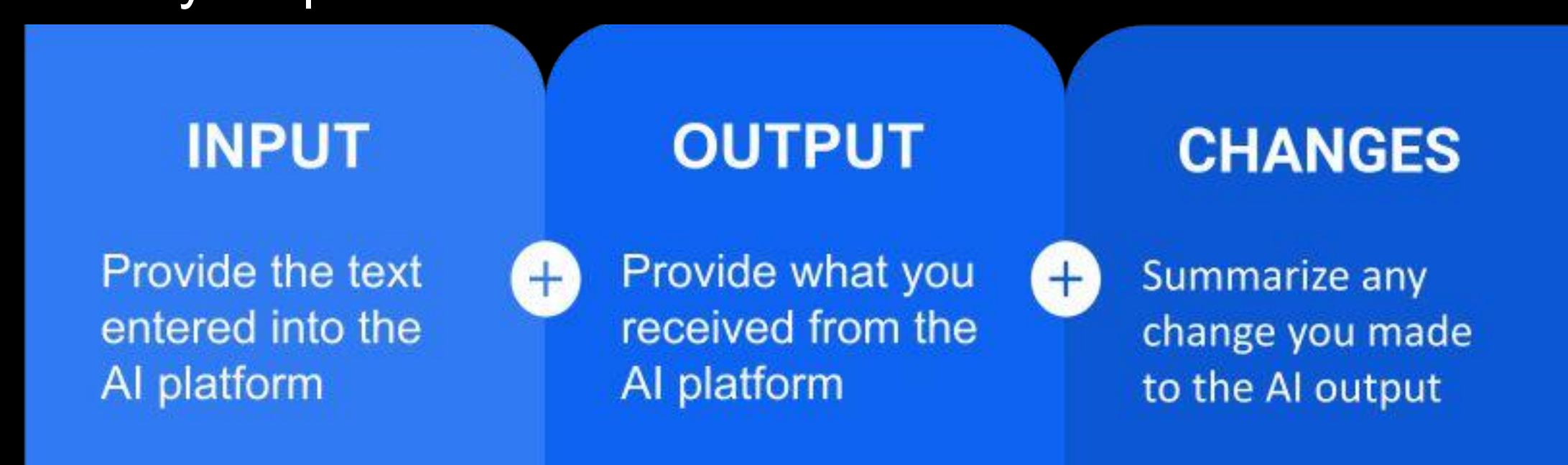
Methods

Students used three different free AI platforms as a part of three different required course assignments.

- ChatGPT to develop patient goals and objectives
- Leonardo.ai to generate a visual scene to support clients in practicing speech perception and production
- CANVA to create a presentation evaluating a video game's application to learning pedagogy

Students evaluated AI in a three step process.

1. Completed a table that included (1) the prompt entered into the AI, (2) the output generated by the AI, and (3) a summary of any output modifications needed.



2. Completed Likert Scale ratings of six questions adapted from the twelve item criteria from Nauta et al. (2023). Ratings for each question fell along a five-point scale (1 = very poor to 5 = very good). The six questions were grouped into three categories: content, presentation, and user.

		1= very poor	2= Poor	3= Acceptable	4= Good	5= Very Good
Category	Question	1	2	3	4	5
Content	Correctness - Accuracy of output					
	Continuity - Similar inputs have similar outputs					
Presentation	Compactness - Output has significant detail balanced with compactness and correctness					
	Composition - Presented with increased clarity - Considers format, organization, and structure					
User	Context - Output is relevant to user's needs and level of expertise					
	Coherence - Consistent with relevant background, knowledge, beliefs, and general consensus					

3. Completed reflections including responses to two prompts:
 - Include a statement of how the AI assignment helped you grow in your planned career path and/or how the artificial intelligence experience could potentially impact your future career goals
 - Evaluate the input/output of artificial intelligence and its usefulness for your intended profession.

Quantitative Results

A non-parametric Kruskal-Wallis rank sum test revealed a significant difference ($p=0.03$) in ratings on the six different Likert Scale questions. A Wilcoxon signed rank test with corrections for multiple testing revealed a significant difference ($p=0.027$) in the overall ratings of Content Continuity (mean = 3.7) with Content Correctness (mean = 3.2), Presentation Compactness (mean = 3.2), and User Context (mean = 3.2). There were no significant differences between ratings of platforms (ChatGPT, CANVA, Leonardo.ai) or categories of Likert Scale questions (content, presentation, user).

Qualitative Results

The content analysis using an inductive approach of the reflection prompt asking students to evaluate the input/output of AI and its usefulness for professional tasks resulted in 74 unique responses across 37 unique assignments. The four most frequently reported themes were that students gained experience with AI (20%), fostered critical thinking skills (19%), facilitated efficiency of work practices (18%), and generated individualized materials (14%).

Reflection Prompt 1: Statement of how the AI assignment helped you grow in your planned career path and/or how the artificial intelligence experience could potentially impact your future career goals

Theme	Frequency	Example Quote
Gained experience with AI	20% n=15	"This assignment has helped me try out new and potentially beneficial technological tools that I would have never tried otherwise."
Fostered critical thinking skills	19% n=14	"It required me to think critically about my patient and the intervention I wish to provide."
Efficiency of work practice	18% n=13	"I can just generate therapy materials quickly with AI rather than spending valuable time making something myself or searching the internet."
Individualized materials	14% n=10	"...can help create pictures that are more individualized for your clients"

The inductive content analysis of the second reflection prompt asking students to evaluate the input/output of AI and its usefulness for their profession resulted in 65 unique responses across 36 unique assignments. One student did not respond to this prompt. The four most frequently reported themes were that AI was a useful tool (29%), concerns about output accuracy existed (25%), input was accurate (17%), and AI was efficient and easy to use (12%).

Reflection Prompt 2: Evaluate the input/output of artificial intelligence and its usefulness for your intended profession.

Theme	Frequency	Example Quote
Useful tool	29% n=19	"By entering just a few key words, a creative image was generated that I can use with a client in treatment."
Concerns about output accuracy	25% n=16	"...there were many changes that I needed to make for the goal to be appropriate, which results in each goal being changed almost completely."
Accurate input	17% n=11	"The information that was AI generated was very accurate and can be easily applicable to my intended profession."
Efficient/Easy to use	12% n=8	"Instead of taking potentially a half hour or longer to create my own image, an image was created for me in less time."



Conclusions

Results suggest the structured process described in the current study was feasible for students to evaluate AI output from a variety of platforms. Students had a framework for evaluating AI that generated critical thinking. While Likert Scale ratings of the three AI platforms were acceptable and the themes indicated benefits of AI for individualization and creation of clinical and educational materials, there were lingering concerns about accuracy. The process seemed to facilitate a shift in time allocation for educational and clinical tasks for prospective speech language pathologists and teacher candidates. Rather than spending time searching for materials, time was reallocated to critiquing and applying expertise. Providing students a structure for this critical evaluation of AI output perhaps further supports efficiency.

Future Work

Future work may further standardize the AI evaluation process to support student and future professional's use of AI tools. The major focus of the current study was evaluating AI output. Future work may seek to evaluate various frameworks for prompt engineering. Combining an evidence based process for creating effective AI input paired with an evidence based process for evaluating AI output has potential to create efficiency and accuracy for AI integration in educational and clinical tasks.

Acknowledgements

This study was funded by the University of Wisconsin-Whitewater Teaching and Learning grant, Artificial Intelligence Category.

Resources



References

