Developing Critical AI Literacy Through Mathematical Inquiry

Undergraduate Students Investigating the Collatz Conjecture
with Generative AI Partners

Zheng Yang, Michael Todd Edwards, and Carlos Augusto Lopez Gonzalez
Technology Educator Alliance

December 2025

Abstract

This chapter examines how undergraduate mathematics students develop critical AI literacy while investigating the Collatz conjecture through collaboration with generative AI. Grounded in critical AI literacy theory (Lee et al., 2021), cognitive offloading frameworks (Risko & Gilbert, 2016), and hybrid intelligence principles (Dellermann et al., 2019), we designed a four-phase inquiry that positions students as authentic research mathematicians: developing conjectures, testing them computationally, and using AI as a collaborative tool rather than an answer-generating oracle. By exploring parameter variants before the Collatz conjecture is formally named, students encounter genuine mathematical uncertainty requiring verification of computational claims, critical interrogation of AI outputs, and strategic decisions about cognitive offloading. Analysis of student transcripts, reflections, and artifacts reveals how this approach cultivates habits of mind characteristic of mathematical practice—emphasizing verification over acceptance, interrogation over consumption, and human-AI partnership over algorithmic dependence.

1 Introduction: The Critical AI Literacy Challenge

The integration of generative AI into mathematics classrooms presents a pedagogical tension: while LLMs offer unprecedented computational power, uncritical reliance risks displacing the mathematical reasoning we aim to cultivate (Long & Magerko, 2020). This tension is especially pressing in inquiry-based learning, where conjecture, testing, and refutation form the heart of mathematical practice. If students treat AI as an authoritative oracle providing answers rather than a tool requiring verification, we undermine the reasoning habits central to mathematical thinking.

We argue this tension can be effectively addressed through intentional design grounded in three complementary frameworks. **Critical AI literacy** (CAL) provides foundation for developing students' capacity to "think beyond the output" (Lee et al., 2021)—treating AI as a computational resource whose contributions demand verification, questioning, and healthy skepticism. **Cognitive offloading theory** (Risko & Gilbert, 2016) explains *why* strategic human-AI partnerships enhance rather than diminish reasoning: by deliberately offloading routine computation, students free cognitive resources for higher-order thinking.

Hybrid intelligence frameworks (Dellermann et al., 2019) characterize effective collaboration as partnership built on complementary strengths—human intuition paired with machine computation—rather than replacement or dependency.

The Collatz conjecture offers ideal context for putting these ideas into practice. The problem is conceptually accessible—iterate a simple rule until reaching 1—yet computationally intensive and mathematically unsolved. Unlike problems such as factoring or limits where solution paths exist, Collatz has no known proof, which forces students to rely on verification and questioning rather than correctness-checking against established methods. This combination creates authentic uncertainty that demands critical engagement rather than answer-seeking. Students cannot simply accept AI's results as truth; they must develop verification strategies, question generalizations based on finite evidence, and maintain appropriate caution about pattern-based reasoning.

2 Theoretical Framework

Our study integrates the three frameworks as *design principles* shaping every instructional decision. Critical AI literacy emphasizes students' capacity to critically evaluate AI outputs, with particular focus on **verification literacy**—treating AI results as hypotheses requiring proof—and **interrogative stance**—habits of questioning claims and asking for justification (Long & Magerko, 2020). For mathematics, computational verification differs fundamentally from logical proof; students must recognize this distinction.

Cognitive offloading theory examines how people delegate mental work to external tools (Risko & Gilbert, 2016). **Strategic offloading**—deliberate, metacognitively aware delegation—differs from default offloading where tools simply substitute for thinking. When students thoughtfully choose *what* to offload (tedious iteration) versus *what* to retain (pattern synthesis, conjecture formation), they develop metacognitive awareness about tool use. For example, a student might ask AI to compute sequences for starting values 1–100, then personally analyze the convergence patterns to identify which parameters produce cycles versus divergence. Research shows strategic offloading can enhance learning by freeing resources for deeper processing (Storm & Stone, 2015).

Hybrid intelligence characterizes human-AI collaboration along dimensions including **complementarity** (leveraging different strengths), **role negotiation** (explicitly defining boundaries), and **synergy** (combined results exceeding individual capabilities) (Dellermann et al., 2019). In our Collatz investigation, complementarity emerges when students leverage AI's computational speed to test hundreds of starting values while contributing human insight to notice which parameter combinations produce stable patterns—whether convergence, cycles, or bounded behavior—worth investigating further. Effective partnerships require humans to understand AI's capabilities and limitations while maintaining control over decision-making (Jarrahi, 2018).

