Integrating Generative AI Chatbots to Foster the Entrepreneurial Mindset in Engineering Education

Xing Gao, University of Illinois at Urbana - Champaign

Xing Gao is a Master student in Computer Science at the University of Illinois at Urbana-Champaign. She holds a Bachelor of Science in Computer Science from University of Delaware. Her research focuses on knowledge graph optimization with LLMs for personalized learning.

Abdulrahman AlRabah, University of Illinois at Urbana - Champaign

Abdulrahman AlRabah is a Ph.D. student in the Siebel School of Computing and Data Science at the University of Illinois Urbana Champaign, where he also earned his M.S. in Computer Science. He received his B.S. in Mechanical Engineering from California State University Northridge and has industry experience including oil and gas and co founding a food and beverage company. His research focuses on AI for computing education, using customized large language models and learning analytics, including knowledge graph based concept modeling, to deliver better feedback, improve student learning, and support instructors.

Dr. Sotiria Koloutsou-Vakakis, University of Illinois at Urbana - Champaign

Dr. Sotiria Koloutsou-Vakakis holds a Diploma in Surveying Engineering (National Technical University of Athens, Greece), a M.A. in Geography (University of California, Los Angeles), and M.S. and Ph.D. degrees in Environmental Engineering (University of Illinois Urbana-Champaign). Her interests include air quality, environmental policy, and supporting student learning and professional development.

Prof. Tomasz Kozlowski, University of Illinois at Urbana - Champaign

Dr. Tomasz Kozlowski is a Professor and Associate Head for the Undergraduate Programs in the Department of Nuclear, Plasma, and Radiological Engineering at the University of Illinois Urbana-Champaign. He received a Ph.D. in Nuclear Engineering from Purdue University in 2005, where he worked on spatial homogenization methods for neutron transport calculations, multi-physics coupling, and PARCS code development. Dr. Kozlowski has been an instructor in professional workforce development courses and summer schools, including MeV, FJOH, SUNCOP, and NRS-HOT (since 2006). He has been an active member of the American Nuclear Society (since 1997) and the American Society for Engineering Education (since 2016). He is a member of the OECD/NEA Nuclear Science Committee Working Party on Reactor Systems (NSC/WPRS) (since 2010) and is serving on the Scientific Board and Technical Committee of the OECD/NEA Benchmark for Uncertainty Analysis in Best-Estimate Modelling for Design, Operation and Safety Analysis of Light Water Reactors (LWR-UAM) (since 2008). He is an Associate Editor of Nuclear Technology (since 2021) and has been a reviewer for several journals in the fields of nuclear engineering, fluid flow and heat transfer, and uncertainty quantification (since 2006).

Dr. Volodymyr Kindratenko, University of Illinois at Urbana - Champaign

Volodymyr Kindratenko is the Director for the Center for Artificial Intelligence Innovation (CAII) at the National Center for Supercomputing Applications (NCSA). He is also an Adjunct Associate Professor in the Departments of Electrical and Computer Engineering and a Research Associate Professor in the Siebel School of Computing and Data Science. His research interests lie at the intersection of AI systems and applications, with a focus on generative AI solutions for science and education. His team is developing the uiuc.chat chatbot platform, which facilitates rapid knowledge distillation.

Dr. Abdussalam Alawini, University of Illinois Urbana-Champaign

Dr. Alawini is a Teaching Associate Professor in the School of Computing and Data Science at the University of Illinois at Urbana-Champaign. He holds a bachelor's degree from the University of Tripoli and master's and doctoral degrees in Computer Science from Portland State University. His research focuses on data management systems and computing education. Dr. Alawini is passionate about building data-driven, AI-based systems for improving teaching and learning.

Integrating Generative AI Chatbots to Foster the Entrepreneurial Mindset in Engineering Education

Abstract

Generative AI offers a scalable way to support the Entrepreneurial Mindset (EM: Curiosity, Connections, and Creating Value) in engineering education, yet many classroom deployments rely on generic chat interfaces that deliver technical answers without prompting EM related reflection. This study evaluates a custom Retrieval Augmented Generation (RAG) chatbot deployed across four engineering courses at a public research university. A distinguishing feature is an optional "EM Info Box" that appends a short, structured metacognitive prompt to selected technical responses. We analyzed n = 3,187 interaction turns from a 16 week semester using log based observational methods. Generalized Estimating Equations modeled the association between EM aligned responses and immediate follow up behavior while accounting for within-student clustering, complemented by a longitudinal analysis of prompt length. Because the EM Info Box is activated by a content based classifier rather than randomized assignment, all reported effects are associative.

Findings indicate that EM aligned scaffolding varies by disciplinary context: in one course, EM aligned turns were associated with lower follow up probability, suggesting single turn informational resolution. In another, "Creating Value" triggers reframed technical metrics as expressions of stakeholder value. Longitudinal patterns revealed modest within-student growth in prompt length with a spike near a major deadline. These findings provide behavioral trace evidence of when EM aligned scaffolding appears and how it co-occurs with interaction behaviors, without directly measuring changes in students' underlying mindset.

1 Introduction

The rapid emergence of Generative Artificial Intelligence (GenAI) has introduced both opportunity and tension in engineering education [1, 2, 3]. Large Language Models (LLMs) such as ChatGPT can support students with code generation, conceptual explanations, and rapid problem solving. At the same time, concerns persist that generic chat interfaces may reduce learning to the passive consumption of automated answers, potentially displacing reflective practices central to professional formation [4, 5]. For engineering educators, the challenge is not merely whether students use AI, but how AI systems can be intentionally designed to support habits of professional reasoning rather than bypass them.

