2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #51620

AI-Assisted Machine Learning Result Interpretation: Effects on Student Learning and Confidence in an Undergraduate ML Course

Dr. Vinod Kumar Ahuja, Florida Gulf Coast University

Dr. Vinod Kumar Ahuja is an Assistant Professor of Computing and Data Science at Florida Gulf Coast University. He holds a Ph.D. in Information Technology from the University of Nebraska Omaha (2023) with a concentration in Human-Centered Computing. His research examines the impact of artificial intelligence and large language models on engineering education, particularly in machine learning courses. Dr. Ahuja teaches undergraduate Machine Learning and has active research exploring how AI tools affect student learning outcomes. He has published at ICIS, AMCIS, and FLAIRS conferences and mentors multiple undergraduate research assistants in AI and data science projects.

AI-Assisted Machine Learning Result Interpretation: Effects on Student Learning and Confidence in an Undergraduate ML Course

Abstract

The rapid proliferation of generative AI tools in education presents both opportunities and challenges for engineering educators, particularly in machine learning (ML) courses where students develop complex result interpretation skills. This study examines how AI assistance affects undergraduate students' ability to interpret machine learning results through a semester-long empirical investigation in an authentic classroom context. Using a repeated-measures pre-post design, we collected data from 19 students across five sequential ML assignments, measuring confidence, multidimensional understanding (technical accuracy, metrics interpretation, real-world applications, model limitations, causal reasoning), and self-reported learning versus dependency through paired questionnaires administered before and after AI usage. Results demonstrate that AI assistance produced moderate but meaningful improvements in student confidence (9.4% average increase, Cohen's d=0.47), though individual experiences varied substantially: 41% improved, 48% showed no change, and 11% declined. AI effectiveness varied dramatically by concept complexity, with Support Vector Machines and Neural Networks showing statistically significant improvements ($p=0.012$, Cohen's $d=0.816$) while basic concepts showed minimal effects. Analysis across understanding dimensions revealed AI particularly supports higher-order interpretive skills, especially causal reasoning (+15.6%) and metrics interpretation (+15.3%), while showing limited impact on domain-specific contextualization. Students' baseline confidence increased naturally across the semester, providing evidence against harmful dependency. These findings suggest evidence-based practices for AI integration: structured phasing requiring independent work before AI assistance, explicit instruction in critical AI evaluation, strategic deployment for mid-to-high complexity concepts, and monitoring individual responses. As AI becomes ubiquitous in engineering education, our research demonstrates the importance of nuanced, empirically-grounded integration strategies.

1. Introduction

The rapid proliferation of generative artificial intelligence tools has created both unprecedented opportunities and significant challenges for engineering education. Large language models (LLMs) such as ChatGPT, Claude, and Gemini have become readily accessible to students, offering instant assistance with understanding complex technical concepts, debugging code, and interpreting analytical results [1]. In machine learning (ML) education specifically, where students must develop not only technical implementation skills but also the ability to interpret, explain, and critique model outputs, these AI tools present a particularly compelling use case and a particularly vexing pedagogical dilemma [2].

The challenge facing educators is multifaceted. On one hand, AI tools offer potential scaffolding for developing interpretive competencies that are often difficult to teach and assess. Students can receive immediate explanations of unfamiliar metrics, explore connections between technical results and real-world implications, and identify potential limitations in their analyses; all capabilities that mirror the kind of expert guidance traditionally provided through office hours and individual mentoring [3]. On the other hand, unrestricted AI access raises legitimate concerns about superficial learning, over-dependence on automated assistance, and the potential for students to bypass the productive struggle that leads to genuine understanding [4].

Unlike many technical skills in computer science education, ML result interpretation requires integrating multiple forms of knowledge: statistical reasoning about model performance metrics, domain understanding to contextualize predictions, awareness of methodological limitations and potential biases, and communication skills to explain technical findings to diverse audiences [5]. This complexity makes ML interpretation an ideal context for examining AI assistance effects, as the task demands more than simple recall or procedural knowledge; it requires the kind of higher-order thinking that educators most want to cultivate [6].

Current institutional responses to generative AI in education range from outright prohibition to unrestricted acceptance, with most institutions occupying an uncertain middle ground characterized by vague guidance and inconsistent enforcement [7]. These policy decisions are being made largely in the absence of empirical evidence about how AI tools affect student learning in specific disciplinary contexts [8]. While research has begun examining AI's impact on writing instruction [9] and introductory programming [10], the effects on technical result interpretation, a critical skill across STEM disciplines, remain largely unexplored.

