

Designing for Discernment: A Mixed-Methods Case Study of GenAI-Supported Critical Thinking

Yueyue Li¹, Jie Wu², & Zheng Yang³

¹University of Electronic Science and Technology of China, Chengdu, Sichuan, China;

²Chengdu University, Chengdu, Sichuan, China;

³Sichuan University, Chengdu, Sichuan, China

Abstract

Against the backdrop of the rapid proliferation of generative AI tools, which bring both new opportunities and challenges to higher education, particularly in cultivating critical thinking within STEM fields, this paper probes into how integrating a Generative AI (GenAI) tool into an undergraduate engineering course impacts students' critical thinking and teamwork, and examines strategies for instructors to facilitate ethical, effective use of AI in learning. In a one-semester case study involving 69 students, a novel "GenAI-CT" instructional model was deployed, integrating GenAI-driven activities with critical thinking exercises. Employing a mixed-methods approach, quantitative tests gauged improvements in critical thinking skills, while qualitative analyses of student reflections and group interaction logs elucidated their experiences of engaging with GenAI. The integration of GenAI was generally feasible and beneficial. Students showed improvement in certain critical thinking tasks (such as problem analysis and idea synthesis), and reported that GenAI support enriched group collaboration by offering timely feedback and diverse perspectives. However, significant challenges were observed—many students struggled with tasks requiring complex logical structuring even with AI assistance, and there was a tendency for some to over-rely on AI-generated answers without sufficient critical evaluation. This study introduces a GenAI-CT framework, which delineates the ethical and pedagogical integration of AI tools into STEM education. Our findings furnish empirical evidence that GenAI can augment learning when utilized responsibly, and offer practical insights for educators striving to balance innovative AI assistance with the pivotal goal of fortifying students' independent critical thinking skills.

Keywords: Critical Thinking, Generative Artificial Intelligence, STEM Education, Higher Education, Pedagogical Strategies

1 Introduction

The rapid advancement of generative artificial intelligence (GenAI) presents distinct challenges for undergraduate STEM education, particularly in the teaching of critical thinking. As students increasingly rely on GenAI as an authoritative resource for solving complex problems, educators have raised concerns that this trend could undermine independent reasoning—a foundational element of STEM disciplines.

In STEM education, independent inquiry and rigorous logical reasoning are essential, yet traditional "standard answer"-focused instruction has often failed to cultivate these skills systematically. If

students depend too heavily on GenAI, this deficit can be compounded: they may begin to accept algorithmic outputs as valid reasoning, relying on cognitive shortcuts that bypass thorough problem decomposition and logical verification. This conflation of AI-generated answers with objective truth reinforces a “technology-first” bias that undermines both critical thinking and practical problem-solving, twin pillars of engineering competence. For example, students might rapidly generate experimental code but lack the understanding to justify their algorithmic choices, or they may enumerate communication protocols without the ability to discuss their trade-offs. These gaps strike at the core of engineering practice, where justifying solutions and navigating contextual trade-offs are essential for addressing real-world uncertainties (Ahern et al., 2019; Wu, 2023).

While GenAI can stimulate inquiry and provide rapid access to information, educators must emphasize its limitations. Overreliance on GenAI can atrophy the development of independent critical thinking (Dasari et al., 2024; Chanpradit, 2025; UNESCO, 2023). Moreover, despite its potential to surface valuable insights, GenAI is prone to propagating misinformation and biased narratives, posing a critical pedagogical challenge. Students must learn to distinguish algorithmic outputs from verified knowledge, a skill central to cultivating intellectual integrity (Ji et al., 2023). Original insight, ethical judgment, and context-sensitive problem-solving are hallmarks of rigorous intellectual training and emerge from iterative human inquiry rather than from the passive absorption of AI-generated solutions.

Cultivating critical thinking in an era of widespread GenAI use requires innovative pedagogical frameworks. STEM disciplines increasingly grapple with complex, open-ended challenges—from ethical algorithm design to context-specific problem solving—making human-centered reasoning indispensable. Educational strategies must therefore not only integrate AI-driven tools but also actively promote critical appraisal of their outputs.

To address these needs, we present a case study of critical thinking instruction integrated with GenAI, introducing a “GenAI-CT” model. This model systematically builds students’ skills in logical analysis, dimensional identification, and consensus-building through a phased implementation process involving problem screening and verification, categorization framework debates, and sub-topic reconstruction. Guided by both quantitative (pre/post-test scores) and qualitative (student reflection logs, group collaboration records) data from an exploratory implementation of GenAI-CT, we investigate three core research questions:

1. Is integrating GenAI into science and engineering critical thinking teaching activities feasible?
2. How does GenAI impact the development of students’ critical thinking and team collaboration capabilities?
3. To what extent can GenAI augment or redefine traditional teacher roles (guidance, instruction, feedback, assessment) in STEM education?

2 Literature Review

2.1 GenAI in Higher Education

Recent literature emphasizes that generative artificial intelligence (GenAI) is profoundly reshaping higher education, bringing both opportunities and challenges. Scholars widely agree that GenAI can significantly support teaching and learning. For instance, Zhang et al. (2024) found that GenAI can function as both a “learning assistant” (helping students understand complex material) and a “teaching enabler” (supporting instructors in designing coursework). Similarly, Salameh (2024) highlights GenAI’s potential to enhance learning processes and student experiences. These studies suggest GenAI creates novel educational opportunities, from adaptive feedback systems to innovative assessment models.

However, researchers also warn against uncritical adoption of GenAI. Generative models often produce imperfect outputs: they can exhibit hallucinations, overgeneralizations, or random errors (De Paoli, 2023; Alghamdi, 2024; Hu et al., 2024), which risk misleading learners if not carefully evaluated. Ethical concerns have also been raised. For example, Salameh (2024) points out risks such as algorithmic bias and privacy breaches, and cautions that AI-driven systems could “depersonalize” learning and widen social inequalities. Educators are therefore advised to balance GenAI’s advantages with safeguards that preserve academic standards and fairness.

The scholarship on GenAI in education is diverse in scope and methodology. Some studies use systematic reviews to map the landscape: Zhang et al. (2024) and Khan et al. (2024) review hundreds of articles to identify common themes. Others propose practical frameworks: Shailendra et al. (2024), for instance, offer a step-by-step model for university adoption of GenAI, including curriculum redesign and faculty training. Discipline-specific analyses add further nuance. Khan et al. (2024) report that GenAI is valuable for personalized learning in engineering (for example, offering coding assistance and design support), while Qu et al. (2024) found that students in applied fields (like engineering) use GenAI more extensively than those in theoretical fields, suggesting that integration strategies must align with each discipline’s norms and epistemology.

Future research should bridge these divergences. Experts recommend focusing on context-sensitive integration strategies, refining ethical guidelines, and leveraging GenAI’s strengths without compromising educational integrity. Developing discipline-aligned AI tools and establishing policies to ensure that GenAI enhances rather than undermines learning outcomes will be critical steps going forward.

2.2 Integration of Generative AI with Critical Thinking Instruction

In STEM and engineering education, fostering critical thinking remains a core goal even as GenAI becomes widespread. Recent studies suggest that GenAI can be integrated into critical thinking instruction to good effect when done thoughtfully. For example, Yi Wu (2024) argues that educators should shift from merely transmitting knowledge to explicitly cultivating students’ analytical and evaluative skills in the AI era. Empirical research supports this shift:

- **Enhanced student thinking:** A large-scale study by Hong and Guo (2024) found that GenAI-supported teaching strategies (such as encouraging students to critically evaluate AI-generated responses) significantly improved learners’ critical-thinking skills.
- **Student perceptions:** Ruiz-Rojas et al. (2024) report that 64% of surveyed students believed that using AI tools strengthened their own critical-thinking abilities.
- **Cognitive support:** Hong et al. (2025) found that GenAI can help students develop analytical and reflective skills by providing feedback and helping to construct arguments.