One widely adopted framework for professional formation in engineering is the Entrepreneurial

Mindset (EM), articulated through the habits of Curiosity, Connections, and Creating Value (the "3Cs") [6, 7, 8, 9, 10]. The EM framework emphasizes exploration of the unknown, integration of knowledge across domains, and consideration of societal and stakeholder impact. While GenAI tools can improve technical efficiency, most commercially available interfaces function as general purpose answer engines and do not explicitly scaffold these entrepreneurial habits. As a result, although GenAI adoption in higher education has expanded rapidly, existing empirical research has largely focused on usage frequency, academic integrity, or performance on discrete technical tasks. Few studies have examined how custom designed AI interfaces can explicitly scaffold professional frameworks such as EM across contrasting engineering disciplines [11, 12]. Moreover, prior quantitative studies of AI interaction data often rely on aggregate descriptive statistics without accounting for the nested structure of educational data (interactions within students, students within courses), raising concerns about pseudo replication and inflated inferences.

To address these conceptual and methodological gaps, this study evaluates a custom built, EM aligned AI chatbot deployed across four engineering courses at a public research university. Unlike standard LLM interfaces, the system includes an optional, visible "EM Alignment Info Box" appended to selected responses. The Info Box labels one of the 3Cs, provides a brief rationale for the alignment, and suggests a next reflective step. Importantly, the system does not measure mindset directly. Rather, it generates structured EM aligned prompts that become observable in interaction logs, allowing analysis of behavioral traces of EM aligned interaction. Using 3,187 interaction turns collected over a 16 week semester, we conduct a log based observational study. We apply Generalized Estimating Equations (GEE) to model associations between EM aligned responses and immediate follow up behavior while accounting for within-student clustering. We also analyze longitudinal changes in inquiry articulation (operationalized as prompt length after removing code blocks). Because EM alignment is triggered by a content based classifier rather than randomized assignment, all findings are associative and should not be interpreted causally. The comparison between EM aligned and non EM responses reflects a design feature of the system rather than an experimental condition.

This study contributes three primary insights. First, it provides empirical evidence about when and how EM aligned scaffolds appear in student AI interactions across disciplinary contexts. Second, it examines whether EM aligned responses are associated with different patterns of immediate engagement. Third, it characterizes longitudinal trends in inquiry articulation over the semester. Rather than claiming evidence of mindset development, we frame our contribution as an analysis of behavioral traces associated with EM aligned AI scaffolding and discuss implications for the design of AI tools intended to support professional frameworks in engineering education.

The study is guided by the following research questions:

- **RQ1 (Immediate Follow up Association):** To what extent is EM aligned AI feedback associated with students' immediate engagement behavior? Specifically, does the presence of EM tags correspond to different follow up probabilities compared to standard responses, and does this association vary by course context?

- **RQ2 (Disciplinary Divergence):** How do EM aligned interaction patterns vary across engineering disciplines? Are certain 3C triggers more prevalent or differently associated with follow up behavior in computational versus physical engineering contexts?
- **RQ3 (Inquiry Articulation Over Time):** Over the course of the semester, how does student inquiry articulation (measured as prompt length after removing code blocks) change within students, and how do these longitudinal patterns co occur with EM aligned interactions?

By reframing the contribution around observable interaction patterns rather than latent mindset change, this work aims to provide a more precise and methodologically grounded account of how EM oriented AI scaffolding functions in practice.

2 Related Work

Ariza et al. [13] and Yan et al. [14] both map the broader GenAI in-education space. Ariza et al. [13] focus on engineering and computing education and find the empirical literature is still heavily perception driven (often around ChatGPT), with relatively few well specified, evidence based teaching methodologies, which results in challenges in comparing outcomes. Yan et al. [14] incorporate prior work by task (e.g., feedback and grading) and warn that adoption is slowed by low readiness, limited transparency, and privacy issues. These works motivate our research by highlighting gaps in evidence based GenAI pedagogy and practice.

Albadarin et al. [15] complements early studies of ChatGPT in education and show that students use it mostly for Q&A and immediate support tasks. Students often report it as helpful, but accuracy problems and the need for enough prior knowledge come up repeatedly. Kao et al. [16] explores a "Socratic" approach where the AI guides students with questions instead of revealing answers, and they test it within an Argument Driven Inquiry lesson using a randomized design with 10th graders. They find the AI group improved on reasoning and engagement measures, but not on self regulation, which may require more time or additional supports. We build on Socratic, non revealing feedback, but implement them via a course grounded RAG chatbot with EM aligned prompts and guardrails.

Shailendra et al. [17] present a 4E framework for GenAI adoption and a matrix for tracking roles, rollout, and outcomes. Lee and Low [18] discusses GenAI should push students to think, not just copy answers. In their approach, students use ChatGPT and then instructors or peers challenge the responses using structured questions and reasoning frameworks. We put these principles to practice in a course integrated tool rather than a purely conceptual framework.

Vittorini et al. [19] build an AI assessment tool for data science homework that grades R code + short written explanations, giving students immediate formative feedback while reducing instructor grading load. Haldeman et al. [20] extend autograding to produce error bucket labels + concept linked hints (mapped to a knowledge map). Their method uses test suite outcome signatures

to classify common logical errors, achieving 90% categorization accuracy and improving completion after initial incorrect submissions. Liffiton et al. [21] present CodeHelp, an LLM based programming helper designed to give tutor like guidance without copyable solutions. They implement "guardrails" via multi step prompting (sufficiency checks, candidate scoring, rewriting to remove code blocks), and a course deployment shows sustained use and positive student perceptions. We build on course grounded RAG responses with EM aligned scaffolding rather than test signature or generic LLM.

Herden et al. [22] highlights that ChatGPT-3.5 works well in intro programming labs for explanation, formatting, and style tasks, but it is less reliable for messy code and refactoring with logical errors, so students must verify and iterate. Chen [23] describes a teams based tutor using GPT-4 + Copilot Studio for real time support and personalized feedback, plus instructor insights; reported benefits exist, but results are limited by single cohort scope and practical usability/stability constraints. Our work provides a course specific RAG design and a longitudinal log based evaluation across multiple courses.

