# Asynchronous Discussion Forums Show More Analytical but Less Diverse Student Engagement in the Age of AI

Zhen Xu*
Columbia University
New York, NY, USA
zx2393@tc.columbia.edu

Yijun Dai*
Columbia University
New York, NY, USA
yg2854@tc.columbia.edu

Chenxi Shi
Columbia University
New York, NY, USA
cs4477@tc.columbia.edu

Siyan Li
Columbia University
New York, NY, USA
siyan.li@columbia.edu

Chengyuan Yao
Columbia University
New York, NY, USA
cy2706@tc.columbia.edu

Renzhe Yu
Columbia University
New York, NY, USA
renzheyu@tc.columbia.edu

## Abstract

Asynchronous discussion forums have been one of the most widely used instructional tools in postsecondary education, enabling multiple pedagogical functions such as question clarification, collaborative learning, and social interaction. The rapid diffusion of generative AI (GenAI) tools, however, raises new questions about whether these forums continue to support meaningful student engagement in the same ways. In this work in progress, we analyze 371,383 discussion posts from 2021 to 2024 at a public US university and identify mixed but concerning shifts in students' cognitive and social engagement across different types of discussion tasks. In academic and social discussions, posts become longer and more analytical, but also more similar in content; peer interaction increases overall but becomes more concentrated among fewer students. Across cognitive complexity levels, tasks with lower cognitive demand show declines in cognitive engagement, whereas creativity-focused tasks exhibit greater content similarity and reduced peer interaction. Together, our findings provide preliminary evidence that traditional discussion forums may be becoming more polished but less pedagogically productive, highlighting the need to redesign discussion activities to better support meaningful learning in the age of AI.

## CCS Concepts

• **Applied computing** → **Interactive learning environments**; E-learning; • **Computing methodologies** → *Information extraction.*

## Keywords

Asynchronous Discussion Forums, Student Engagement, Postsecondary Education, Generative AI, Instructional Design

**ACM Reference Format:**
Zhen Xu, Yijun Dai, Chenxi Shi, Siyan Li, Chengyuan Yao, and Renzhe Yu. 2026. Asynchronous Discussion Forums Show More Analytical but Less Diverse Student Engagement in the Age of AI. In *Proceedings of the Thirteenth ACM Conference on Learning @ Scale (L@S '26), June 29–July 3, 2026, Seoul, Republic of Korea.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3774398.3811556

*Both authors contributed equally to this research.


ACM ISBN 979-8-4007-2293-6/2026/06
https://doi.org/10.1145/3774398.3811556

## 1 Introduction

Asynchronous discussion forums are a key feature of modern learning management systems which are commonly adopted by higher education institutions [4]. Their capacity to support multiple instructional functions, including question clarification, collaborative learning, and social interaction, has made them a common pedagogical activity across disciplines and course formats [3, 5]. The emergence of generative artificial intelligence (GenAI), however, raises concerns about student engagement in this longstanding instructional environment. The effectiveness of discussion-based learning depends on students' cognitive and social engagement, which are essential for promoting metacognition, idea development, and collaborative knowledge construction. GenAI tools' capacity to produce instant responses may reduce the cognitive effort these activities are designed to cultivate, while easy access to AI-generated answers may diminish students' motivation to interact with peers or sustain extended dialogues.

Research on large public knowledge-sharing platforms suggests that GenAI has reshaped user behavior around information seeking, content creation, and social interaction. For example, there have been declines in forum participation, particularly among newer users, as they increasingly turn to AI for immediate answers rather than engaging in community-based knowledge exchanges [1, 2]. If similar patterns exist in academic discussion forums, they may become repositories of AI-assisted responses rather than spaces for collaborative knowledge construction, undermining their pedagogical value.

We present one of the largest empirical studies of student engagement in asynchronous discussion forums in the age of AI. Drawing on 371,383 discussion posts from a public U.S. university between 2021 and 2024, we depict temporal dynamics of students' cognitive and social engagement. We further examine how these changes vary across discussion task characteristics. These analyses offers timely empirical evidence on how routine forms of learning activities evolve amid the diffusion of GenAI, contributing to broader debates about AI's implications for authentic teaching and learning and informing future-ready instructional design.