This paper addresses this gap through an empirical study examining how undergraduate students in a ML course used AI assistance for result interpretation across five sequential assignments spanning an entire semester. Using paired pre-post self-reflection questionnaires administered for each assignment, we measured changes in student confidence, understanding across multiple dimensions, AI interaction patterns, and self-reported learning versus dependency. Our research questions are:

RQ1: How does AI assistance affect student confidence and understanding when interpreting ML results?
RQ2: What interaction patterns do students use with AI tools, and what challenges do they encounter?
RQ3: Do students perceive AI assistance as enhancing learning or creating dependency?

Our findings reveal nuanced patterns in AI's educational impact. AI assistance produced a moderate but meaningful improvement in student confidence when interpreting ML results, with an average 9.4% increase representing a moderate effect size (Cohen's d=0.47). However, individual experiences varied substantially; while approximately 40% of students showed measurable improvements, nearly half experienced no change, and about 10% saw declines in confidence. Effectiveness varied dramatically across different ML concepts, with more complex topics showing significantly stronger benefits than foundational concepts.

Analysis of understanding across five dimensions: technical accuracy, metrics interpretation, real-world applications, model limitations, and causal reasoning, revealed that AI assistance was particularly effective for higher-order interpretive skills, especially helping students understand *why* specific results occurred. Conversely, AI showed a more limited impact on connecting technical results to practical applications, suggesting domain knowledge limitations of general-purpose AI tools.

Students overwhelmingly preferred ChatGPT (used in over 80% of cases), though tool choice did not correlate with learning outcomes. Interaction patterns revealed balanced usage between seeking comprehension support (clarifying terminology and metrics) and pursuing higher-order analysis (identifying limitations and applications). Students reported specific challenges, including information overload, unfamiliar AI-generated terminology, and AI-induced doubt in their own analysis, yet 81% indicated intent to continue using AI for future work.

When asked directly about learning versus dependency, the majority of students perceived AI as primarily enhancing their learning, with about one-quarter reporting balanced perspectives acknowledging both benefits and concerns. Strong dependency concerns were rare. Importantly, students' baseline confidence, measured before AI assistance, increased naturally across the semester, providing evidence against harmful dependency. If students were becoming dependent rather than developing genuine skills, their pre-AI confidence would have stagnated or declined. Instead, the data suggests AI scaffolding complements rather than replaces skill development.

This research makes several contributions to engineering education. First, we provide semester-long empirical evidence about AI's actual effects on technical learning in an authentic classroom context, moving beyond speculation to systematic measurement. Second, our multidimensional assessment framework captures confidence and understanding across technical accuracy, causal reasoning, real-world applications, and limitation identification, revealing which aspects of ML interpretation benefit most from AI scaffolding. Third, our repeated-measures design across different ML topics reveals that AI assistance effects vary substantially by concept complexity, with particularly strong benefits for advanced topics. Finally, our findings inform evidence-based policies for AI integration in technical courses, helping educators design assignments and guidance that leverage AI's benefits while building genuine student competency.

2. Literature Review

2.1. Artificial Intelligence in Engineering Education

The integration of generative AI tools in education represents the latest evolution in computer-supported learning, but one that differs fundamentally from previous educational technologies in its apparent ability to engage in natural language dialogue, provide personalized explanations, and generate novel content [11]. Early research on ChatGPT and similar tools in education focused primarily on detecting AI-generated text [12], assessing academic integrity implications [13], and documenting faculty concerns about cheating [14]. However, this deficit-focused perspective has begun to shift as educators recognize that prohibition is neither technically feasible nor pedagogically optimal [15].

Recent work has started examining productive uses of AI in education. Research has found that structured AI assistance in writing courses improved student revision processes when combined with explicit instruction on effective tool use [16]. Other studies demonstrated that AI-generated explanations in introductory programming helped students debug code more efficiently, though questions remained about whether students developed independent debugging skills [17]. Educational researchers argue that rather than asking whether to allow AI tools, educators should focus on how to integrate them in ways that enhance rather than replace human learning processes [2].