These findings indicate practical ways GenAI can empower learners. However, most case studies of GenAI-assisted critical thinking instruction come from the humanities or social sciences. In many STEM courses, traditional assessments emphasize obtaining correct answers rather than analyzing reasoning. Consequently, critical thinking has often been treated as a peripheral “soft skill” rather than a core educational outcome. This gap suggests that critical thinking must be made a central priority in STEM education.

At the same time, scholars caution against potential pitfalls of GenAI integration. Yi Wu (2024) likens excessive reliance on GenAI to a “crutch” that can weaken students’ motivation for independent reasoning. Lee and Low (2024) note that students express concerns about academic integrity and the loss of learning value when they overuse AI tools. Additionally, imperfections in GenAI outputs can mislead learners if not monitored. Discipline differences matter too: Qu et al. (2024) observed that students in applied fields (e.g., engineering) use GenAI more extensively than those in theoretical fields, implying that teaching strategies should be adapted to each field’s context.

To address these challenges, experts recommend an “empowering, not replacing” approach to using GenAI. Practical teaching strategies include:

- **Teaching AI literacy:** Incorporating lessons on AI fundamentals and how to critically evaluate AI-generated content (as Zhang et al., 2024, suggest) so students learn to question and verify information from GenAI.
- **Supplementary use of GenAI:** Using GenAI as a supporting tool in collaborative, problem-based activities rather than as the primary source of answers. For example, instructors might assign group projects where students first generate ideas with GenAI and then critique or expand upon them, as demonstrated in the case study by Lee and Low (2024).
- **Structured reflection:** Designing tasks that require students to reflect on and discuss AI-generated suggestions, thereby practicing higher-order thinking about both the content and its sources.

By strategically designing instruction around these principles, educators can leverage GenAI’s strengths (such as rapid information generation) while still cultivating rigorous critical-thinking habits.

2.3 Research Gaps and Future Directions

The literature review highlights several important gaps and avenues for future research:

- **Longitudinal impact studies:** Current research primarily examines short-term outcomes. Few longitudinal studies have tracked how GenAI exposure affects students’ cognitive development, disciplinary identity, or career readiness over time. Both Zhang et al. (2024) and Marchena Sekli et al. (2024) note this as a critical gap.
- **Cross-cultural and contextual analyses:** Few studies deeply analyze how cultural, pedagogical, or regulatory differences shape GenAI integration. Preliminary work (Siraj et al., 2025; Zhao et al., 2025) points to regional variation in student attitudes, but more comprehensive cross-cultural comparisons are needed.
- **Disciplinary specificity:** Many studies generalize across fields, but the interaction between GenAI and discipline-specific epistemologies is underexplored. Beyond engineering (Khan et al., 2024), little is known about GenAI’s role in pure science, mathematics, or other domains. Future work should examine how GenAI tools can be aligned with diverse curricular goals and ways of thinking in each discipline.
- **Ethical and practical frameworks:** Although some authors (Salameh, 2024; Zhang et al., 2024) identify risks like bias and academic integrity issues, few studies propose actionable guidelines for classroom use. Future research should co-design ethical guidelines and practical policies with educators and policymakers to ensure that GenAI promotes equity and quality.
- **GenAI and metacognition:** The relationship between GenAI use and metacognitive skills (such as self-regulation and independent problem-solving) requires deeper study. Early work (Wu, 2024; Hong et al., 2025) suggests that AI tools can support metacognition (for example, by providing feedback), but optimal instructional designs remain unclear. Investigating adaptive GenAI models that respond to individual learners’ needs could help personalize support without undermining students’ own cognitive engagement.

Addressing these gaps will be critical for ensuring that GenAI advances educational quality and equity. By conducting longitudinal and cross-disciplinary studies and by co-developing context-sensitive integration strategies, researchers and educators can help transform GenAI from a disruptive novelty into a constructive, equitable driver of educational innovation.

3 Methodology

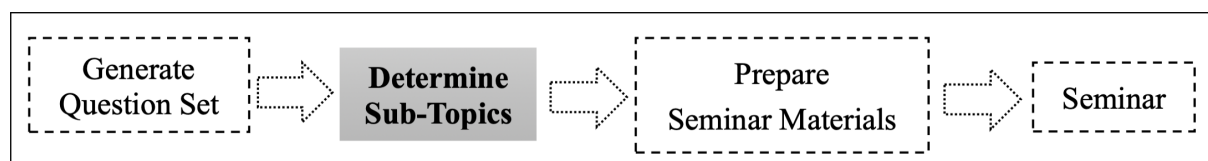
The IoT Application System Design course (a third-year undergraduate course in Internet of Things engineering) is a 48-hour core class that combines 32 hours of theory with 16 hours of lab

work. Its goals are to integrate foundational knowledge (from prior courses on embedded systems, communications, etc.) into mastering system design concepts and to cultivate students' ability to design innovative, industry-aligned IoT systems. In recent years, instructors have emphasized critical thinking in this course, but three main challenges emerged: (1) a traditional lecture-heavy format led students to passively absorb information rather than actively inquire; (2) group activities were often superficial because roles and tasks were unclear, so debates and deep analysis did not occur; (3) large class sizes (60+ students, ~10+ teams) prevented instructors from giving timely, individualized guidance, resulting in uneven development of critical thinking across teams.

To address these bottlenecks, this study designs a GenAI-enabled critical thinking pedagogy (GenAI-CT). By leveraging generative AI tools, the pedagogy creates a learning environment that explicitly fosters critical inquiry and reflection while students work through a structured design task. This offers an innovative model for STEM courses: students learn to treat AI as a cognitive assistant rather than a crutch, developing both analytical rigor and collaborative skills. As a case study, the GenAI-CT model is applied to one phase of the course: Requirements Analysis, which covers four primary teaching activities (see Figure 1). Instead of a lecture on requirements analysis, the course is organized into student-driven activities. In particular, the second activity of the Requirements Analysis phase (spanning 4 class hours) asks teams of 5–6 students to analyze a question set from prior tasks, ultimately defining seminar topics for deeper discussion. GenAI tools are introduced as part of this process.

Figure 1

Primary teaching activities in the “Requirements Analysis” phase.



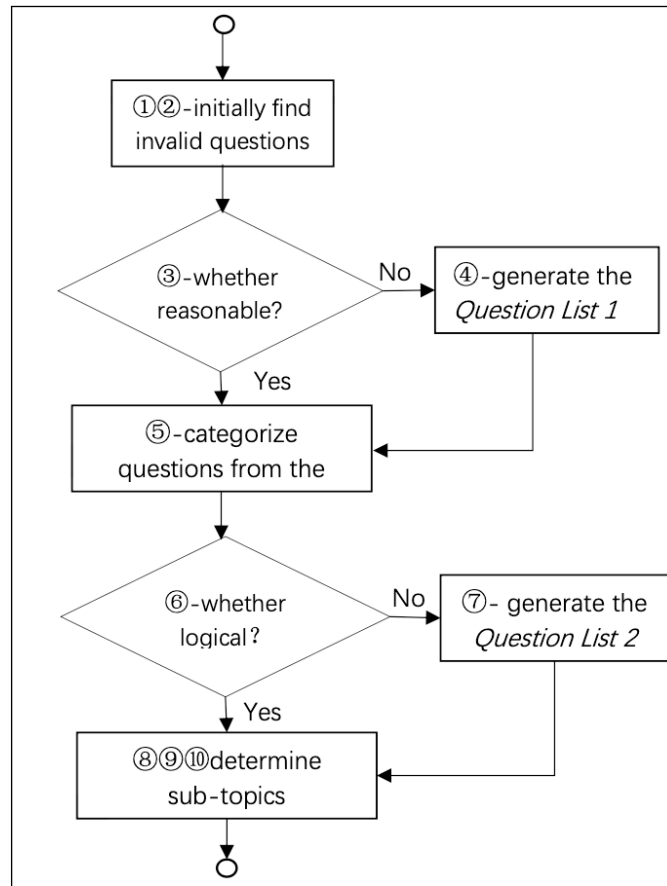
Upon completing the activity, students will be able to:

- **Enhance critical thinking skills:** make attempts at applying logical analysis to identify and prioritize relevant issues, and categorize or generalize problems to extract common features and build discussion frameworks.
- **Use GenAI tools effectively:** gain experience in employing AI platforms (e.g., Kimi, DeepSeek, Doubao, Wenxin Yiyan) to support course-related tasks, while critically evaluating AI output.
- **Improve teamwork and decision-making:** engage in structured group reasoning, balancing human–human and human–AI collaboration dynamics.