**Table 1: Measures of cognitive engagement**

| Measure | Description |
|---|---|
| Content similarity | Semantic similarity of posts within a discussion, measured using two embedding-based metrics. (1) *Embedding variance* is computed as the average variance across embedding dimensions, and (2) *Semantic spread* is derived from the eigenvalue spectrum of the cosine similarity matrix. Each post is encoded with all-MiniLM-L6-v2 (384 dimensions). |
| Average content length | Average number of words per post within a discussion topic. |
| Average content depth | Analytical thinking score computed using LIWC [6], capturing higher-order cognitive process such as logical reasoning and structured analysis in the post content. |
| Average content width | Informativeness, defined as the amount of new information introduced relative to prior posts. Computed as one minus the proportion of word-level tokens overlapping with earlier posts within the same discussion topic [7]. |

**Table 2: Measures of social engagement**

| Measure | Description |
|---|---|
| Received reply rate | Proportion of participating students who receive at least one reply. |
| Average thread depth | Mean number of follow-up replies along reply chains across all posts within a discussion topic. |
| Maximum thread depth | Maximum number of follow-up replies observed in any single reply chain within a discussion topic. |
| Average thread breadth | Mean number of distinct participants contributing replies to each post within a discussion topic. |
| Maximum thread breadth | Maximum number of distinct participants contributing replies to any single post within the discussion topic. |

# 2 Methodology

## 2.1 Data and Sample

Through an existing collaboration, we analyze anonymized discussion forum records from Canvas learning management system at a four-year public university in the United States. These records include the content and response structures of individual posts. The dataset spans a three-year period from Fall 2021 through Spring 2024. We use the public release of ChatGPT at the end of Fall 2022 term as the time marker to distinguish pre-AI and post-AI periods. Because students create discussion posts in the context of discussion tasks (topics) specified by their instructor(s), we restrict our analyses to discussion tasks that were identical across the two periods for fair comparison. This sample accounts for 18.3% of the original dataset and includes 371,383 student posts across 7,667 discussion tasks.

## 2.2 Measuring Student Engagement and Discussion Task Characteristics

Building upon prior research, we measure students' cognitive and social engagement by quantifying features of their written posts and peer interactions at the discussion topic level. Table 1 and Table 2 summarize all the measures used in this study.

We further derive a series of discussion task characteristics from an instructional design perspective. First, we classify each discussion task by its instructional purpose: academic, logistic and/or social. Since some tasks may have multiple purposes at the same time, we use allow each task to have more than one purpose label in our annotation. Second, for all academic discussion tasks, we use a grounded theory approach to further identify four types of discussion questions: *Open Q&A*, which invites clarification questions; *Integration & Application*, which asks students to synthesize, apply, evaluate, or reflect; *Collaborative Learning*, which encourages peer interaction and collaboration; and *Personal Experience*, which draws on students' own experiences and contexts. Finally, we classify each academic discussion task by their cognitive complexity using Bloom's taxonomy. Because of small sample sizes, the level of *Remember* is merged into *Understand*, which leaves five levels in our analysis: *Understand*, *Apply*, *Analyze*, *Evaluate*, and *Create*. For all three dimensions above, we use a human–AI collaborative annotation pipeline validated against expert-labeled ground truth (all Cohen's $\kappa > 0.8$).