However, most existing research examines relatively simple tasks: essay writing, basic programming exercises, or factual question answering [18]. The effects of AI assistance on complex technical skills requiring integration of multiple knowledge types, such as ML result interpretation, remain largely unexplored. Furthermore, few studies have employed designs that track students' learning across multiple interactions with AI tools, potentially missing important patterns in how dependency or expertise develops over time [19].

2.2. Machine Learning Education and Result Interpretation

Machine learning education faces distinctive pedagogical challenges compared to traditional programming or software engineering courses [20]. Students must develop not only coding proficiency but also statistical reasoning, domain knowledge for contextualization, awareness of ethical implications, and, critically, the ability to interpret what their models are actually telling them [21]. This interpretive skill is often underdeveloped, with students able to implement algorithms and generate predictions but struggling to explain what those predictions mean, why they occurred, or when they should be trusted [22].

Research has found that undergraduate ML students focus disproportionately on maximizing performance metrics while neglecting interpretation of why specific metrics are appropriate, what trade-offs exist, or what limitations their models have [23]. This "black box" approach to ML is problematic not just pedagogically but professionally, as graduates entering the workforce need to communicate technical findings to non-technical stakeholders, justify model choices, and identify potential biases or failures [24].

Several pedagogical approaches have been proposed to address interpretation deficits in ML education. Research advocates for problem-based learning where students work with real, messy datasets requiring extensive interpretation of unexpected results [25]. Other work emphasizes the importance of explicit instruction in model criticism and failure mode analysis [23]. However, these approaches are resource-intensive, requiring significant individual instructor attention, precisely the resource that AI tools might potentially augment.

The intersection of AI assistance and ML education remains largely unexplored. While some research has examined AI code generation tools in ML courses, they focused on implementation rather than interpretation [26]. The question of whether AI tools can effectively scaffold the development of interpretive competencies, or whether they short-circuit the cognitive processes necessary for developing such competencies, remains unanswered. Our study addresses this gap by examining AI assistance specifically for result interpretation rather than code generation, measuring effects across multiple dimensions of understanding.

2.3. Self-Reflection and Metacognitive Awareness in Learning

Self-reflection and metacognitive awareness, the ability to monitor and evaluate one's own understanding, are recognized as critical components of deep learning [27]. Foundational work on reflective practice in professional education emphasized that expertise develops not just through experience but through systematic reflection on that experience [28]. In technical education specifically, research has found that engineering students who regularly engaged in structured self-reflection demonstrated stronger problem-solving skills and more accurate self-assessment compared to peers who did not [29].

Structured self-reflection questionnaires provide a methodological approach for both promoting and measuring metacognitive development [30]. By asking students to explicitly evaluate their understanding before and after specific learning activities, these instruments serve dual purposes: they prompt students to engage in metacognitive monitoring, and they provide researchers with data about students' perceptions of their own learning trajectories [31].

In the context of AI-assisted learning, self-reflection takes on additional importance. Students using AI tools must develop metacognitive skills to evaluate AI-generated content, recognize when AI explanations align with or contradict their understanding, and decide when to accept, question, or seek alternative perspectives on AI suggestions [32]. Research on cognitive processes suggests that reliance on external tools can reduce engagement of slower, more deliberate cognitive processes, precisely the processes associated with deep learning and skill development [33].

However, self-reported measures have known limitations [34]. Students may lack accurate awareness of their own understanding (the Dunning-Kruger effect [35]), may provide socially desirable responses, or may struggle to distinguish between feeling confident and actually being correct. To address these limitations, our study combines self-reflection questionnaires with documentation of actual AI interactions, triangulates quantitative confidence scores with qualitative descriptions, and examines patterns across multiple assignments rather than relying on single-point measurement.

2.4. Research Gap

Despite growing interest in AI's educational impact, significant gaps remain in understanding how these tools affect the learning of complex technical skills. Most research has examined AI tools in relatively simple contexts, focused on perceptions rather than learning patterns, or relied on cross-sectional designs [36][37][38]. This study addresses these gaps by examining AI-assisted ML result interpretation across an entire semester, measuring both confidence and multidimensional understanding, tracking patterns across students with varying experience levels, and investigating which specific aspects of complex interpretation tasks benefit most from AI scaffolding.