AlRabah et al. [24] analyze semester long logs from a course grounded RAG chatbot across five engineering courses. They find usage is course dependent (conceptual questions dominate, programming help rises in programming heavy courses) and identify policy risks like solution seeking and copy pasted prompts, but they do not measure learning impact. Qiu et al. [25] propose an evaluation framework and test a Socratic style chatbot in higher ed; they find no significant post test difference between chatbot and control in their small study, highlighting that measured learning gains may depend on intervention intensity and study power. We examine how EM settings shape interaction behavior across different subjects while tracking policy risk patterns.

3 Methodology

This study uses a log based observational design to examine how EM aligned scaffolding in a course integrated Retrieval Augmented Generation (RAG) chatbot is associated with student interaction patterns. The chatbot was deployed across four engineering courses at a large public university during the Fall 2025 semester. Across the four courses (total enrollment = 756 students), we collected n = 3,187 interaction turns over 16 weeks. Unless otherwise noted, sample sizes (n) refer to interaction turns (the unit of analysis). Because the EM Info Box is triggered by a content based classifier rather than randomized assignment, comparisons between standard and EM aligned responses are interpreted as associative.

3.1 System Architecture and Implementation

Unlike single model apps, our system uses a mixed architecture for practical validity. Figure 1 shows the pipeline combining input processing, vector retrieval, and a dynamic model routing layer.

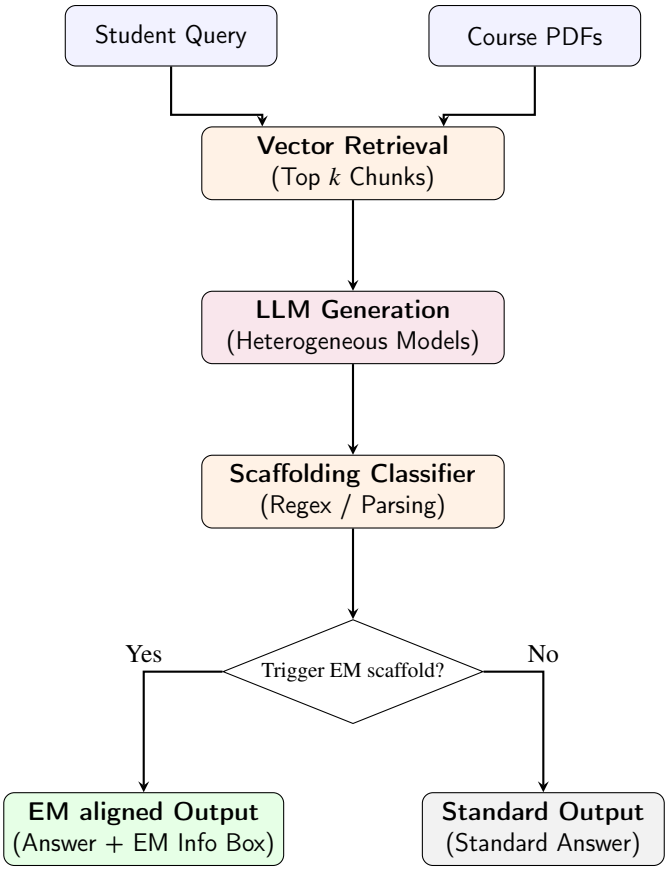

Figure 1: System Logic Flow. The EM Info Box is conditionally triggered by a content based classifier rather than randomized assignment, producing either an EM aligned or a standard response mode.

3.1.1 Dynamic Model Routing

For discipline specific reasoning, the system used a mixed model setup. Logs confirm LLMs varied by course for advanced open source capabilities:

- **Parallel Programming (PP):** Powered by `Qwen-2.5-VL-72B-Instruct` [26] ($> 90\%$ usage) for complex CUDA/C++ reasoning.
- **Database Systems (DS):** Used a hybrid setup of `Llama-3.1-70B` [27] and `Qwen-2.5`, evolving mid semester to improve SQL generation.
- **Nuclear Engineering (NEF):** Powered by `GPT-4.1-Mini` [28], serving as a stable baseline for physics queries.
- **Civil Engineering (EPS):** Powered by `DeepSeek-R1` [29] and `Qwen-2.5`, using chain of thought reasoning for design inquiries.

This heterogeneous deployment reflects authentic course practice; accordingly, course level differences observed in later analyses should be interpreted as emerging from the combined ecology of course context and model affordances rather than being attributed to disciplinary culture alone.

3.1.2 Response Modes: Standard vs EM aligned

The chatbot returns a *standard* technical answer by default. When a content based classifier detects an opportunity to connect the response to the KEEN 3Cs, the system appends an *EM Info Box:* a short, structured addendum containing (1) an EM-3C label (Curiosity, Connections, or Creating Value), (2) a brief rationale ("Why this fits"), and (3) a suggested next step. The trigger is not randomized and is therefore correlated with prompt content and course context. Examples of cues used to trigger the EM Info Box include prompts that (i) ask for explanation or alternatives (Curiosity), (ii) involve trade offs or cross concept constraints (Connections), or (iii) reference performance, cost, safety, or user needs (Creating Value).

Concrete response example (de identified; trimmed): In EM aligned mode, the EM Info Box is appended to the technical answer; in standard mode, the assistant returns the same technical answer without this addendum. Table 1 illustrates the difference using one interaction turn.

Table 1: Example of standard vs EM aligned output (same interaction turn; excerpts).

