Real-Time Scaffolding of Critical Thinking: AI Agents for Cognitive Presence Detection in Discussions

Fengjiao Tu
University of North Texas
FengjiaoTu@my.unt.edu

Haihua Chen
University of North Texas
Haihua.Chen@unt.edu

Ji Hyun Yu
University of North Texas
Jihyun.Yu@unt.edu

ABSTRACT

This study designs, implements, and evaluates an AI-based discussion facilitation agent that promotes critical thinking in graduate courses by detecting cognitive presence (CP) in real time. Grounded in the Community of Inquiry (CoI) framework, the agent embeds a fine-tuned RoBERTa transformer model (trained on 17,627 human-coded MOOC sentences; macro-F1 ≈ 0.70) into live graduate-level online discussions. The system classifies each student post into one of four CP phases—Triggering, Exploration, Integration, or Resolution—and immediately generates theory-informed, formative prompts to scaffold learners toward deeper inquiry. A design-based research methodology is used to iteratively refine the model and evaluate its pedagogical impact in instructional design courses. Three research questions address model accuracy, student and instructor perceptions of utility, and technical and pedagogical feasibility. Expected contributions include an interpretable NLP approach to CP detection, a replicable blueprint for real-time AI scaffolding, and empirical evidence that automated nudges can elevate online discourse beyond the Exploration plateau commonly observed in CoI studies.

Keywords

Cognitive presence, Community of Inquiry, AI agent, natural language processing, real-time scaffolding, online discussion

1. INTRODUCTION

Cognitive presence (CP) is defined as the extent to which learners construct and confirm meaning through sustained reflection and discourse, and is a central component of the Community of Inquiry (CoI) framework [1]. CP develops through four phases of practical inquiry: Triggering Event, Exploration, Integration, and Resolution. Together, these phases describe how critical thinking can emerge in online and hybrid learning environments.

Across two decades of CoI research, a consistent pattern has been reported: most student posts remain at the Exploration phase, and few reach Integration or Resolution [3, 4]. The pedagogical implication is that online discussions often underdeliver on their critical-thinking promise. Prior work suggests, however, that this pattern is responsive to instructor facilitation. When instructors actively guide students to synthesize and resolve, learners do advance to higher CP phases [2]. The difficulty is that such facilitation is hard to sustain in courses with many discussions and many students. Instructors have limited time, and most posts do not receive immediate, individualized guidance in the moment of writing.

In parallel, recent advances in natural language processing (NLP) have made automated CP classification feasible. Transformer-based models can label CP phases with accuracy that approaches human coders [3, 5]. Our prior work fine-tuned RoBERTa on 17,627 human-coded MOOC sentences and reached macro-F1 ≈ 0.70. These classifiers, however, have been used mainly for retrospective analysis of completed courses rather than for supporting learners as they write. The pedagogical potential of CP detection—using it to actually influence discourse in the moment—remains largely untested.

This study addresses that gap. It embeds a validated CP classifier in a live AI discussion agent that monitors graduate-level discussion forums, classifies each new post in real time, and returns immediate, gentle, theory-grounded prompts intended to encourage deeper inquiry. The study examines whether such real-time scaffolding can move student posts toward Integration and Resolution, and under what technical and pedagogical conditions this is feasible.

This study is guided by three research questions:

- **RQ1 (Accuracy):** How reliably can the AI agent classify discussion posts into CP stages compared to human coders? We will benchmark the model against expert-coded data using F1 and Cohen's κ.
- **RQ2 (Utility):** How do students and instructors perceive the agent's real-time prompts? Do they find the feedback clear, timely, and helpful for deepening discussion? Surveys and interviews will gauge acceptability and usefulness.
- **RQ3 (Feasibility):** What technical and pedagogical challenges arise in deploying this real-time agent? We

Fengjiao Tu, Ji Hyun Yu, and Haihua Chen. Real-Time Scaffolding of Critical Thinking: AI Agents for Cognitive Presence Detection in Discussions. In Anthony Botelho, Maria Mercedes T. Rodrigo, Adish Singla, Hiroaki Ogata, Hyojeong So, and Young Hoan Cho (eds.) Proceedings of the 19th International Conference on Educational Data Mining, Seoul, Republic of Korea, June, 2026, pp. 881–883. International Educational Data Mining Society (2026).

https://doi.org/10.5281/zenodo.21040044

will examine integration with learning platforms, system responsiveness, instructor workload, and any unintended effects on discourse dynamics.

This project makes three interrelated contributions to the EDM and AIED communities. Theoretically, it extends the CoI framework from a retrospective coding scheme into an automated real-time in- tervention, demonstrating how a well-established learning theory can be operationalized in an AI system. Methodologically, it demonstrates an explainable transformer NLP approach—aug- mented with SHAP analyses and CoI-theoretic features—that makes model decisions transparent and trustworthy for educational stakeholders. Instructionally, it offers a concrete blueprint for deploying real-time AI scaffolding in online and hybrid graduate courses, with documented lessons on integration, usability, and pedagogical impact. By anticipating that real-time CP nudges will increase the propor- tion of Integration/Resolution posts and that students will perceive the agent positively, this research also builds empirical evidence for the promise of AI-driven formative feedback in higher education. Findings will be disseminated to both technical (EDM/AIED) and educational (Learning Sciences, instructional design) communities, reinforcing the interdisciplinary spirit of the EDM conference

2. RELATED WORK

This study draws on two lines of prior research: empirical CoI work on cognitive presence in online discussions, and NLP work on automated CP detection. We review each below before identifying the gap that motivates this study.