The activity is structured into three sequential phases, each integrating GenAI at key points (see Figure 2). Phase 1 (Steps 1–4) focuses on logically analyzing raw material. Phase 2 (Steps 5–7) involves classifying the identified problems using a chosen framework. Phase 3 (Steps 8–10) centers on synthesizing these categories into coherent seminar topics. In each phase, students first discuss as a group to reach consensus on their current output, then submit their work to a GenAI tool for evaluation and feedback. This iterative human–human / human–AI workflow allows teams to collectively assess the rationality and coherence of each step’s decisions, deepening their critical thinking at every stage.

Figure 2

The teaching activity based on GenAI-CT method.



(1) Step 1: Identify Flaws in the Raw Question Set. Students download a “Question List” of ~150 questions related to requirements analysis. In a short time, teams skim the list to spot duplicate items, unclear phrasing, or irrelevant questions (for example, similar entries like “What is requirements analysis?” versus “What does a requirements analysis include?”). This step intentionally uses imperfect materials to challenge students’ initial assumptions and surface superficial reasoning. Confronting a flawed question set motivates them to actively question and scrutinize content rather than accept it at face value. Table 1 shows examples of questions:

Table 1. Examples of questions from Raw Question Set.

#	Question	#	Question
1	What information should a requirements analysis include?	2	What processes does a requirements analysis consist of?
3	At what stage should a requirements analysis be modified?	4	What is the impact of requirements analysis on the final outcome?
5	What factors should be considered when establishing a requirements analysis?	6	What specific functions does our design need to achieve?
7	What is requirements analysis?	8	How should a requirements analysis be written?
9	What is the importance of project requirements analysis?	10	What steps need to be taken for project requirements analysis?
11	What is an IoT application?	12	What is requirements analysis?
13	What aspects does a requirements analysis cover?	14	What components make up a requirements analysis?

(2) Step 2: Filter Out Invalid Questions. Teams perform a logical analysis to remove invalid items from the question set. For each question, they decide whether it is irrelevant or redundant and document their reasoning. For instance, if two questions partially overlap, the team might retain one and delete the other; if a question is off-topic, it is discarded. This phase concretely trains critical thinking: students apply classification and causal reasoning, label flaws, and justify each exclusion. By turning abstract “critical thinking” into tangible actions (like generating a deletion/retention table, shown as in Table 2), students practice structured reasoning that yields clear, defensible outputs.

Table 2. *Rationale and Recommendations for Item Removal/Retention.*

Marked Items	Reasons	Suggestions
1, 5, 13	partial overlap	retain 13
2, 10	partial overlap	retain 2
6	unclear phrasing	delete
7, 12	complete overlap	retain 7
11	Irrelevant to the topic	delete

(3) Step 3: Consult GenAI for Feedback on Screening. Teams select a GenAI tool (such as Kimi, DeepSeek, Doubao, or Wenxin Yiyan) and provide it with three inputs: (i) a description of their procedure so far (Steps 1–2), (ii) the original problem list, and (iii) their current filtered list. They then prompt the AI to evaluate the rationality of their screening decisions and to suggest optimizations. For example, the AI might highlight that “How to write a requirements analysis?” is too vague or suggest merging overlapping questions about components of requirements. At this stage, GenAI acts as an objective advisor: it surfaces logical loopholes and offers novel perspectives that break students’ thinking stereotypes. Students must critically compare these AI suggestions with their own reasoning. This interaction forces them to dialectically evaluate algorithmic feedback rather than follow it blindly. It also partly compensates for instructors’ limited time by giving teams immediate external feedback on their logic. Table 3 shows sample outputs by GenAI:

Table 3. *Sample outputs by GenAI (Kimi, DeepSeek and Doubao).*

[Kimi] Questions Requiring Further Consideration:

- Questions 8 and 9: Although Question 8 ("How to write a requirement analysis?") and Question 9 ("The importance of project requirement analysis?") are not identical, Question 8 is too vague and needs to specify the standards, format, or key elements of "writing". Question 9 should also be more specific, such as pointing out the importance of requirement analysis for project success, cost control, and time management.
- Question 14: It overlaps with Question 13 to some extent. The focus of the two can be distinguished. For example, Question 13 can focus on the aspects covered by requirement analysis, while Question 14 can focus on the content structure.
- Question 4: The wording is inaccurate and can be rephrased as "What is the impact of requirement analysis on the final outcome?" for clarity.

[DeepSeek] Potential Oversights & Refinements:

- Q14 Should Also Be Treated as Redundant: Q14 ("What content primarily constitutes requirements analysis?") is semantically identical to Q1/Q5/Q13 (all asking about components/content). To fully eliminate overlap, delete Q14 or retain it as the sole representative of this category.

[Doubao] Omission -- Removable Questions:

- Question 8: The question "How to write requirement analysis" is related to requirement analysis, but it belongs to the execution-level question of "Writing requirement analysis documents". If the course first focuses on basic concepts like "What is requirement analysis and what does it include", it can be retained temporarily; if strictly screening basic questions, it can be classified as a "process/execution-type" question for subsequent refinement. It is not deleted for now but its priority needs to be considered.

(4) Step 4: Evaluate AI Feedback and Finalize Question List 1. After reviewing the AI's comments, students hold a structured discussion to decide which suggestions to adopt. They debate questions like "Why should this question be retained or removed?" and finalize a consensus list of valid questions (Question List 1). For example, a group might end up keeping seven core questions out of the original set. This phase emphasizes that students must *validate* AI input: they learn to accept useful advice and reject irrelevant or incorrect AI output. By justifying their decisions to peers (and responding to AI-generated advice), students externalize their reasoning and recognize blind spots. The resulting consensus codifies clear criteria for what makes a question valid, moving the group from vague intuitions to explicit, shared standards.

(5) Step 5: Choose a Classification Framework. Teams select a method to organize the questions in List 1. Recommended strategies include keyword-driven clustering or the 5W1H heuristic (Who, What, When, Where, Why, How). Instructors demonstrate these methods, but students apply the approach of their choice. (Groups unfamiliar with such strategies are encouraged to start with keyword clustering.) This phase balances autonomy and guidance: students structure fragmented questions into a logical matrix, building a conceptual scaffold for system knowledge. Importantly, it reinforces a core principle of critical thinking: effective reasoning relies on systematic, intentional frameworks, not ad hoc sorting. By choosing and applying a framework, students learn how deliberate organization enhances clarity.

(6) Step 6: Consult GenAI for Classification Feedback. Students again use their GenAI tool, submitting: (1) the classification task instructions from Step 5, (2) Question List 1, and (3) their proposed categorization of these questions. They ask the AI to check the logical consistency of their classification and to suggest improvements (for example, merging similar categories or adjusting category boundaries). The AI's evaluation reveals implicit connections among the problems. When the AI proposes, say, combining two subcategories into one or reassigning questions, students must re-examine the rigor of their framework. This step sharpens their ability to scrutinize classification logic: they learn to treat AI recommendations as prompts for reflection. For Table 4, the included content is drawn from the learning journals that students submitted, presenting a case of student-AI interactions during the logical classification phase. As the original journals were composed in Chinese, a translation into English was conducted to facilitate analysis.