Standard output (technical answer excerpt)	**EM Info Box addendum (excerpt)**
The estimated number of warps that will have control divergence is around 80. This is based on the assumption that boundary blocks contain threads falling outside image boundaries, causing divergence in the last warps.	*EM-3C Focus: Creating Value.* Why this fits: Understanding control divergence in CUDA kernels supports performance optimization by reducing unnecessary computations and maximizing parallel execution. Next step: Experiment with different block sizes and boundary conditions to observe the impact on divergence.

Note: In standard mode, the assistant provides the technical answer without this EM Info Box addendum.

3.2 Data Collection and Participants

The University Institutional Review Board (IRB) approved this study. For privacy, we hashed user identifiers with SHA 256 and removed personal data before analysis. The dataset includes 3,187 turns over 16 weeks. Table 2 shows engagement by course, with most activity in advanced Parallel Programming (n = 2,672 turns).

Table 2: Dataset Composition: Verified Enrollment and Turn Counts (n = 3,187 turns)

Course Context	**Discipline**	**Enrollment**	**Turns**	**% of Total**
Parallel Programming (PP)	ECE / CS	126	2,672	83.8%
Database Systems (DS)	Computer Science	472	167	5.2%
Civil Engineering (EPS)	Civil Engineering	124	168	5.3%
Nuclear Fundamentals (NEF)	Nuclear Eng.	34	180	5.7%
Total Dataset (*Fall 2025 Semester*)		**756**	**3,187**	**100.0%**

3.3 Data Preparation and Processing

For validity, we used a structured preprocessing pipeline (Python `pandas`) to separate student inquiry from technical noise. We define an interaction *turn* as a student prompt paired with the subsequent assistant response within a session. Data preparation proceeded in three stages: (1) session level normalization to ensure temporal consistency across courses, (2) content level filtering to remove system artifacts and low information messages (noise), and (3) feature extraction to operationalize engagement and inquiry depth for statistical analysis.

3.3.1 Algorithmic Cleaning and Feature Extraction

We extracted interaction features from raw JSONL logs by parsing nested message lists for each session. Timestamps were converted to UTC using `pandas.to_datetime` to ensure longitudinal consistency over the 16 week semester. To accurately measure student inquiry depth, we used a regular expression to strip Markdown code blocks from each message, isolating only the student's natural-language text for length analysis.

3.3.2 Variable Operationalization

We operationalized three primary variables for quantitative analysis. First, **Has_EM** (an EM Info Box indicator) indicates response mode: interactions were coded as 1 if the assistant response included an EM Info Box (i.e., contained an explicit EM-3C label) and 0 otherwise; this was implemented via strict regex matching of output metadata. Second, the dependent variable **Follow up** was a binary flag indicating whether a student continued the session with a subsequent non noise message. Finally, the covariate **Log Prompt Length** was calculated as $\ln(\text{length} + 1)$ to normalize the character counts of student prompts (after code block removal).

3.4 Statistical Analysis Models

Because multiple interaction turns are nested within students, we used marginal models that account for within-student clustering to avoid pseudo replication. For RQ1 and RQ2, we fitted a population average logistic model via Generalized Estimating Equations (GEE), which yields robust standard errors for correlated observations. Throughout, we emphasize that Has_EM is content triggered rather than randomized; therefore, coefficients are interpreted as associations.

3.4.1 Generalized Estimating Equations (GEE)

For RQ1 and RQ2, we fitted a GEE model using `statsmodels.genmod`. We selected a Binomial family with a logit link function because the outcome variable (Follow up) is binary. An Exchangeable correlation structure was chosen to assume a constant correlation between interactions from the same student, which provides robust estimates when students engage with the system multiple times over a semester.

The model is:

$$\text{logit}(P_{followup}) = \beta_0 + \beta_1 X_{EM} + \beta_2 X_{Course} + \beta_3 (X_{EM} \times X_{Course}) + \beta_4 X_{Length} \quad (1)$$

The interaction term β_3 tests whether the association between Has_EM and follow up varies by course context. Course context was encoded as a categorical variable using treatment (dummy) coding via `C(Course)` in `statsmodels`; DS was the reference level (omitted category), so the intercept and main Has_EM term correspond to DS. Because NEF had Has_EM = 0 for all turns (standard only interface), the Has_EM $\times$ NEF interaction is not identifiable and is reported as NA.

3.4.2 Growth Trend Analysis

For RQ3, we tracked inquiry depth evolution. For each student (i), we calculated the linear slope (λ_i) of Prompt Length for each student over 16 weeks. We used the nonparametric Wilcoxon Signed Rank Test (`scipy.stats.wilcoxon`) to check if the median slope exceeded zero. This method was selected for its robustness against non normal distributions and outliers, providing a conservative estimate of longitudinal articulation growth.

4 Results

In this section, we present findings from the analysis of n = 3,187 interaction turns. Using the disciplinary contexts from our Methodology, we used GEE to quantify associations between EM scaffolding and follow up (RQ1), examined trigger distribution (RQ2), and tracked inquiry depth trends over time (RQ3).

4.1 RQ1: Context Dependent Follow up Patterns

To examine how EM aligned feedback is associated with immediate engagement, we analyzed follow up probability. The GEE estimates suggest that the association between scaffolding and follow up varies by disciplinary context (Figure 2; Table 3).