Our central research questions ask: How does Collatz investigation through human-AI collaboration shape students' development of (1) critical AI literacy practices, particularly verification and questioning; (2) strategic cognitive offloading patterns; and (3) effective human-AI partnership characterized by complementarity and clear roles?

3 Proposed Study Design

This design-based research study (Design-Based Research Collective, 2003) will be conducted as a pilot implementation with undergraduate students (N=15) enrolled in one section of a discrete mathematics course at Sichuan University, taught by the lead author. Rather than a comparative quasi-experiment, we adopt a design-based research approach that prioritizes depth over breadth—emphasizing rich qualitative data, detailed process analysis, and iterative refinement. While N=15 limits statistical generalizability, it provides ideal conditions for fine-grained examination of how students develop critical AI literacy practices through structured mathematical inquiry. The study operationalizes our theoretical framework through a four-phase investigation where task design necessitates verification and questioning.

3.1 The Pedagogical Innovation: Parameter-Based Discovery

Our key innovation addresses a fundamental challenge: how to prevent students from simply asking AI "What is the Collatz conjecture?" and receiving the answer. We designed a parameter-based discovery sequence where students explore *variants* before the conjecture is formally introduced. Table 1 summarizes the four phases.

Table 1: *Investigation Phases and Theoretical Operationalization*

Phase	**Primary Constructs**	**Operationalization Strategy**
Phase 1: Parameter Exploration	Strategic Offloading; Hybrid Intelligence	Different groups explore parameter variants; AI computes, students analyze
Phase 2: Cross-Group Comparison	Interrogative Stance; Verification Literacy	Groups compare findings; students verify AI claims and form conjectures
Phase 3: Historical Context	Algorithmic Awareness	Formal Collatz introduction; AI research on proof attempts and limitations
Phase 4: Metacognitive Reflection	Strategic Offloading; Role Negotiation	Explicit reflection on cognitive division of labor and partnership quality

Phase 1: Students work in small groups, each receiving different parameter triplets (d, m, a) defining recursive functions (divide by d when even; multiply by m and add a when odd). Group A receives the classic Collatz parameters (2,3,1) without knowing they are special; other groups receive mathematically distinct variants: (2,3,-1) produces predictable cycles, (2,5,1) typically diverges to infinity, and (3,2,1) exhibits irregular convergence patterns. These choices ensure the uncertainty is mathematically real, not contrived—different parameter sets genuinely produce qualitatively different behaviors. Using AI as computational partner, students test starting values to determine convergence/divergence behavior, recording results in data tables. AI handles iteration; students analyze patterns and formulate preliminary conjectures as written statements. This creates authentic uncertainty—students cannot dismiss divergence as "error" but must explain *why* parameters behave differently.

Phase 2: Cross-group presentations reveal Group A's (2,3,1) parameters produce unique convergent behavior compared to the cycles and divergence other groups observe. A structured Critical Questioning Protocol scaffolds questioning habits: "How many values tested? Did you verify manually? Could there be exceptions?" Students must hand-verify AI sequences before conjectures are accepted, reinforcing the idea that computational evidence requires justification.

Phase 3: The instructor reveals students have discovered the famous unsolved Collatz Conjecture. Students use AI to research the problem's history and failed proof attempts, developing algorithmic awareness by distinguishing computational verification (testing trillions of cases) from mathematical proof. This develops appropriate caution about computational boundaries.

Phase 4: Students explicitly reflect on their cognitive division of labor across all phases, evaluating partnership quality and articulating what they offloaded versus retained.

Each phase yields student artifacts—data tables, conjecture statements, verification notes, and reflection responses—that feed into transcript and reflection analysis, providing triangulated evidence of developing critical AI literacy practices.