Table 4. *An example of student-AI interactions during the logical classification phase.*

Student: This is my RA question list, which includes various questions raised during my requirement analysis and design of the IoT application system. Now, I will classify these questions using the 5W1H method. Here is my classification: **Who** -- questions 2, 3, 4, 5, 81, 89, 98, 112; **What** -- questions 3, 14, 24, 30, 37, 39, 46, 51, 54, 140, 144, 145; **When** -- questions 38, 137, 139; **Where** -- questions 131, 138; **Why** -- 9, 142; **How** -- questions 1, 9, 18, 19, 20, 21, 27, 28, 40, 43, 44, 52, 57, 65, 67, 79, 86, 90, 91, 94, 103, 107, 109, 115, 126. How is my classification according to the 5W1H method? What's your opinion?

Doubao: Question Classification Optimization

- **Multi-Dimension Intersection Issues:** For questions involving multiple dimensions, they can be repeatedly listed in the table or described in the notes to clarify their multi-dimensional nature.
 - Example: Question 9 ("How to balance function requirements and budget reduction") can appear in both the "What" and "Why" dimensions.
 - Example: Question 38 ("Can the system be launched under poor information conditions") can appear in both the "When" and "Where" dimensions.

(7) Step 7: Reconcile AI Feedback and Form Question List 2. Building on the AI’s input, teams hold another discussion to align human reasoning with AI insights. They map each question into predefined categories, producing a final Question List 2. For example, using the 5W1H framework, each question is assigned to one of the six dimensions (Who, What, etc.), filling tables under each category. This iterative process (student ↔ student ↔ AI) requires students to defend their logic to peers while addressing AI critiques. It embodies the idea of AI as a “cognitive disruptor”: by pointing out overlaps or inconsistencies (such as a question fitting multiple categories), AI pushes the group to refine their framework. Through this three-way interaction, students expose inconsistencies in their thinking, which builds decision-making resilience stronger than standard group work alone.

(8) Step 8: Synthesize Subtopics from Classified Questions. With the categorized questions in hand, teams engage in structured discussion to distill 3–5 subtopics. They integrate related questions under meaningful headings. For instance, a group might define subtopics such as “Concepts and Significance of Requirements Analysis” (covering questions on definitions and importance), “Tasks and Content of Requirements Analysis”, and “Process of Requirements Analysis”. This step shifts students from “problem deconstructors” to “knowledge constructors”: they abstract scattered details into broader themes. It trains information integration and high-order thinking, reinforcing categorization as a cognitive skill. By naming subtopics, students transform their classified data into a structured outline, a crucial step in generating insightful solutions.

(9) Step 9: Consult GenAI for Subtopic Feedback. Teams submit to the AI: (1) the requirements/guidelines for generating subtopics, (2) Question List 2, and (3) their draft subtopic headings. They ask the AI to evaluate how well each heading summarizes its underlying questions and to suggest refinements. The AI provides exemplar answers. For example, if students propose “Needs of different users,” the AI might respond with a more precise topic like “Multidimensional Analysis of User Requirement Characteristics”. Comparing student proposals with AI’s systematic formulations helps learners recognize the gap between intuitive labeling and academic abstraction. This feedback guides them to more precise, comprehensive phrasing, effectively bridging from concrete examples to abstract concepts.

(10) Step 10: Finalize Seminar Topics. Finally, students review the AI’s suggestions and discuss to finalize the seminar content. They reach consensus on which AI-optimized headings to adopt and ensure each subtopic accurately reflects its mapped questions. At the end of this step, the seminar topics (with corresponding question numbers) are set. This concluding phase emphasizes systematic reflection and clarity: students judge whether the final headings cover the essence of the classified questions (in terms of scope and accuracy) and incorporate AI’s improvements where appropriate. The process brings their thinking from vague, initial ideas to precise, well-defined solutions, achieving the ultimate goal of advanced critical reasoning.

Significance of the GenAI-CT Design

This ten-step activity is valuable in multiple ways:

- **Cultivating critical thinking:** The design systematically builds students’ analytical skills. During screening and verification (Steps 2–4), students learn to evaluate information, distill key details, identify semantic and logical errors, and make reasoned judgments. Through classification tasks (Steps 5–7), they develop systematic analysis and evidence-based decision-making. In abstraction tasks (Steps 8–10), they practice generalizing ideas and articulating them precisely. Throughout, continuous reflection is encouraged: by debating with peers and AI, students become more aware of their own reasoning and learning strategies, promoting autonomous and effective learning.
- **Bridging traditional and AI-enhanced learning:** The GenAI-CT model blends classic pedagogical methods with modern AI tools. Paired phases (e.g., Step 2 with 3, Step 5 with 6, Step 8 with 9) show how logical analysis can be coherently integrated with GenAI assistance.

This hybrid approach prepares students for an AI-rich future: they gain confidence using AI as a cognitive partner without losing core analytical skills. It also helps instructors gradually adapt: they can leverage GenAI within familiar teaching sequences.

- **Fostering collaboration and human-AI synergy:** The pedagogy enriches communication on two fronts. First, students continuously interact with each other, jointly debating and deciding at every phase. Second, the approach emphasizes a collaborative stance toward AI. In Steps 3 and 6, for example, students learn to view GenAI as a support tool, not a replacement for thought. In Steps 4 and 7, they practice persuading peers of their reasoning and critiquing AI insights. This three-dimensional interaction (student–student–AI) hones decision-making resilience more effectively than traditional group discussions alone. By the end, students see the value of human–AI teaming, a critical capability for the AI-driven STEM landscape.

In summary, the GenAI-CT teaching design not only addresses the specific challenges of the IoT course but also demonstrates a general model for STEM education in the AI era. It encourages active critical inquiry, effective use of AI tools, and robust peer collaboration—essential skills for developing innovative, independent thinkers.

4 Results

The GenAI-CT teaching model was piloted in a spring 2025 course with 69 students (organized into 12 teams). A mixed-methods approach was used to evaluate GenAI’s role in fostering critical thinking and teamwork. Quantitative measures (e.g., performance on critical thinking tasks) were combined with qualitative evidence (student reflection logs, group discussion records) to capture both skill development and experiential insights. Notably, students completed the full ten-step GenAI-CT process without any instructor intervention to ensure an authentic assessment of their current critical thinking abilities and GenAI utilization. Despite the limited sample size, clear patterns emerged regarding how GenAI integration impacted students’ critical thinking competencies and team collaboration. The results are organized below by these two focal areas.

GenAI’s Impact on Students’ Critical Thinking Development

Quantitative Outcomes: Students’ critical thinking performance varied notably across different cognitive skill dimensions (logical analysis, logical classification, and topic condensation). Table 5 summarizes the assessment results for each phase of the GenAI-CT process (Steps 2–4 focused on logical analysis, Steps 5–7 on logical classification, and Steps 8–10 on topic condensation). The majority of students performed well in the logical analysis stage: 75% achieved a “good” performance, demonstrating they could identify most irrelevant or duplicate problems, while none achieved a perfect screening of all issues. In contrast, logical classification proved challenging—over 90% of students only reached an “average” level, with significant overlaps or gaps in their categorized problem frameworks. Not a single group produced a fully clear and mutually exclusive classification scheme (ideal criteria), and 8.3% failed to organize the problems logically at all. Topic condensation outcomes were more encouraging: 41.7% of students achieved an “excellent” level (highly concise topics covering all core problem directions) and an additional 33.3% at a “good” level. Still, about a quarter of the students struggled with this synthesis task (16.7% showed weak relevance in their condensed topics, and 8.3% produced off-target or no meaningful summary).