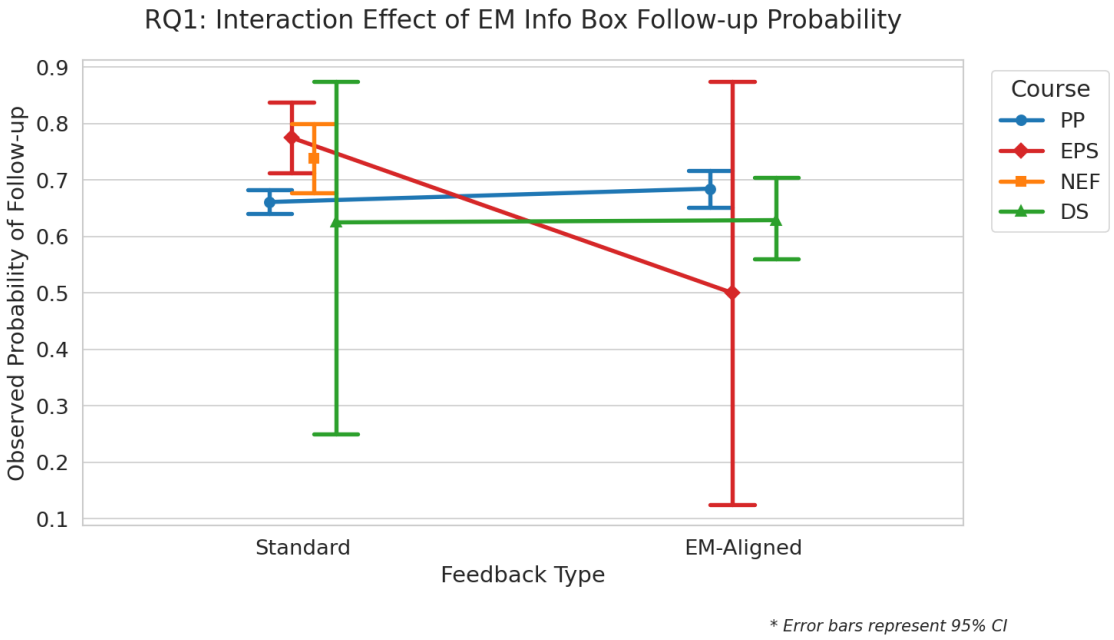

Figure 2: **Interaction Effect of the EM Info Box on Follow up Probability (with 95% CI).** The **Blue line (PP)** shows a near null association with relatively tight uncertainty. The **Red line (EPS)** shows lower follow up probability for EM aligned turns, but wide confidence intervals reflect sparse EM triggers. The **Green line (DS)** shows a negative point estimate, with high uncertainty due to few standard turns. *Note: NEF (Orange) used a standard only interface (no EM triggers) and is shown as a baseline course.*

Informational Resolution in Design (EPS): In EPS, follow up probability was lower for EM aligned turns than for standard turns (0.50 vs. 0.78; Figure 2). The Has_EM $\times$ EPS interaction was negative ($p = 0.097$, Table 3) but estimated with wide uncertainty due to sparse EM triggers (4.8% of EPS turns). One plausible interpretation is informational resolution: by explicitly linking technical constraints to value considerations in the EM Info Box, the system may help students reach a satisfactory stopping point within a single turn. However, given the observational design and small EM sample in EPS, alternative explanations (e.g., prompt type or task completion) remain plausible.

Saturation in Database Systems (DS): In DS (the GEE reference group), the estimated Has_EM coefficient was negative (-0.49) but not statistically significant ($p = 0.288$, Table 3). Interpretation is limited because EM scaffolds appeared in 95% of DS turns, leaving only $n = 8$ standard turns and therefore high uncertainty.

Table 3: GEE Analysis of Follow up Probability (Reference Group: DS). Note: NEF had no EM triggers; Has_EM × NEF is not estimable.

Predictor	**Coef.**	**Std. Err.**	**P value**	**95% CI**
Intercept (DS Baseline)	0.66	0.60	0.272	[-0.52, 1.84]
Has_EM (Effect in DS)	-0.49	0.46	0.288	[-1.39, 0.41]
Log_Length	0.01	0.03	0.720	[-0.05, 0.07]
Interaction Terms (Difference from DS):				
Has_EM × EPS (Civil)	-0.91	0.55	0.097	[-1.99, 0.17]
Has_EM × PP (Parallel Prog.)	0.42	0.47	0.372	[-0.50, 1.34]
Has_EM × NEF (Nuclear)	*NA*	*NA*	*NA*	*NA*

Task Driven Resilience in Computing (PP): In PP, follow up probabilities were similar across response modes, and the Has_EM × PP interaction was not statistically significant ($p = 0.372$, Table 3). This pattern is consistent with many interactions being debugging oriented, where students prioritize functional resolution and may treat metacognitive prompts as secondary.

4.2 RQ2: Disciplinary Divergence

While RQ1 modeled the association between the presence of any EM Info Box and student follow up behavior, this section shifts the focus to the *content* of EM aligned turns. Specifically, we report how often each EM dimension (Curiosity, Connections, and Creating Value) was triggered across courses. Because these triggers were produced by prompt engineered rules and a content based classifier rather than randomized assignment, the results in Table 4 describe the distribution of scaffolding opportunities rather than causal effects.

Table 4: Disciplinary Divergence: EM Trigger Frequency by Course (n = 3,187 turns)

Course Context	**Standard**	**EM Info Box Triggers**			**Total Turns**
	No EM Info Box	**Curiosity**	**Connections**	**Creating Value**	
PP (Parallel Prog.)	1,784 (67%)	**293 (11%)**	**179 (7%)**	**416 (16%)**	2,672
DS (Database Sys.)	8 (5%)	76 (46%)	28 (17%)	55 (33%)	167
EPS (Civil Eng.)	160 (95%)	6 (4%)	0 (0%)	2 (1%)	168
NEF (Nuclear Eng.)	180 (100%)	0 (0%)	0 (0%)	0 (0%)	180
Total	**2,132**	**375**	**207**	**473**	**3,187**

Note: Percentages are row-wise (within course) and may not sum to 100% due to rounding.