3. Methodology

3.1. Research Design

This study employed a repeated-measures pre-post design within a single undergraduate ML course. For each of the five individual assignments, students completed a paired self-reflection questionnaire before and after using AI tools for result interpretation. This structure allowed for the measurement of immediate AI effects while identifying longitudinal patterns across various topics.

The study received Institutional Review Board (IRB) approval. To ensure ethical standards, all students completed the coursework and questionnaires as part of the standard curriculum; however, only data from those who provided informed consent were included in the research. This approach guaranteed that participation decisions had no impact on grades or learning opportunities.

3.2. Setting and Participants

The study was conducted during the Fall 2025 semester at a regional comprehensive university in the southeastern United States. As a requirement for Computing and Data Science majors, the course covers supervised and unsupervised learning techniques, including regression, neural networks, and clustering. Students implemented these algorithms in Python using Visual Studio Code, utilizing GitHub Copilot for coding assistance and result interpretation. Of the 21 enrolled students, 19 (90.5%) consented to the use of their data for research. The cohort was predominantly novice, with 67% reporting no prior ML experience and only 9% having completed previous internships or projects. Regarding prior AI tool usage for learning, 24% of participants had never used such tools, while the majority (57%) reported only occasional use, followed by 14% regular and 5% extensive users.

3.3. Course Structure and Assignment Design

The course followed a structured three-phase approach designed to integrate AI assistance while maintaining academic rigor. For each individual assignment:

- **Phase 1 (Independent Work):** Students independently implemented the required ML algorithm, trained their model on a provided dataset, generated initial results, and documented their baseline interpretations of the model's outputs. This phase ensured students engaged with technical implementation before seeking AI assistance.
- **Phase 2 (AI-Assisted Interpretation):** Students used AI tools of their choice to help interpret results. They completed a pre-AI questionnaire assessing their initial confidence and understanding, interacted with AI tools while documenting their interactions, and then completed a post-AI questionnaire assessing changes.
- **Phase 3 (Deliverable Submission):** Students submitted their implementation code, results, and a comparative analysis detailing their initial interpretations and how these were refined or expanded following AI assistance. They also provided documentation of their AI prompts and workflows, with grading rubrics assessing both technical accuracy and the depth of their interpretive reflections.

The five assignments covered: (1) Linear and Polynomial Regression, (2) Logistic Regression and Classification, (3) Tree-Based Models and Random Forests, (4) Support Vector Machines and Neural Networks, and (5) Unsupervised Learning and Clustering.

3.4. Measurement Instrument

The self-reflection questionnaire was developed based on literature on ML education competencies [21][22][39] and metacognitive assessment [30][31]. This instrument was automatically de-identified with research codes and captured data across several constructs:

- **Background & Confidence:** Prior experience and a comparative assessment of confidence in explaining results before and after AI use (5-point Likert scale).
- **Multidimensional Understanding:** Pre- and post-ratings across five dimensions: technical accuracy, metric interpretation, real-world implications, model limitations, and causal reasoning.
- **Task-Specific Competencies:** Self-reported ability to use proper terminology, identify patterns, and connect results to the original problem, measured both before and after the AI intervention.
- **AI Experience & Reflection:** Post-intervention feedback on the helpfulness of the AI and open-ended reflections on how their understanding changed.

3.5. Data Analysis

Only de-identified data from consenting students was included in the analysis. We employed a mixed-methods approach to evaluate the intervention:

1. **Quantitative Analysis:** Paired t-tests compared the pre- and post-intervention scores within the Qualtrics survey. Effect sizes (Cohen's) were calculated to assess the magnitude of change.
2. **Qualitative Analysis:** We performed thematic analysis on the open-ended survey reflections and conducted content analysis on the AI interaction logs submitted in the course deliverables to identify common usage patterns.
3. **Integration:** Quantitative confidence shifts were triangulated with qualitative descriptions to validate self-reported changes against the actual interaction data.

4. Results

Findings are presented based on 75 observations from 19 consenting participants across five assignments. These results evaluate the impact of AI on student self-efficacy, the nature of student-AI interactions, and the perceived balance between learning enhancement and tool dependency.

4.1. Effects of AI Assistance on Student Confidence and Understanding

4.1.1. Overall Confidence Effects

AI assistance produced a statistically and practically significant improvement in student confidence. Mean confidence increased from 3.734 before AI to 4.086 after AI (5-point scale), representing a +0.352 point gain (9.4% increase) with moderate effect size (Cohen's d≈0.47). However, individual variation was substantial: 41.3% improved, 48.0% showed no change, and 10.7% declined.