Table 5. *Critical Thinking Performance by Phase (Quantitative Assessment).*

Critical Thinking Phase (Step Range)	Performance Criterion (Quality of Outcome)	% of Students Achieving Criterion
Logical Analysis (Steps 2–4)	Precise filtering of problems—identified all duplicates/irrelevancies; retained 100% of core relevant problems.	0%
	Minor omissions (<30%)—mostly correct elimination with only slight misses; core problems reasonably screened.	75%
	Major omissions (>30%)—notable gaps or some retained problems deviating from the theme.	16.7%
	Chaotic or no filtering—criteria unclear, many irrelevancies kept; or no participation in this phase.	8.3%
Logical Classification (Steps 5–7)	Clear & systematic grouping—categories well-defined, mutually exclusive, covering all dimensions of the problem set.	0%
	Mostly reasonable grouping—some overlaps or omissions (<30%) in categories.	0%
	Vague or flawed grouping—obvious category overlaps, logical loopholes (>30% gaps/overlaps).	91.7%
	Chaotic or no grouping—unable to reflect problem relevance in any structured way; or no participation.	8.3%
Topic Condensation (Steps 8–10)	Excellent synthesis—subtopics highly concise, capturing all core problem themes.	41.7%
	Good synthesis—subtopics cover main directions with only minor redundancies.	33.3%
	Weak synthesis—subtopics have weak relevance to problems or omit key issues.	16.7%
	Off-target or none—subtopics detached from problem list; or no participation in summarizing.	8.3%

Note. Each phase corresponds to a segment of the GenAI-CT process conducted without instructor help. Logical Analysis required teams to scrutinize and filter an initial problem list; Logical Classification required organizing problems into a coherent framework; Topic Condensation required distilling the essence into concise themes.

The above results indicate that GenAI’s assistance had uneven effects on different critical thinking skills. GenAI appeared to support students in simpler analytical tasks (e.g., identifying irrelevant problems) but was less effective in aiding complex structuring tasks like categorization. In logical analysis, even without GenAI, most science and engineering students can adequately eliminate invalid problems (a skill they likely possess from prior training). On the other hand, the logical classification stage demands more systematic integration of information, an area where students struggled despite using GenAI. In fact, the data suggest GenAI only offered limited improvement for classification: virtually no group achieved a fully sound classification, and the vast majority produced flawed category schemes. Two underlying causes were identified from qualitative evidence. First, many students lacked sufficient training in structured thinking; their understanding and application of logical classification methods were unsystematic and shallow. Second, students did not know how to effectively leverage GenAI as a tool—analysis of the GenAI query logs showed that many prompts failed to convey the student’s true intent, leading to GenAI outputs misaligned with what students actually needed. As a result, students perceived GenAI’s utility in this phase to be quite limited.

Notably, topic condensation (synthesizing subtopics) was the one area with relatively high performance, yet the qualitative data revealed a cautionary insight. Several groups’ final answers exhibited signs of “pseudo-excellence,” where the polished responses did not fully reflect the students’

own thinking. For instance, at least four different teams submitted an identical subtopic titled “Main Users and Stakeholders,” a phrase that originated from a GenAI suggestion rather than the students’ initial ideas. Those groups had originally brainstormed titles like “Involved Main User Groups” or “User Roles,” but after seeing GenAI’s output, they assumed the AI-generated title was more “professional” and adopted it without further scrutiny, discarding their original conclusions. This behavior suggests that some students may overly defer to GenAI in complex tasks, substituting AI-generated content for their own reasoning. In such cases, GenAI’s positive impact on outcome quality (e.g., producing a concise, professional-sounding summary) comes at the expense of students’ independent critical thinking, as they trust the tool’s answer over their own judgment.

Qualitative Insights—Reflective Awareness and Inquiry: The extent to which students critically evaluated GenAI’s contributions was limited. Table 6 summarizes two specific metacognitive behaviors observed: reflective awareness (recognizing and questioning contradictions between GenAI outputs and the team’s reasoning) and active inquiry (deliberately probing or testing GenAI for explanations and reliability). These behaviors were expected at key checkpoints in the process (after initial problem screening, after classification, and after topic synthesis—corresponding to Steps 3, 6, and 9). However, only a minority of groups displayed such behaviors in practice.

Table 6. *Student Reflective and Inquisitive Behaviors with GenAI.*

Metacognitive Behavior	Description of Expected Behavior	Example Log Entries	Occurrence Rate at Key Steps
Reflective Awareness	Critically examine GenAI’s suggestions: identify contradictions or discrepancies between GenAI output and human reasoning; compare AI’s answers with the team’s own conclusions. Students may partially use GenAI’s ideas but do not treat them as unquestionably correct.	E.g., GenAI suggested merging two issues into one category, but after discussion the team noted each issue had a distinct focus. They chose not to merge them, saying “We decided to keep it because [each question’s focus is different]...The students did not copy GenAI’s classification blindly.”	33.3% of groups at Step 3; 25% at Step 6; 25% at Step 9 showed evidence of this awareness.
Active Inquiry	Proactively engage GenAI in dialogue: ask GenAI to explain its reasoning or justify answers; test GenAI’s reliability by providing deliberate misinformation or edge cases. This behavior reflects deeper curiosity about why GenAI gave certain outputs and how it might fail.	E.g., A student asked GenAI, “Why was this question categorized under ‘Who’ rather than ‘How’?”, seeking an explanation. Another team intentionally input a flawed query to observe how GenAI would respond to an error, thereby testing the AI’s fault tolerance.	16.7% of groups at Step 6 engaged in such inquiry; 0% at Steps 3 and 9 (no teams asked “why” or tested GenAI in those stages).

Overall, these findings show that relatively few students adopted a critical or probing stance toward GenAI outputs. Only about one-third of teams demonstrated reflective awareness by double-checking GenAI’s suggestions against their own reasoning, and this occurred mostly in the early phase (Step 3) with declining occurrence in later phases. Even fewer students engaged in active inquiry—for example, hardly any teams asked GenAI to justify its classifications or purposely challenged the AI’s answers to gauge its reliability. This lack of sustained skepticism is notable given that current GenAI tools can produce inconsistent or erroneous information. In fact, different GenAI platforms sometimes gave divergent answers to the same question, and occasionally GenAI introduced incorrect

or biased viewpoints. When students fail to question such outputs, there is a risk they will accept false or suboptimal information as truth. The observations suggest that without instructor prompts or prior training in critical interrogation of AI, many students tended to accept GenAI's output at face value. This could impair their objective judgment and decision-making over time. Thus, while GenAI provided useful content and suggestions, its presence did not automatically instill stronger critical thinking habits in students—in fact, some students became over-reliant on GenAI's answers, underscoring the need for explicit guidance on reflecting and questioning AI-generated information.

GenAI's Role in Enhancing Team Collaboration

Baseline Collaboration Attitudes: To contextualize GenAI's impact on teamwork, students' initial attitudes toward group work and collaboration challenges were surveyed before the course (see Appendix). Results indicated a mixed enthusiasm for team-based learning. Only about 52.6% of students reported being generally willing to participate in group work, while roughly 24.8% admitted they did not particularly enjoy group projects (but would cooperate if required). Moreover, nearly 44.6% of the class preferred to act as regular team members rather than take on leadership roles, suggesting a reluctance among many to lead discussions. Students also identified several pain points that hinder effective collaboration: about 45.5% cited social anxiety or shyness in group interactions; 44.6% were frustrated by unproductive disputes during team discussions; nearly half (49.5%) noted a lack of proper guidance or direction in group tasks; and 47.5% felt they often defaulted to independent learning habits, finding it hard to truly collaborate. Despite these challenges, students were largely optimistic about using GenAI as a collaborative aid—over 90% supported integrating GenAI into group work (with 26.6% expressing strong support). They did voice some reservations: for example, 67.9% felt GenAI's responses might be overly formulaic or “too standardized,” 46.2% worried about technical issues or AI instability, and 43.2% pointed out that an AI lacks human-like emotional understanding, which could limit its usefulness in teamwork.