## 2.3 Estimating Changes in Student Engagement

We estimate changes in student cognitive and social engagement using fixed effects regression models with cluster-robust standard errors. We standardize all engagement metrics using z-scores for comparability across metrics and consistent interpretation of effect sizes. For each metric, separate regressions are estimated with *PostAI* as the key explanatory variable. The baseline model for overall engagement changes is:

$$Y_{it} = \beta_0 + \beta_1 PostAI_t + \beta_2 ClassSize_{it} + \beta_3 PostReplyRatio_{it} + \lambda_i + \epsilon_{it} \quad (1)$$

where $Y_{it}$ represents a standardized engagement metric for discussion task $i$ implemented in term $t$, $PostAI_t$ is a binary indicator for the post-GenAI period, and control variables include the number of students participating in the discussion ($ClassSize_{it}$), the ratio of original posts to replies in the discussion thread ($PostReplyRatio_{it}$), and task fixed effects ($\lambda_i$).

To depict engagement dynamics under different instructional conditions, we apply this baseline model to subsamples of discussion tasks defined by different instructional purposes, discussion question types, and cognitive complexity levels.

## 3 Results

Table 3 presents estimated changes in student engagement across discussion tasks with different instructional purposes. In academic-purpose discussions, posts become more similar in content, longer in length, and cognitively deeper, while their level of informativeness remains largely unchanged. Specifically, semantic spread declines significantly in *Academic Only* discussions (-0.068 SD, $p < .05$) and *Academic + Social* discussions (-0.077 SD, $p < .01$). At the same time, posts become more elaborated: average word count increases in *Academic Only* discussions (0.041 SD, $p < .05$). Analytical thinking also rises across academic discussion types, with the largest effect observed in *Academic + Logistics* (0.225 SD, $p < .001$). In contrast, informativeness does not vary significantly across discussion categories. In terms of social interaction, peer response rates increase in academically oriented discussions, including *Academic Only* (0.063 SD, $p < .001$) and *Academic + Social* discussions (0.05 SD, $p < .05$). However, this rise in activity is accompanied by a decline in maximum interaction width in both *Academic Only* (-0.055 SD, $p < .001$) and *Academic + Social* discussions (-0.179 SD, $p < .01$), suggesting that the increase in replies may be concentrated among a smaller subset of highly active students.

Table 4 presents estimated changes in student engagement across different discussion question types. Open Q&A discussions show minimal changes, except for a decline in maximum discussion thread width (-0.21 SD, $p < .01$), with no significant content changes. In contrast, Integration and application discussions become more similar in content (semantic spread -0.059 SD, $p < .05$), longer (0.054 SD, $p < .05$), and more analytical (0.204 SD, $p < .001$), alongside higher reply rates (0.044 SD, $p < .001$). Collaborative discussions follow a similar pattern, with reduced semantic spread (-0.06 SD, $p < .05$), increased analytical thinking (0.236 SD, $p < .001$), and greater peer interaction (0.077 SD, $p < .001$). Personal experience discussions also show gains in analytical thinking level (0.192 SD, $p < .001$) and reply rates (0.064 SD, $p < .001$), but without increased content similarity.

Table 5 presents estimated changes in student engagement across different levels of cognitive complexity. For lower-cognitive complexity discussion tasks (*Understand*, *Apply*, *Analyze*), students' posts become shorter, less analytical, and less interactive. Higher-level tasks show more complex patterns. *Evaluate* discussions become longer (0.079 SD, $p < .01$), more analytical (0.244 SD, $p < .001$), and more interactive (e.g., *Reply rate*: 0.092 SD, $p < .001$). In contrast, *Create* tasks show greater content similarity (embedding variance -0.08 SD, $p < .05$), lower informativeness (-0.168 SD, $p < .05$), and reduced peer interaction (e.g., *Reply rate*: -0.075 SD, $p < .05$).

## 4 Discussion and Future Work

We present a large-scale, longitudinal analysis of asynchronous discussion forums, a commonly deployed instructional tool in higher education. By examining posting content and response patterns under identical discussion tasks before and after the diffusion of GenAI, we identify several concerning changes in student cognitive and social engagement. While posts were longer and more analytical in academic discussion tasks in the post-GenAI era, the content was also more similar on average. In academic and social discussions, peer response rates increased, but interaction became more concentrated among a smaller group of students compared to pre-GenAI times. Under discussion tasks with lower cognitive demand, student engagement declined, and creativity-focused tasks showed more similar responses and less peer interaction. In contrast, tasks that involve personal reflection or evaluation appeared more resistant to the potential influences of GenAI diffusion.