Figure 1 illustrates confidence trajectories across assignments. Students' pre-AI confidence naturally increased from 3.429 (A1) to 3.833 (A5), a +0.404 point gain indicating natural skill development. This baseline improvement provides evidence against harmful dependency—if AI created dependence rather than supporting learning, pre-AI confidence would have stagnated or declined. The widening gap between pre-AI and post-AI lines demonstrates that AI impact strengthened over the semester.

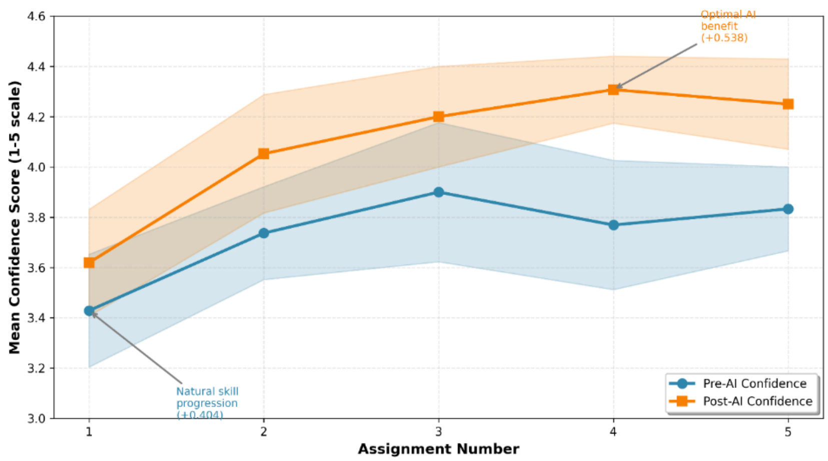

FIGURE 1: Confidence Trajectories Across Assignments

4.1.2. Assignment-Specific Patterns and Statistical Significance

Figure 2 and Figure 3 reveal dramatic variation in AI effectiveness across assignments. Assignment 1 (Regression) showed minimal improvement (+0.190, p=0.493, Cohen's d=0.152, ns), with only 28.6% of students improving. Assignments 2 and 3 showed moderate gains (+0.316 and +0.300, respectively) but remained statistically non-significant.

Assignment 4 (SVMs and Neural Networks) stands out dramatically as the only assignment achieving statistical significance (t=2.941, p=0.012, Cohen's d=0.816, large effect). It produced the highest improvement (+0.538), the highest improvement rate (46.2%), and remarkably, zero confidence declines. This suggests optimal conditions for AI-assisted learning, concepts complex enough to benefit from scaffolding yet structured enough for effective AI explanation.

Assignment 5 (Clustering) showed strong improvement (+0.417, p=0.054, Cohen's d=0.623), approaching significance with half of the students (50.0%) improving, the highest improvement rate across all assignments.