Table 4 highlights substantial heterogeneity in when the EM Info Box was triggered. In DS, the EM Info Box appeared in 95.2% of turns, leaving limited standard only comparisons; in EPS,

EM triggers were rare (4.8%), contributing to wide uncertainty in RQ1; and in NEF, the interface operated in a standard only mode by design (0% EM). In PP, Creating Value was the most common EM dimension (416 instances; 16% of PP turns and 88% of all Creating Value triggers). In this computational context, EM Info Boxes often connected performance and efficiency trade offs (e.g., latency, throughput, speedup) to practical value considerations such as resource cost, scalability, or user impact. Because EM triggering is content based, these differences likely reflect both prompt distributions and classifier coverage across courses. Because PP accounts for 83.8% of all turns in the dataset, aggregate trigger totals are also shaped by this course's much higher interaction volume.

4.3 RQ3: Longitudinal Inquiry Articulation Trends

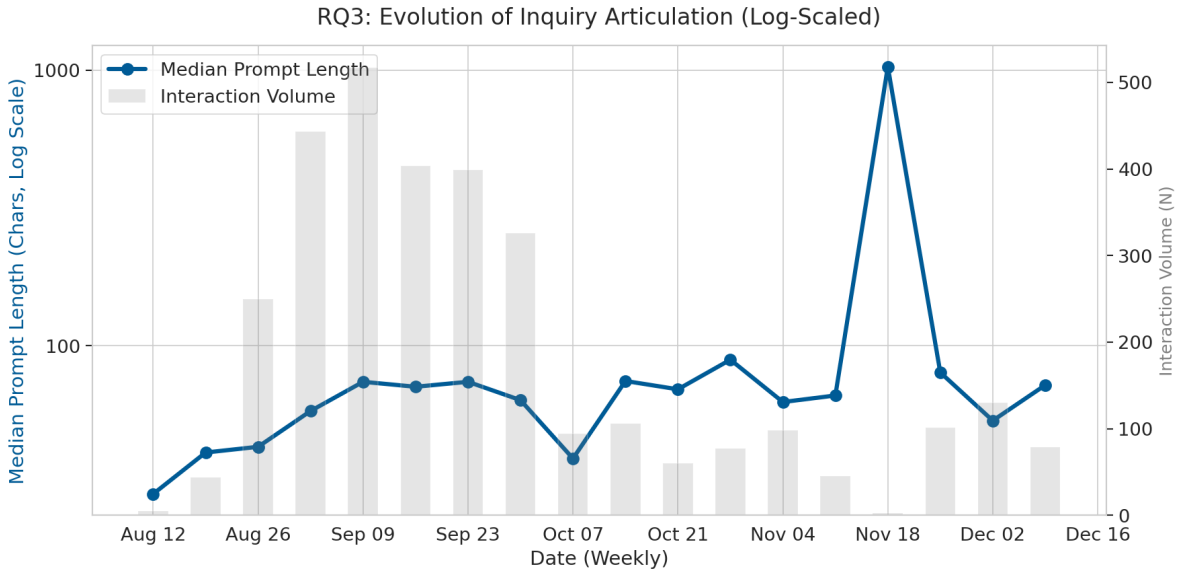

Figure 3: **Longitudinal Evolution of Inquiry Articulation.** The blue line tracks weekly median prompt length (log scaled), and gray bars show interaction volume. The November spike (week of Nov 18) aligns with major project deadlines.

Our analysis shows a positive within student trend in natural language inquiry articulation. Across 661 users, the median number of turns per user was 2 (interquartile range, IQR: 1–5), indicating highly skewed usage. Here, users refer to distinct chatbot account IDs with at least one recorded interaction turn during the study period. To reduce survivorship bias in longitudinal inference, we estimated a growth trajectory for each student with sufficient repeated observations (329 users with ≥ 3 turns). The median slope was +2.6 characters/week (Wilcoxon signed rank test, $p = 0.012$). Because preprocessing removed Markdown code blocks, these changes reflect growth in students' natural language problem descriptions rather than increased code length.

Figure 3 also shows a pronounced November spike (week of Nov 18): the weekly median prompt length increased from a median of 67.8 characters in the prior four weeks (range: 62.5–89.0) to 1,027 characters in the spike week, then returned toward the pre spike range (median: 72.0;

range: 54.0–80.0 over the subsequent four weeks). This temporal pattern aligns with major project deadlines and suggests a shift toward troubleshooting oriented prompts that include substantially more contextual detail (e.g., logs or specifications).

5 Discussion

This study examines where and how EM aligned scaffolding appears in a course integrated AI assistant and how it co occurs with interaction behavior across disciplines. Using GEE and longitudinal descriptive analyses of n = 3,187 interaction turns, we highlight three patterns: context dependent follow up associations in design contexts, discipline specific interpretations of "Creating Value" in computing, and deadline related shifts in inquiry depth. Because scaffolding was triggered by a content based classifier rather than randomized assignment, these findings should be interpreted as associative.

5.1 Informational Resolution: Efficiency over Engagement

In EPS, EM aligned turns were associated with lower follow up probability (Figure 2; Table 3), though estimates are imprecise due to sparse EM triggers. A plausible explanation is informational resolution: by explicitly linking technical constraints to value considerations in the EM Info Box, the system may help some students reach a satisfactory stopping point within a single turn. We interpret this possibility through the lens of need for closure [30]. However, we do not directly measure cognitive states, and follow up behavior can also reflect factors such as prompt type, task completion, or fatigue; therefore, this interpretation should be treated as a hypothesis for future validation.

5.2 Redefining Value in Computing

A striking descriptive result is the concentration of Creating Value triggers in Parallel Programming (416 of 473 instances; 88%). In this course context, EM Info Boxes often framed optimization, throughput, and efficiency trade offs as value, suggesting that students may articulate "creating value" in discipline specific terms (e.g., efficiency as-value in high performance computing). In RQ1, follow up probabilities in PP were similar across response modes, indicating limited evidence that the presence of an EM Info Box changes immediate follow up behavior in this debugging oriented context. Together, these results suggest that EM aligned prompts can surface value language in computing, but their relationship to immediate engagement may depend on task context and timing.