2.1 The Exploration Plateau in CoI Research

The CoI framework conceptualizes meaningful online learning as the interaction of social, teaching, and cognitive presence [1]. Empirical CP research consistently reports that discussion contributions concentrate at the Exploration phase, with relatively few posts reaching Integration or Resolution [4]. Facilitation is identified as the primary driver of advancement: when instructors explicitly guide students to resolve problems, participants do progress to higher CP phases [2]. Two implications follow. First, the Exploration plateau is not absolute; it responds to deliberate scaffolding. Second, scaffolding of this kind has so far been documented mainly in small, instructor-intensive settings, where post-by-post facilitation is feasible.

2.2 Automated Detection of Cognitive Presence

A parallel line of work uses NLP to automate CP classification. Transformer architectures, such as BERT and RoBERTa, have achieved strong performance on educational text classification tasks. [3] introduced MOOC-BERT for automated phase labeling. [5] extended this line with annotation-guideline-based knowledge augmentation for educational text classification. [2] report that GPT-class large language models can approximate human coders on CP content analysis. Our prior work fine-tuned RoBERTa with embedded CoI features and reached macro-F1 ≈ 0.70; the same analysis also revealed a correlation between higher-order CP posts and downstream outcomes including grades and self-reported knowledge transfer.

2.3 The Gap This study Addresses

Although both lines of work are well developed, they have rarely been combined. The CoI facilitation literature shows that scaffolding can move learners beyond Exploration but does not scale to the post level. The NLP literature shows that machines can detect CP reliably but has used detection almost entirely as a post-hoc analytic tool. To our knowledge, no prior study has embedded a CP classifier in a live discussion environment to provide real-time, theory-grounded prompts to students as they write. This study is positioned at exactly this intersection: it asks whether automated, in-the-moment CP scaffolding can support the kind of advancement toward Integration and Resolution that human facilitation can produce but cannot sustain at scale.

3. METHODOLOGY

This project employs a design-based research (DBR) approach, iteratively developing the AI agent and testing it in authentic classroom settings, then refining the model and prompts based on real classroom feedback. The work comprises two interlocking components.

3.1 Model Refinement

We begin with an existing transformer classifier (RoBERTa fine-tuned on MOOC data with embedded CoI features). To adapt it to graduate seminar discussions, we apply transfer learning: initializing from MOOC-trained weights, then retraining on a smaller set of manually coded graduate-level posts. Approximately 20% of collected posts will be double-coded by experts to create a gold-standard test set. Model performance will be evaluated using precision, recall, and F1 for each CP category, along with inter-rater agreement (Cohen's κ) between model and human coders. Shapley Additive Explanations (SHAP) analyses [6] will be used to identify the linguistic features driving classification decisions, rendering the model interpretable to both technical and educational audiences.

3.2 AI Agent Development

The refined classifier is embedded into an interactive agent within the learning management system (LMS) or chat platform. During weekly forum activities, the agent listens for new posts. When a student submits a contribution, the agent immediately classifies it into a CP phase and generates a tailored, formative prompt. These prompts are co-designed with instructors and grounded in CoI theory. For example, if a post is rated as Exploration, the agent might respond: *"You have shared some great ideas. How might these connect to other concepts we have covered?"* If a post is already at the Integration level, it might offer: *"Nice synthesis of ideas. Can you take this further by applying it to a real-world problem?"* Crucially, prompts are phrased as open questions, not evaluations, and students are free to revise their post or ignore the suggestion with no grading penalty. Instructors have oversight through a dashboard that logs each classification and can disable or customize feedback as needed.

3.3 Evaluation Design

The AI agent will be deployed in several graduate instructional design courses over a semester (approximately six discussions per course, targeting roughly 300 posts in total).

Data collection from Fall 2025 (without the CP-supported agent) will serve as a baseline comparison. The agent-supported implementation is planned for Fall 2026. Both quantitative and qualitative data will be collected: system logs will capture processing time, uptime, and integration issues (RQ3); post-semester surveys will assess student perceptions of clarity and helpfulness (RQ2); and semi-structured interviews will probe student and instructor experiences in depth (RQ2–RQ3). Model accuracy metrics will be computed against expert-coded samples (RQ1).

4. REFERENCES

[1] D. Randy Garrison, Terry Anderson, and Walter Archer. 2001. Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education* 15, 1 (2001), 7–23.

[2] Daniela Castellanos-Reyes, Larisa Olesova, and Ayesha Sadaf. 2025. Transforming online learning research: Leveraging GPT large language models for automated content analysis of cognitive presence. *The Internet and Higher Education* 65 (2025), 101001.

[3] Zhi Liu, Xi Kong, Hao Chen, Sannyuya Liu, and Zongkai Yang. 2023. MOOC-BERT: Automatically identifying learner cognitive presence from MOOC discussion data. *IEEE Transactions on Learning Technologies* 16, 4 (2023), 528–542.

[4] Yuanyuan Hu, Claire Donald, and Nasser Giacaman. 2023. Can multi-label classifiers help identify subjectivity? A deep learning approach to classifying cognitive presence in MOOCs. *International Journal of Artificial Intelligence in Education* 33, 4 (2023), 781–816.

[5] Shiqi Liu, Sannyuya Liu, Lele Sha, Zijie Zeng, Dragan Gašević, and Zhi Liu. 2025. Annotation guideline-based knowledge augmentation: Towards enhancing large language models for educational text classification. *IEEE Transactions on Learning Technologies* (2025).

[6] Scott M. Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* 30 (2017), 4765–4774.