Observed Effects of GenAI on Teamwork: During the course, GenAI tools were introduced at specific stages of the team project to help mitigate the above difficulties and to observe GenAI's actual influence on team collaboration. Students were encouraged to use GenAI as needed when they encountered hurdles (e.g., confusion, deadlock in discussion, lack of ideas). Analysis of students' collaboration logs and instructors' in-class observations indicates that GenAI had several positive effects on team dynamics, especially in scenarios of delayed feedback, uneven participation, and prolonged disputes. Table 7 highlights these key roles of GenAI in the group work context, and subsequent paragraphs illustrate each with examples.

Table 7. *Common Collaboration Challenges and GenAI's Supportive Roles.*

Team Challenge	Typical Issue in Student Teams	GenAI's Support Role
Delayed Feedback	Teachers, managing many teams, cannot always provide help at each step in real-time. Groups often stalled at critical points waiting for instructor input; unrecognized logical errors could persist and mislead subsequent work.	Teaching Assistant: GenAI filled this gap by automatically checking the team's work for omissions or logical flaws and offering immediate suggestions. For example, GenAI could detect if a key question was accidentally omitted and prompt the team with a revision ("Based on your classification, I suggest that"). By providing on-demand feedback, GenAI helped prevent discussion stagnation. In practice, all 12 teams were able to progress through each step without overrunning time, as GenAI's timely guidance kept them on track.
Uneven Participation	In many groups, a few outspoken members tend to dominate the discussion, while quieter members contribute little (sometimes out of fear their ideas will be rejected or lack of confidence). This creates a "discourse power imbalance," where some voices go unheard and the team's output reflects only a subset of perspectives.	Information Provider: GenAI served as an equalizing force by equipping quieter or less confident students with information and ideas to share. For instance, a student hesitant to speak could query GenAI and introduce a point by saying "According to GenAI, ..." using the AI as a knowledge back-up. This practice lowered the barrier for shy members to join in. According to discussion records, nearly every student ultimately participated in the group dialogues, with each member contributing at least one idea or viewpoint during the project. GenAI's presence thus helped distribute participation more evenly, reducing the executors-vs.-dominators dynamic.
Unproductive Disputes	Teams sometimes got bogged down in arguments over conflicting approaches or frameworks rather than moving forward. For example, one group spent 20 minutes debating whether to apply a "5W1H" vs. a "SWOT" framework, with neither side yielding. Such deadlocks not only waste time but also strain team relationships, especially if debates become personal.	Neutral Mediator: In heated standoffs, GenAI acted as an impartial third-party mediator. Students could input a summary of the dispute (e.g., "Classmate A insists on X, while Classmate B insists on Y") and ask GenAI for an analysis or suggestion. The AI would then provide an objective comparison of the two perspectives or propose a compromise approach. This external viewpoint often helped teams refocus on the problem's essence rather than on who was "right." In the above example, consulting GenAI yielded a balanced analysis that guided the team to combine elements of both 5W1H and SWOT, reaching a consensus more efficiently than if the argument continued unchecked.

Through these roles, GenAI functioned as a valuable collaborative aide. It effectively supplemented the teacher's limited availability by giving real-time feedback, empowered all team members to contribute knowledge, and diffused conflicts with neutral analysis. The net result was that teams stayed on schedule and engaged more members in the problem-solving process than they might have otherwise. Classroom observations and student feedback logs consistently noted these benefits. One instructor's observation was that "nearly all members were actively participating with GenAI's help," and several student groups reported that using the AI "made quieter members more comfortable speaking up" (as recorded in their reflection journals).

However, the introduction of GenAI to team work also surfaced new collaboration risks that warrant careful management. One issue observed was "pseudo-collaboration." In some groups, although every member appeared to be interacting (and even adding content generated by GenAI), the interactions remained superficially coordinated. For example, chat transcripts showed that while all students added some points (including ideas borrowed from GenAI), their contributions were simply listed one after another in the final document without real integration or debate. In such cases, the final "consensus" was not truly co-created through discussion; instead, the team leader just merged inputs arbitrarily to form a report. This indicates that mere use of GenAI does not guarantee deep collaboration—teams might still operate in silos and then stick pieces together, giving an illusion of collaboration. Another concern was a tendency towards over-reliance on GenAI's authority. Some students appeared to treat GenAI's suggestions as inherently correct or more credible than their own ideas, a mindset described in logs as the "let AI decide" mentality. In extreme cases, a few groups deferred to GenAI even when their own reasoning was superior: instructors noted instances where a team's independent solution was more accurate or creative than GenAI's answer, yet the team abandoned their solution in favor of the AI output simply because it came from an "expert" source. Such over-trust in GenAI can undermine students' sense of responsibility and confidence, dampening their willingness to contribute ideas.

These findings underscore that while GenAI can enhance team collaboration by addressing practical challenges (timeliness of feedback, inclusion of all members, conflict resolution), it does not automatically instill effective collaboration habits. In fact, without guidance, students might misuse the AI in ways that reinforce poor practices—for instance, using GenAI to avoid tough discussions or to offload decision-making. In the study, the course instructors mitigated these risks through reflective debriefings: after the project, they provided feedback highlighting where a group's independent ideas had outperformed GenAI and praised those contributions to reinforce human expertise. The instructors also reminded students that GenAI is a support tool rather than an infallible oracle, encouraging a balanced partnership where human critical judgment remains paramount. Such scaffolding will be important in future implementations to ensure GenAI's positive impacts on collaboration are fully realized without eroding students' collaborative ownership and critical thinking.

5 Ethical Considerations

Informed Consent and Participant Rights: This study adhered to strict ethical guidelines to protect all participants. Before the study began, each student received a comprehensive explanation of the research purpose, procedures, potential risks and benefits, and their rights as participants. In particular, they were informed of their right to voluntarily participate and the option to withdraw from the study at any time without penalty or negative consequences. This ensured that students participated with full understanding and autonomy.

Data Privacy and Confidentiality: Robust measures were implemented to safeguard personal data. All information collected from participants was anonymized to remove identifying details, and the coded data were stored securely on the university's servers with access limited to the research team. No participant-identifiable information appears in any reports or publications resulting from this work. By protecting confidentiality and data security in this manner, the study complied with institutional and international standards for research ethics, assuring students that their privacy would be respected.

Responsible Use of GenAI Tools in Learning: In addition to protecting participants, the study emphasized ethical practices in using generative AI (GenAI) as part of the learning process. Both instructors and students were guided to treat GenAI as a supportive tool to augment human thinking rather than a replacement for it. Practically, instructors encouraged students to use AI-generated content as a springboard for ideas and problem-solving, while always subjecting those outputs to critical scrutiny and personal verification. This approach fostered ethical awareness by helping students maintain academic integrity—for example, avoiding uncritical acceptance or plagiarism of AI-generated answers—and ensuring that reliance on GenAI did not erode their own independent reasoning. Instructors modeled this behavior by highlighting potential biases or errors in AI-provided information and by setting clear guidelines on when and how to integrate AI assistance into coursework. By cultivating such habits, the use of GenAI tools in the classroom was kept transparent and accountable, reinforcing principles of honesty, accountability, and intellectual integrity for both students and educators.