5.3 The November Spike: Deadline Driven Troubleshooting

The November spike illustrates how assessment pressure moderates tool use. In the week of Nov 18, median prompt length exceeded 1,000 characters, compared with roughly 70–90 characters in nearby weeks, suggesting a shift from short conceptual questions to longer troubleshooting prompts containing extensive context. This highlights a design implication: EM oriented AI tools may benefit from distinct interaction modes (e.g., conceptual coaching vs. debugging support) that accommodate deadline related behavior without over interpreting it as mindset change.

5.4 Limitations and Behavioral Inference

A primary limitation of this study is that we do not directly measure Entrepreneurial Mindset (EM) as a latent construct; instead, we analyze behavioral traces from interaction logs. As a result, we cannot claim mindset development, and observed patterns should be interpreted as interaction level associations. We also do not report per user usage distributions (e.g., turns per user) beyond the subset used for RQ3; future work should characterize usage heterogeneity and assess how it may shape interaction level inferences. A second limitation is that EM Info Boxes were triggered by a content based classifier rather than randomized assignment. This makes Has_EM endogenous to prompt content and course context: differences in follow up may reflect prompt type, task completion, timing, and heterogeneous underlying models (Figure 1) in addition to the presence of scaffolding, even after controlling for prompt length. Finally, we report GEE results under an exchangeable working correlation. As sensitivity checks (Appendix A), we refit the model assuming independence and refit excluding NEF; in these DS referenced specifications, the Has_EM $\times$ EPS contrast remained negative with similar magnitude (with $p \approx 0.10$). To address DS near saturation (95% EM triggers), we also refit using PP as the reference on the PP+EPS subset; the contrast remained negative. This PP+EPS refit is included as a diagnostic sensitivity check (not a primary analysis), since effect sizes and p-values can shift with the reference group and sample composition in an observational, content triggered setting. However, the Log_Length association varied across specifications, so coefficients should be interpreted cautiously. Future work should test the stability of these findings under alternative specifications and triangulate log patterns with validated EM assessments (e.g., surveys or interviews).

6 Conclusion

This study evaluates an Entrepreneurial Mindset (EM)-aligned, course integrated RAG chatbot across four engineering courses using n = 3,187 interaction turns. Because EM Info Boxes were triggered by a content based classifier rather than randomized assignment, our findings describe associative patterns in interaction logs rather than causal effects or measured mindset change. For RQ1, the association between EM aligned responses and immediate follow up varied by context. In the civil engineering design course, EM aligned turns were associated with lower follow up probability, a pattern consistent with informational resolution in single turn responses but estimated with wide uncertainty due to sparse EM triggers. For RQ2, EM Info Boxes most fre-

quently triggered Creating Value in Parallel Programming (88% of Creating Value instances), often framing efficiency and performance trade offs as value. For RQ3, prompt length showed modest within student growth and a pronounced deadline related spike, indicating shifts between conceptual questioning and troubleshooting oriented use. Overall, the study contributes behavioral trace evidence and design implications for EM oriented AI tools: scaffolds may need to be context aware, discipline specific, and robust to deadline related usage. Future work should triangulate log patterns with validated EM assessments and adopt designs with stronger identification (e.g., randomized or balanced comparisons) to test causal impacts.

A Robustness and Sensitivity Checks

To assess the sensitivity of the follow up model to specification choices, we refit the GEE under alternative working correlation assumptions and refit excluding NEF (which used a standard only interface and yields a non estimable Has_EM $\times$ NEF interaction). Table 5 summarizes key coefficients for the primary terms of interest.

Table 5: Sensitivity checks for the GEE follow up model (key coefficients). Cells report coefficient (standard error, SE), p-value. "Exchangeable" is the main specification; "Independence" assumes working independence; "Exchangeable (no NEF)" refits the model excluding NEF; "Exchangeable (PP+EPS only)" refits on the PP+EPS subset with PP as the reference group (reporting the Has_EM $\times$ EPS contrast).

Term	Exchangeable	Independence	Exchangeable (no NEF)	PP+EPS only (PP ref.)
Has_EM (reference)	-0.491 (0.462), 0.288	0.076 (0.744), 0.919	-0.505 (0.449), 0.261	-0.083 (0.107), 0.435
Has_EM $\times$ EPS	-0.914 (0.551), 0.097	-1.326 (0.811), 0.102	-0.900 (0.539), 0.095	-1.320 (0.316), < 0.001
Has_EM $\times$ PP	0.423 (0.474), 0.372	0.021 (0.759), 0.978	0.427 (0.462), 0.355	*NA*
Log_Length	0.011 (0.032), 0.720	0.086 (0.043), 0.044	0.013 (0.032), 0.684	0.002 (0.031), 0.939

Note: The PP+EPS only refit is a diagnostic sensitivity check and should not be interpreted as stronger evidence than the main DS referenced specification. A first order autoregressive (AR(1)) working correlation (using week index) was attempted but did not converge and is not reported.

References

[1] U. Mittal, S. Sai, V. Chamola, and D. Sangwan, "A comprehensive review on generative ai for education," *Ieee Access*, vol. 12, pp. 142 733–142 759, 2024.

[2] A. Yusuf, N. Pervin, M. Román-González, and N. Noor, "Generative ai in education and research: A systematic mapping review," *Review of Education*, vol. 12, no. 2, p. e3489, 2024.

[3] S. Feuerriegel, J. Hartmann, C. Janiesch, and P. Zschech, "Generative ai," *Business & Information Systems Engineering*, vol. 66, no. 1, pp. 111–126, 2024.

[4] S. Lau and P. Guo, "From" ban it till we understand it" to" resistance is futile": How university programming instructors plan to adapt as more students use ai code generation and

explanation tools such as chatgpt and github copilot," in *Proceedings of the 2023 ACM Conference on International Computing Education Research-Volume 1*, 2023, pp. 106–121.