3.2 Data Collection and Analysis

This design-based research study emphasizes rich qualitative data collection across multiple sources: student-AI chat transcripts capturing real-time interactions, structured reflection journals targeting specific theoretical constructs (verification literacy, strategic offloading, questioning habits), think-aloud interviews with purposive sample (approximately 5–6 students representing diverse engagement patterns), and post-study surveys. Analysis employs iterative framework-based coding organized by key constructs: *verification-seeking*, *limitation-acknowledgment*, *strategic-offloading*, *complementarity*, *interrogative-stance*. We identify critical moments demonstrating development (or absence) of target practices, analyze patterns in cognitive offloading decisions across phases, and characterize the evolution of human-AI partnership quality along hybrid intelligence dimensions. The small sample size enables detailed case-by-case analysis impossible with larger samples, allowing us to trace individual students' developmental trajectories and identify mechanisms underlying observed practices.

4 Anticipated Findings and Implications

We anticipate observing developmental trajectories in critical AI literacy across the four phases. Students will likely demonstrate gradual growth in verification literacy—initially accepting AI outputs uncritically, then developing manual checking practices after Phase 2's verification protocol makes this norm explicit. We expect critical questioning to emerge iteratively, with students first relying heavily on the structured protocol, then progressively internalizing these questioning habits. Strategic offloading patterns should become increasingly metacognitively sophisticated as students articulate rationales for delegation decisions and recognize complementary strengths in human-AI partnership.

The fine-grained qualitative data will allow us to identify specific mechanisms underlying these developments: Which pedagogical moves trigger shifts in verification behavior? What kinds of AI errors or inconsistencies prompt students to question outputs? How do students negotiate cognitive division of labor, and what factors influence these decisions? By analyzing transcripts and reflections at the individual student level, we can trace the microgenetic development of critical AI literacy practices impossible to detect in larger-scale studies.

The parameter-based design should reveal how task structure shapes AI engagement: creating conditions where students use AI as a computational partner for exploration rather than as an answer source. This addresses a key challenge in AI-integrated instruction—how to design tasks that leverage AI's strengths while preserving opportunities for mathematical reasoning and discovery.

We anticipate Phase 3 will produce crucial moments of algorithmic awareness, particularly when students discover AI can describe but not explain mathematical "why" or generate novel proofs. This distinguishes computational capability from mathematical reasoning, a critical insight for mathematics education in the AI era. The design-based research approach allows us to document these moments in detail, identifying specific pedagogical features that trigger critical insights and refining the intervention iteratively for future implementations.

Implications for Practice: Our findings will inform design principles for AI-integrated inquiry: (1) creating authentic uncertainty that necessitates critical engagement; (2) explicitly scaffolding verification and interrogation through protocols rather than assuming students will spontaneously develop these practices; (3) positioning AI as computational partner requiring oversight rather than authoritative oracle; (4) making cognitive offloading decisions visible and deliberate through metacognitive reflection.

Theoretical Contributions: This study demonstrates how CAL, cognitive offloading, and hybrid intelligence frameworks can function as integrated "architectural blueprint" for pedagogical design rather than post-hoc analytic lenses. Each framework addresses distinct dimension: CAL provides critical dispositions, offloading theory explains cognitive resource management, hybrid intelligence characterizes effective partnership structure. Together, they offer comprehensive approach to AI-integrated learning.

Broader Significance: As AI becomes ubiquitous in education, we need pedagogies that develop students' capacity to work *with* AI critically and strategically rather than depend on it uncritically or reject it entirely. Mathematical inquiry with open problems provides ideal context because authentic uncertainty demands rather than permits critical engagement. Our study offers empirically-grounded model for cultivating AI literacy through disciplinary practice.

5 Conclusion

This chapter examines how intentional design—grounded in critical AI literacy, cognitive offloading, and hybrid intelligence frameworks—shapes students' critical thinking while investigating mathematics with generative AI. Our parameter-based discovery approach creates authentic conditions for critical engagement by preventing oracle-seeking and requiring verification of computational claims.

By documenting students' development of verification literacy, questioning habits, strategic offloading, and effective human-AI partnership, this work contributes design principles and theoretical integration for educators leveraging AI while developing critical thinking for an AI-saturated world.

References

[1] Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, *32*(1), 5–8.

[2] Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering*, *61*, 637–643.

[3] Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577–586.

[4] Lee, V. R., Bain, C., & Barua, A. (2021). Critical AI literacy: A conceptual framework. *TechTrends*, *65*(4), 510–519.

[5] Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16).

[6] Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688.

[7] Storm, B. C., & Stone, S. M. (2015). Saving-enhanced memory: The benefits of saving on the learning and remembering of new information. *Psychological Science*, *26*(2), 182–188.