6 Conclusion

This exploratory case study found that integrating GenAI into an undergraduate STEM course is feasible and can yield notable benefits for student learning, but it also uncovered important challenges that warrant careful attention. Key findings indicate that generative AI can support the development of critical thinking in specific areas: for instance, many students showed strong performance in tasks like problem analysis and idea synthesis when aided by GenAI, suggesting that AI tools can help scaffold higher-order thinking. The introduction of GenAI into group work was likewise well received—most students reported that AI support enriched their team discussions by providing immediate feedback and alternative perspectives, which in turn helped keep collaboration on track. At the same time, the results revealed limitations in students' critical thinking that GenAI alone could not overcome. Over 90% of participants struggled with systematically classifying and organizing complex problems even with AI suggestions, indicating that fundamental skills in structured reasoning still need to be directly taught and practiced. Moreover, some students exhibited a tendency to over-rely on GenAI outputs without sufficient critical evaluation. In several cases, groups adopted AI-generated answers or frameworks at face value—a “let AI decide” mentality—instead of debating or verifying these suggestions against their own insights. Notably, only a minority of students actively questioned the AI's conclusions or sought to test the AI's reasoning during the activities. These patterns underscore the continued necessity of teacher guidance and critical thinking instruction alongside technological tools. In our implementation, instructors had to intervene at times to remind students to reflect on AI-proposed solutions and to validate the students' own reasoning when it was sound. Overall, while GenAI proved to be a valuable assistant that can enhance learning—by quickly answering queries, offering diverse viewpoints, and mediating group challenges—it is not a substitute for human critical inquiry. Effective integration of GenAI in education requires balancing AI assistance with deliberate cultivation of students' skepticism, reflection, and independent problem-solving skills.

Limitations: Several limitations of this study should be acknowledged. First, the sample size (69 students in one course at a single university) was relatively small, and the intervention was conducted over one academic semester. These factors limit the generalizability of the findings—the results may not directly translate to other contexts, disciplines, or longer-term implementations. Second, as an initial exploratory study, there was no long-term follow-up to determine whether improvements in critical thinking or collaboration were sustained beyond the course or whether students continued to apply ethical AI usage principles later on. There is also the possibility of selection bias or novelty effects; students' enthusiasm for GenAI might reflect the excitement of trying a new tool rather than a lasting change in their learning approach. Given these constraints, the conclusions drawn here should be seen as indicative rather than definitive. Future studies with larger and more diverse samples, as well as longer observation periods, will be valuable to reinforce and expand upon these insights.

Future Research Directions: This research opens up several avenues for further investigation. One important direction is to replicate and extend the GenAI-CT model in different educational settings—including other universities, disciplines, or age groups—to test its effectiveness and adaptability at scale. Such studies could provide comparative data on how students from various backgrounds engage with GenAI in learning, and what adjustments might be needed for different subject matters. Another vital area is the development of targeted training programs for both students and instructors on AI literacy and critical usage skills. Our findings suggest that explicit instruction in how to formulate good prompts, recognize AI limitations, and critically evaluate AI-generated outputs could significantly enhance the benefits of GenAI tools. Future experiments could measure whether providing prompt engineering workshops or “AI skepticism” exercises (such as those that have students practice catching AI mistakes) leads to deeper critical engagement and reduces over-reliance on AI. Additionally, long-term and longitudinal research is recommended to observe how continuous exposure to GenAI throughout a curriculum impacts students’ critical thinking growth, problem-solving independence, and teamwork dynamics over time. It would be insightful to track whether early training with GenAI translates into improved performance in advanced courses or in professional settings. From the educators’ perspective, future work should also explore the evolving role of instructors in AI-augmented learning environments. As generative AI takes on functions like providing instant feedback or alternative solutions, instructors may need to shift more into the role of facilitators and ethics coaches. Research on effective professional development strategies can help teachers learn how to guide students in the ethical use of AI, design assignments that incorporate AI appropriately, and intervene when AI use might hinder learning. In summary, continued inquiry in these areas will help build a more comprehensive understanding of how to harness GenAI’s capabilities to enrich education while steadfastly promoting the cultivation of independent critical thinking and ethical reasoning in the AI era.

References

- Ahern, A., Dominguez, C., McNally, C., O’Sullivan, J. J., & Pedrosa, D. (2019). A literature review of critical thinking in engineering education. *Studies in Higher Education*, 44(5), 816–828. <https://doi.org/10.1080/03075079.2019.1586325>
- Wu, H. (2023). Critical thinking in the era of automation and artificial intelligence. *Journal of Yan’an University (Social Sciences Edition)*, 45(6), 27–39. <https://doi.org/10.3969/j.issn.1004-9975.2023.06.005>
- Dasari, D., Hendriyanto, A., Sahara, S., Suryadi, D., Muhaimin, L. H., Chao, T., & Fitriana, L. (2024). ChatGPT in didactical tetrahedron, does it make an exception: A case study in mathematics teaching and learning. *Frontiers in Education*, 8, 1295413. <https://doi.org/10.3389/educ.2023.1295413>
- Chanpradit, T. (2025). Generative artificial intelligence in academic writing in higher education: A systematic review. *Education and Information Technologies*.
- UNESCO. (2023). *Guidance for generative AI in education and research*. Paris: UNESCO.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1–38. <https://doi.org/10.1145/3571730>
- Zhang, X., Zhang, P., Shen, Y., Lu, X., Li, Q., & Tao, D. (2024). A systematic literature review of empirical research on applying generative artificial intelligence in education. *Frontiers in Digital Education*, 1(3), 223–245.
- Salameh, W. (2024). The integration of artificial intelligence (AI) into undergraduate education. *International Journal of Science and Research Archive*, 13(1), 2455–2463. <https://doi.org/10.30574/ijrsra.2024.13.1.1960>

- De Paoli, S. (2023). Performing an inductive thematic analysis of semi-structured interviews with a large language model: An exploration and provocation on the limits of the approach. *Social Science Computer Review*, 42(4), 997–1019. <https://doi.org/10.1177/08944393231220483>
- Alghamdi, R. (2024). Exploring the impact of ChatGPT-generated feedback on technical writing skills of computing students: A blinded study. *Education and Information Technologies*, 29(14), 18765–18790. <https://doi.org/10.1007/s10639-024-12594-2>
- Hu, Y., Goktas, Y., Yellamati, D. D., & De Tassigny, C. (2024). The use and misuse of pre-trained generative large language models in reliability engineering. In *Proceedings of 2024 Annual Reliability and Maintainability Symposium* (pp. 1–7). IEEE.
- Khan, R., Bhaduri, S., Mackenzie, T., Paul, A., Sankalp, K. J., & Chakrabarti, A. (2024). Path to personalization: A systematic review of GenAI in engineering education. *KDD AI4Edu Workshop*, ACM. hal-04673700.
- Shailendra, S., Kadel, R., & Sharma, A. (2024). Framework for adoption of generative artificial intelligence (GenAI) in education. *IEEE Transactions on Education*. <https://doi.org/10.1109/TE.2024.3432101>
- Qu, Y., Tan, M. X. Y., & Wang, J. (2024). Disciplinary differences in undergraduate students' engagement with generative artificial intelligence. *Smart Learning Environments*, 11(1), 51–69. <https://doi.org/10.1186/s40561-024-00334-7>
- Wu, Y. (2024). Critical thinking pedagogics design in an era of ChatGPT and other AI tools—Shifting from teaching “what” to teaching “why” and “how.” *Journal of Education and Development*, 8(1), 1–12.
- Hong, H., & Guo, J. (2024). Cultivation of critical thinking skills: Exploring the impact of generative artificial intelligence-enabled instruction in English essay writing. *Journal of Contemporary Educational Research*, 8(8), 226–232.
- Hong, H., Vate-U-Lan, P., & Viriyavejakul, C. (2025). Perceptions of the generative AI-enabled cognitive offload instruction in English writing. *International Journal of Evaluation and Research in Education (IJERE)*, 14(3), 1761–1769.
- Ruiz-Rojas, L. I., Salvador-Ullauri, L., & Acosta-Vargas, P. (2024). Collaborative working and critical thinking: Adoption of generative artificial intelligence tools in higher education. *Sustainability*, 16, 5367. <https://doi.org/10.3390/su16135367>
- Lee, C. C., & Low, M. Y. H. (2024). Using genAI in education: The case for critical thinking. *Frontiers in Artificial Intelligence*, 7, 1452131. <https://doi.org/10.3389/frai.2024.1452131>
- Marchena Sekli, Giulio, F. C. A., Godo, A., et al. (2024). Generative AI solutions for faculty and students: A review of literature and roadmap for future research. *Journal of Information Technology Education: Research*, 23, 14.
- Siraj, M., Duke, J., & Pltz, T. (2025). The GenAI generation: Student views of awareness, preparedness, and concern. [*Journal to be determined*].
- Zhao, X., Chen, X., Huang, V., et al. (2025). Students' use and attitudes toward generative artificial intelligence: A comparative study between the UK and China. In *Proceedings of the 58th Hawaii International Conference on System Sciences* (pp. 4933–4940).