Our findings have important implications. First, they deepen our understanding of how students are adapting to generative AI within a widely used instructional activity. Even when discussion forums may include specific participation requirements, we observe shifts similar to those found on public knowledge-sharing platforms, including reduced peer interaction and greater content similarity across responses. These patterns suggest that generative AI may be changing not only how students complete academic tasks, but maybe also their learning preference. Rather than engaging in peer interaction-based knowledge building, students may increasingly rely on more personalized, AI-augmented learning processes. This shift highlights the growing need to reconsider how traditional instructional activities support meaningful learning in the age of AI. Second, our results provide empirical guidance for instructional reform. AI does not affect all discussion activities in the same way. Some discussion tasks show clear declines in meaningful engagement, while others remain more stable. By identifying where engagement declines and where it remains robust, our study highlights the critical role of task design and supports evidence-based instructional improvement.

In future work, we plan to extend our analysis in several directions. First, to strengthen causal interpretation, we will more explicitly account for underlying temporal trends in students' discussion forum participation before and after the public release of ChatGPT. Future analyses may adopt stronger quasi-experimental approaches to better distinguish potential generative AI effects from broader shifts in student engagement over time. Second, we plan to complement our quantitative findings with qualitative analyses to further investigate how the content, structure, and nature of student responses are changing. Finally, we will conduct cross-institutional analyses to examine how these adaptation patterns vary across different student populations, instructional contexts, and institutional settings. Together, these efforts will strengthen our understanding on how students adapt to generative AI and provide a stronger empirical foundation for evidence-based instructional design in the age of AI.

## References


[1] Gordon Burtch, Dokyun Lee, and Zhichen Chen. 2024. The consequences of generative AI for online knowledge communities. *Scientific Reports* 14, 1 (2024), 10413. doi:10.1038/s41598-024-61221-0

[2] R Maria del Rio-Chanona, Nadzeya Laurentsyeva, and Johannes Wachs. 2024. Large language models reduce public knowledge sharing on online Q&A platforms. *PNAS Nexus* 3, 9 (2024), pgae400. doi:10.1093/pnasnexus/pgae400

[3] Khe Foon Hew and Wing Sum Cheung. 2012. *Student participation in online discussions: Challenges, solutions, and future research.* Springer Science & Business Media.

[4] Phil Hill. 2025. State of Higher Ed LMS Market for US and Canada: Year-End 2024 Edition. https://onedtech.philhillaa.com/p/state-of-higher-ed-lms-market-for-us-and-canada-year-end-2024-edition

[5] Florence Martin, Ting Sun, and Carl D Westine. 2020. A systematic review of research on online teaching and learning from 2009 to 2018. *Computers & Education* 159 (2020), 104009.

[6] James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The Development and Psychometric Properties of LIWC2015.* Technical Report.

**Table 3: Estimated changes in students' cognitive and social engagement by instructional purpose**

| | Academic Only | Academic + Logistics | Academic + Social | All Three | Logistics Only | Social Only |
|---|---|---|---|---|---|---|
| Embedding variance | 0.013 | -0.006 | -0.077** | 0.047 | 0.103. | -0.107** |
| Semantic spread | -0.068* | -0.056 | 0.081 | 0.255 | 0.128 | 0.143 |
| Word count | 0.041* | 0.043 | -0.047 | -0.013 | 0.258. | -0.012 |
| Analytical | 0.17*** | 0.225*** | 0.103** | 0.048 | 0.46. | 0.103 |
| Informativeness | -0.025 | 0.056 | 0.086 | -0.074 | -0.001 | -0.196 |
| Reply rate | 0.063*** | 0.004 | 0.05* | 0.039 | 0.001 | 0.098. |
| Avg thread depth | 0.007. | -0.017 | 0.007 | -0.085. | 0.018 | 0.023 |
| Max thread depth | 0.034 | 0.014 | -0.031 | -0.174. | 0.207*** | 0.078 |
| Avg thread width | 0.005 | -0.015 | 0.002 | -0.073 | 0.011 | 0.047 |
| Max thread width | -0.055*** | 0.045 | -0.179** | 0.025 | 0.025 | -0.159 |