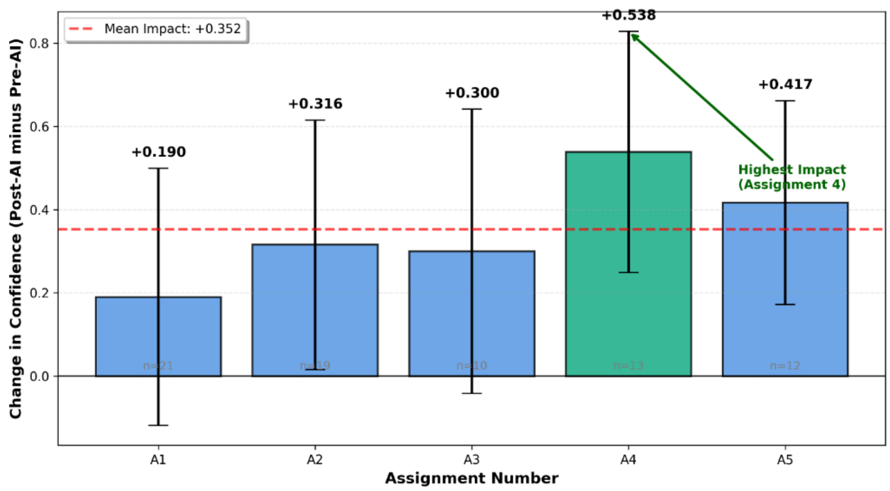

FIGURE 2: Magnitude of AI Impact by Assignment

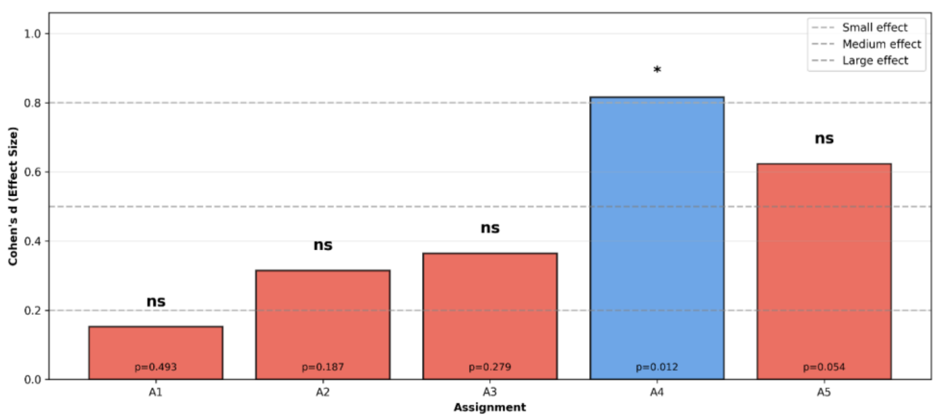

FIGURE 3: Effect Sizes and Statistical Significance by Assignment (*** p<.001, ** p<.01, * p<.05, ns=not significant)

4.1.3. Distribution of Outcomes Over Time

Across all assignments, 41.3% of students improved in confidence, 48.0% showed no change, and 10.7% declined after using AI assistance. Improvement rates increased substantially across the semester: only 28.6% improved in Assignment 1, but 50.0% improved by Assignment 5, a 74% relative increase in students benefiting from AI. Decline rates remained stable around 10% except for Assignment 4 (0% declines) and Assignment 3 (20% peak decline). This progression suggests students developed more effective AI usage strategies over time.

4.1.4. Multidimensional Understanding Effects

Analysis across five understanding dimensions revealed differential AI impacts (Figure 4). Causal Reasoning (understanding why results occurred) showed the strongest improvement (+0.625 points, 15.6%), followed by Metrics Meaning (+0.611, 15.3%) and Model Limitations (+0.596, 14.9%). Technical Accuracy improved by +0.537 (13.4%), while Real-world

Applications showed the smallest gains (+0.474, 11.9%). This pattern indicates AI excels at higher-order interpretive skills and technical explanation but struggles with domain-specific contextualization.