Yueyue Li is affiliated with the School of Communication and Information Engineering at the University of Electronic Science and Technology of China in Chengdu, Sichuan, China. ORCID: 0009-0009-7263-7952.

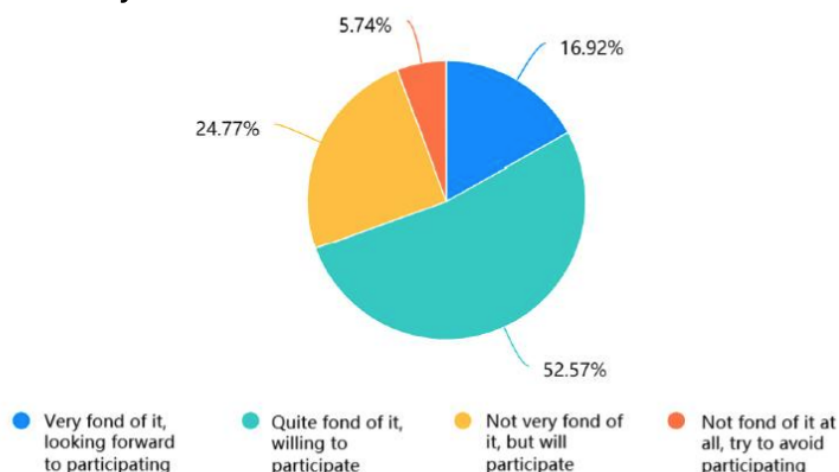
Jie Wu is a faculty member in the College of Computer Science at Chengdu University in Chengdu, Sichuan, China.

Zheng Yang is an Associate Professor of Mathematics at Sichuan University. His research spans commutative algebra, homological algebra, and mathematics education. He teaches calculus, linear algebra, and statistics and is active in interdisciplinary teaching and learning collaborations.

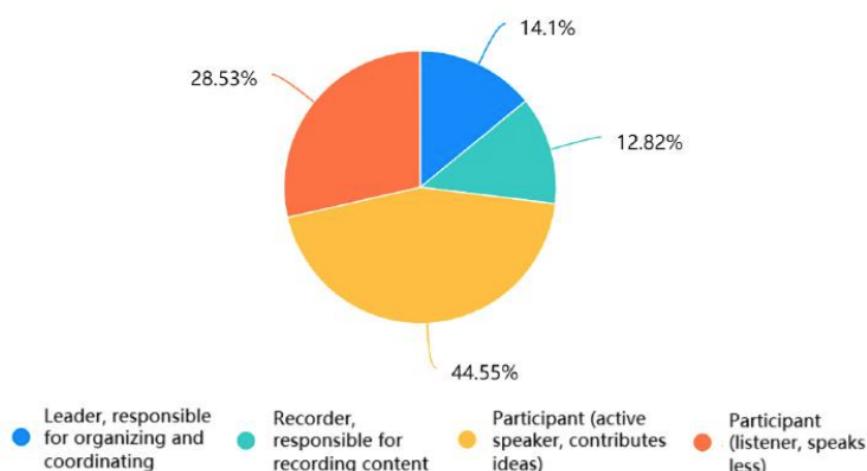
Appendix. A Study on Group Collaboration and attitudes toward GenAI

An investigation on Tem Collaboration-Based Learning was carried out at Chengdu University and Sichuan University of Media and Communications. The research subjects cover multiple grades and majors, with a total of 331 valid samples. In terms of grade distribution, freshmen (first - year undergraduates) are in the majority, accounting for 64.35%. Followed by sophomores (second - year undergraduates) at 21.45% and juniors (third - year undergraduates) at 13.6%. Senior undergraduates (fourth - year undergraduates) and postgraduates have relatively low participation. The gender ratio is relatively balanced, with males accounting for 48.34% and females 51.66%. In terms of academic discipline backgrounds, participants are mostly from engineering and technology - related majors (52.57%) and economics, management, law and business - related majors (39.27%). Additionally, a small number come from majors such as natural science, humanities and social sciences, as well as art and design.

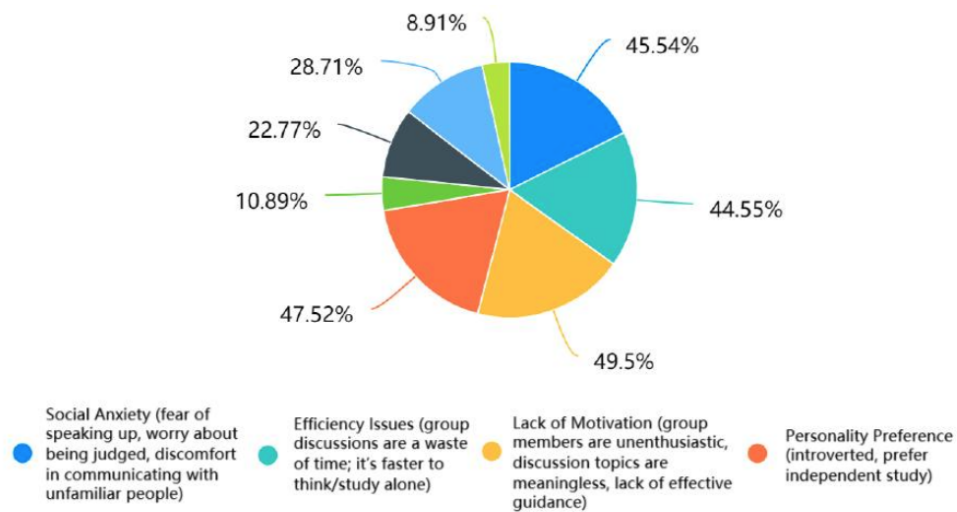
Question: What is your attitude towards team collaboration in the classroom?



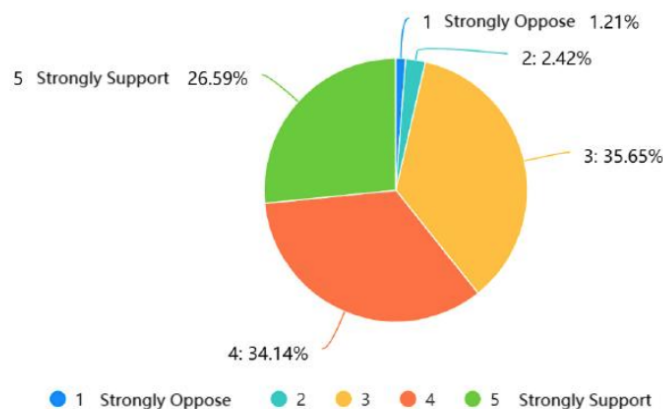
Question: What role do you tend to play in team collaboration?



Question: What is the main reason you are unwilling to participate in group discussions?



Question: To what extent do you support the use of AI as a "partner" in group tasks?



Question: To what extent do you support the use of AI as a "partner" in group tasks?