[5] R. Liu, C. Zenke, C. Liu, A. Holmes, P. Thornton, and D. J. Malan, "Teaching cs50 with ai: Leveraging generative artificial intelligence in computer science education," in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 2024, pp. 750–756.

[6] J. M. Kriwall and K. Mekemson, "Instilling the entrepreneurial mindset in next-generation engineers," in *2010 ASEE Annual Conference & Exposition*, 2010, pp. 15–734.

[7] Engineering Unleashed (KEEN), "The keen framework," https://engineeringunleashed.com/framework, accessed: 2026-01-21.

[8] S. Shrivastav and P. Rao, "Defining, cultivating, and assessing entrepreneurial mindset in higher education: a systematic review and future research directions," *European Journal of Engineering Education*, pp. 1–34, 2025.

[9] E. C. Howard and D. I. Castaneda, "Entrepreneurs in the making: A qualitative exploration of entrepreneurial mindset in engineering learners," in *2025 Systems and Information Engineering Design Symposium (SIEDS)*. IEEE, 2025, pp. 215–220.

[10] M. Morin, B. Hutson, and R. Goldberg, "The impact of a faculty learning community for students' entrepreneurial mindset development in stem education," *The Journal of Faculty Development*, vol. 40, no. 1, pp. 16–28, 2026.

[11] H. Nguyen and V. Allan, "Using gpt-4 to provide tiered, formative code feedback," in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 2024, pp. 958–964.

[12] P. Denny, J. Leinonen, J. Prather, A. Luxton-Reilly, T. Amarouche, B. A. Becker, and B. N. Reeves, "Prompt problems: A new programming exercise for the generative ai era," in *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 2024, pp. 296–302.

[13] J. Á. Ariza, M. B. Restrepo, and C. H. Hernández, "Generative ai in engineering and computing education: A scoping review of empirical studies and educational practices," *IEEE Access*, 2025.

[14] L. Yan, L. Sha, L. Zhao, Y. Li, R. Martinez-Maldonado, G. Chen, X. Li, Y. Jin, and D. Gašević, "Practical and ethical challenges of large language models in education: A systematic scoping review," *British Journal of Educational Technology*, vol. 55, no. 1, pp. 90–112, 2024.

[15] Y. Albadarin, M. Saqr, N. Pope, and M. Tukiainen, "A systematic literature review of empirical research on chatgpt in education," *Discover Education*, vol. 3, no. 1, p. 60, 2024.

[16] S. Kao, P. Grant, and S. Woltering, "Socratic ai in k–12 science classrooms: Effects on critical thinking, motivation, and self-regulation in a randomized controlled trial," 2025.

[17] S. Shailendra, R. Kadel, and A. Sharma, "Framework for adoption of generative artificial intelligence (genai) in education," *IEEE Transactions on Education*, 2024.

[18] C. C. Lee and M. Y. H. Low, "Using genai in education: The case for critical thinking," *Frontiers in Artificial Intelligence*, vol. 7, p. 1452131, 2024.

[19] P. Vittorini, S. Menini, and S. Tonelli, "An ai-based system for formative and summative assessment in data science courses," *International Journal of Artificial Intelligence in Education*, vol. 31, no. 2, pp. 159–185, 2021.

[20] G. Haldeman, M. Babeş-Vroman, A. Tjang, and T. D. Nguyen, "Csf: Formative feedback in autograding," *ACM Transactions on Computing Education (TOCE)*, vol. 21, no. 3, pp. 1–30, 2021.

[21] M. Liffiton, B. E. Sheese, J. Savelka, and P. Denny, "Codehelp: Using large language models with guardrails for scalable support in programming classes," in *Proceedings of the 23rd Koli Calling International Conference on Computing Education Research*, 2023, pp. 1–11.

[22] O. Herden *et al.*, "Integration of chatbots for generating code into introductory programming courses," in *Conference Proceedings. The Future of Education 2024*, 2024.

[23] W.-Y. Chen, "Intelligent tutor: Leveraging chatgpt and microsoft copilot studio to deliver a generative ai student support and feedback system within teams," *arXiv preprint arXiv:2405.13024*, 2024.

[24] A. AlRabah, M. Blumthal, S. Koloutsou-Vakakis, V. Kindratenko, T. Kozlowski, Z. Li, and A. Alawini, "Data-driven insights into ai-powered learning: Analyzing student interactions with ai-bot in engineering education," in *2025 ASEE Annual Conference & Exposition*, 2025.

[25] W. Qiu, C. L. Su, N. B. Jamil, M. Thway, S. S. H. Ng, L. Zhang, F. S. Lim, and J. W. Lai, "A systematic approach to evaluate the use of chatbots in educational contexts: Learning gains, engagements and perceptions," *Computers*, vol. 14, no. 7, p. 270, 2025.

[26] S. Bai, K. Chen, X. Liu, J. Wang, W. Ge, S. Song, K. Dang, P. Wang, S. Wang, J. Tang *et al.*, "Qwen2.5-vl technical report," *arXiv preprint arXiv:2502.13923*, 2025.

[27] A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Yang, A. Fan *et al.*, "The llama 3 herd of models," *arXiv e-prints*, pp. arXiv–2407, 2024.

[28] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

[29] D. Guo, D. Yang, H. Zhang, J. Song, R. Zhang, R. Xu, Q. Zhu, S. Ma, P. Wang, X. Bi *et al.*, "Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning," *arXiv preprint arXiv:2501.12948*, 2025.

[30] D. M. Webster and A. W. Kruglanski, “Individual differences in need for cognitive closure,” *Journal of Personality and Social Psychology*, vol. 67, no. 6, pp. 1049–1062, 1994.