Exploration and Potential of AI Generated Feedback for Promoting Cognitive Presence in Asynchronous Online Discussions

Christine Wusylko, Larisa Olesova, Bojan Lazarevic, & Swapna Kumar

Problem

Students struggle to construct and confirm meaning through reflections and discourse during online asynchronous discussions. This process is called **cognitive presence (CP)** (Garrison et al., 2001)

Can **GPT-4.0 (Microsoft CoPilot)** Large Language Model (LLM) be used to guide students' quality of posts in asynchronous online discussions?

Background

CP can be operationalized by **the Practical Inquiry Model (PIM)** (Garrison et al., 2001; Shea et al., 2010)

Phases	Categories	Descriptions
(1)	Triggering	Become aware of a problem by asking questions
(2)	Exploration	Explore a problem by searching or offering information
(3)	Integration	Integrate interpretations and construction of possible solution
(4)	Resolution	Resolve the problem by critical evaluation of the solution

Giving **personalized feedback** to learners of their level of CP can help them understand what level of CP their posts are and how to improve the quality of the posts.

This is usually done by the content analysis of human qualitative coding based on the CP coding framework by Garrison et al., 2001 and revised by Shea et al. (2010). Castellos-Reyes et al. (2024) adapted the **LLM-Assisted Content Analysis (LACA)** to create an **AI-adapted CP codebook** by Shea et al. (2010) to automatically code CP. This study explores how to apply the AI-adapted CP codebook to analyze students' CP.

Methods

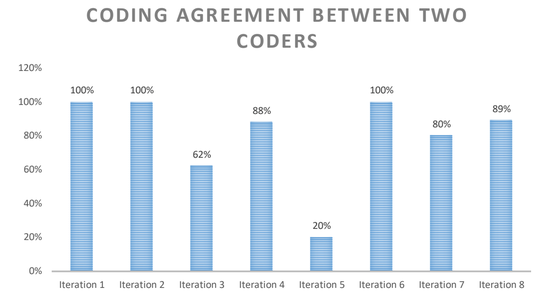
RQ1: How accurate are GPT-4.0 models in classifying students' cognitive presence compared with manual human coding?
RQ2: What other insights do automated content analysis using GPT-4.0 models reveal regarding cognitive presence in asynchronous online discussions?

Data Collection and Analysis

We sampled a random 5% (n=489) of posts from (1) an undergraduate course *Introduction to Instructional Design (n=216)* and (2) a graduate course *Instructional Design (n=273)*

- The first two authors coded the posts using and adapting an **AI-adapted CP codebook** (Castellos-Reyes et al., 2024) in 8 iterations
- We then used **LACA** to code the same posts using *a one-shot (with examples)* and *two-shot approach (without examples*)

Results



Generally, high agreement between the two human coders, which was resolved before comparing with GPT-4.0

Undergraduates and Graduates	Cohen k	p
CoPilot One Shot (Without Examples)	κ=.223	p<0.163
CoPilot Few Shot (With Examples)	κ=.281	p<0.093

Few-shot was more similar to human coders than one-shot

Undergraduates	Cohen k	p
CoPilot One Shot (Without Examples)	κ=.123	p<0.163
CoPilot Few Shot (With Examples)	κ=.120	p<0.093

Undergraduate posts had **poor** strength in agreement with human coders

Graduates	Cohen k	p
CoPilot One Shot (Without Examples)	κ=.350	p<0.001
CoPilot Few Shot (With Examples)	κ=.547	p<0.001

Graduate posts had **fair** and **moderate** strength of agreement with human coders

Discussion and Conclusion

- GPT-4.0 can guide *graduate* students' quality of posts (k=.547) better than *undergraduate* (k=.120).
- Students can use the **few-shot prompt** as it was more accurate and can be applied in online discussions. Our results **differ** from Castellanos-Reyes et al (2024).
- High agreement between the human coders could be because one of the coders is a native English speaker. This may have implications for coding English text (Castellanos-Reyes et al., 2024)