*Notes: Cell values are average estimated changes (standard deviation) in each engagement measure under each instructional purpose. Significance levels:* $p < 0.001(^{***})$, $p < 0.01(^{**})$, $p < 0.05(^{*})$, $p < 0.10(.)$.

**Table 4: Estimated changes in students' cognitive and social engagement by discussion question type**

| | Integration & Application | Open Q&A | Personal Experience | Collaborative Learning |
|---|---|---|---|---|
| Embedding variance | 0.029 | -0.08 | 0.003 | 0.018 |
| Semantic spread | -0.059* | -0.076 | -0.009 | -0.06* |
| Word count | 0.054* | 0.032 | 0.021 | 0.042. |
| Analytical | 0.204*** | -0.045 | 0.192*** | 0.236*** |
| Informativeness | 0.023 | -0.007 | 0.002 | 0.044 |
| Reply rate | 0.044*** | 0.007 | 0.064*** | 0.07*** |
| Avg thread depth | 0.006 | 0.013 | 0.005 | -0.001 |
| Max thread depth | 0.056** | 0.079 | 0.007 | 0.011 |
| Avg thread width | 0.003 | -0.044 | 0.003 | -0.003 |
| Max thread width | 0.003 | -0.21** | -0.036. | -0.063** |

*Notes: Cell values are average estimated changes (standard deviation) in each engagement measure under each question type of academic discussion tasks. Significance levels:* $p < 0.001(^{***})$, $p < 0.01(^{**})$, $p < 0.05(^{*})$, $p < 0.10(.)$.

**Table 5: Estimated changes in students' cognitive and social engagement by levels of cognitive complexity**

| | Understand | Apply | Analyze | Evaluate | Create |
|---|---|---|---|---|---|
| Embedding variance | -0.048 | -0.043 | -0.031 | 0.03 | -0.08* |
| Semantic spread | -0.039 | 0.234 | -0.053 | -0.039 | -0.012 |
| Word count | -0.12*** | -0.024 | -0.147*** | 0.079** | -0.054 |
| Analytical | -0.179*** | -0.205** | -0.185*** | 0.244*** | -0.107. |
| Informativeness | -0.021 | -0.241** | -0.14 | 0.061 | -0.168* |
| Reply rate | -0.064** | -0.121. | -0.089*** | 0.092*** | -0.075* |
| Avg thread depth | 0.002 | -0.025 | -0.012 | 0.008 | -0.015 |
| Max thread depth | -0.017 | -0.026 | -0.111* | 0.056* | -0.061 |
| Avg thread width | 0.001 | -0.022 | -0.009 | 0.007. | -0.022* |
| Max thread width | 0.006 | 0.081 | 0.003 | -0.04. | -0.058 |

*Notes: Cell values are average estimated changes (standard deviation) in each engagement measure under each cognitive complexity level. Significance levels:* $p < 0.001(^{***})$, $p < 0.01(^{**})$, $p < 0.05(^{*})$, $p < 0.10(.)$.

University of Texas at Austin, Austin, TX. https://repositories.lib.utexas.edu/items/705e81ca-940d-4c46-94ec-a52ffdc3b51f

[7] Junling Wang, Jakub Macina, Nico Daheim, Sankalan Pal Chowdhury, and Mrinmaya Sachan. 2024. Book2Dial: Generating Teacher Student Interactions from Textbooks for Cost-Effective Development of Educational Chatbots. In *Findings of the Association for Computational Linguistics: ACL 2024*. Association for Computational Linguistics, Bangkok, Thailand, 9707–9731. doi:10.18653/v1/2024.findings-acl.578