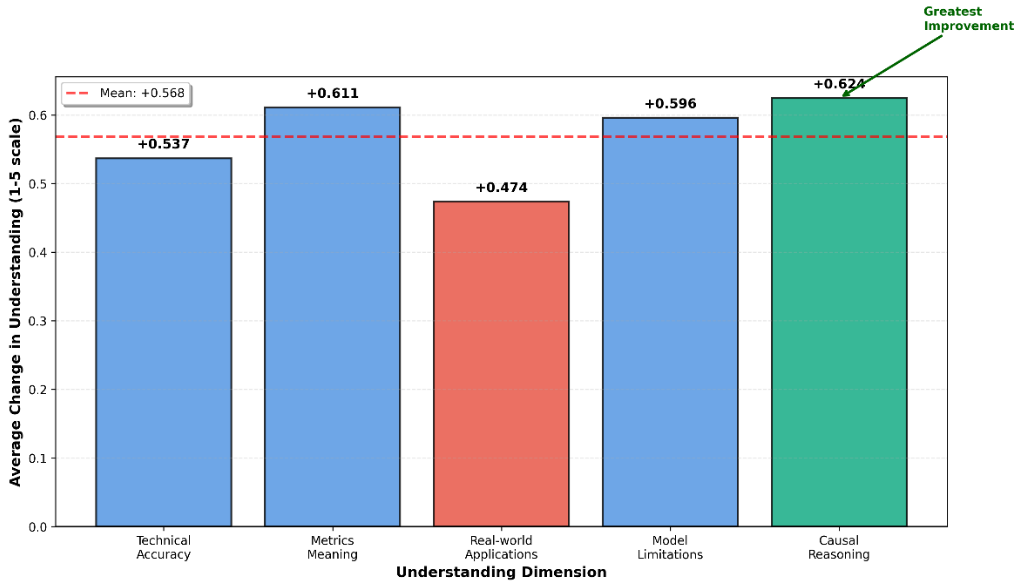

FIGURE 4: AI Impact on Understanding Dimensions

Figure 5's heatmap reveals fascinating assignment-specific patterns. Assignment 2 (Classification) produced dramatic improvements in Metrics Meaning (+0.737, p=0.003), Real-world Applications (+0.789, p=0.001), and Model Limitations (+0.632, p=0.019), with large effect sizes. Assignment 4 showed the most dramatic pattern: Causal Reasoning improved massively (+1.000, p=0.0002, Cohen's d=1.414), a huge effect far exceeding conventional thresholds. One student captured this: *"My understanding of my results after using AI was a lot better. I struggled a lot with understanding the medical terms [in the dataset]... AI was the most helpful this time."*

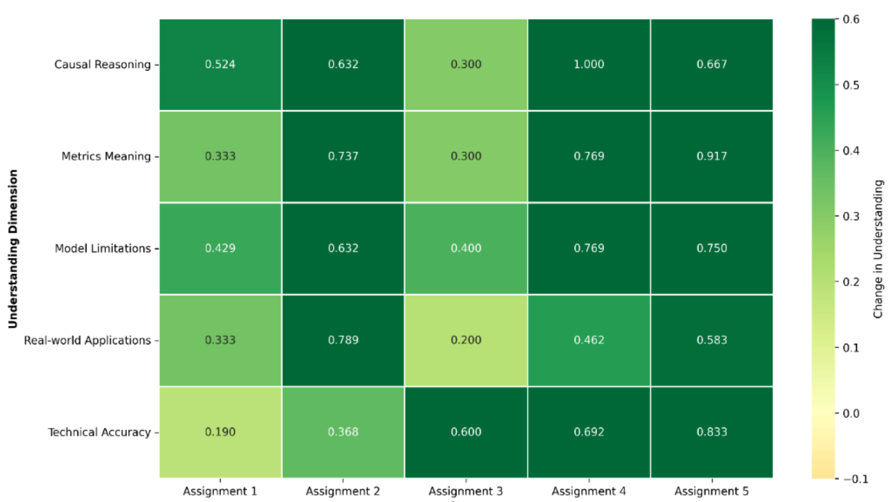

FIGURE 5: Heatmap – Understanding Dimensions Improvement Across Assignments

Assignment 5 continued strong performance with Technical Accuracy (+0.833, p=0.002, Cohen's d=1.161) and Metrics Meaning (+0.917, p=0.002, Cohen's d=1.156) showing very large effects. The pattern confirms AI benefits amplify with concept complexity rather than diminishing.

4.2. Student-AI Interaction Patterns and User Challenges

4.2.1. Tool Preferences

ChatGPT dominated AI tool usage overwhelmingly, used in 61 of 73 responses (83.6%), with Claude being the distant second choice at 12.3%, and other tools accounting for only 4.1%. This ChatGPT dominance likely reflects its first-mover advantage in public awareness, free accessibility, and students' prior familiarity from other contexts. Notably, tool choice did not correlate with learning outcomes; students using ChatGPT exclusively showed similar confidence gains and understanding improvements to those using Claude or multiple tools. This suggests benefits derive from the structured interaction and explanation-seeking process rather than specific capabilities of particular platforms.

4.2.2. Interaction Types

Students' AI queries fell into distinct categories: general result explanations (23%), technical terminology help (20%), limitation identification (17%), metric clarification (16%), real-world applications (14%), and other focused queries (10%). This reveals balanced usage between comprehension support (terminology + metrics = 36%) and higher-order analysis (limitations + applications = 31%).

4.2.3. Challenges and Benefits

Students identified specific challenges: provided too much information (21%), used unfamiliar terminology (21%), made me doubt my analysis (18%), explanations too complex (16%), and generic rather than specific insights (12%). One student noted: *"I could access clearer information using AI; sometimes it offered a proper explanation, but I'm still learning the concepts of ML, so some topics were confusing for me."* Another expressed feeling overwhelmed: *"Claude helps me understand difficult topics, but overwhelms me as the user."*

Conversely, most helpful aspects were clearer technical explanations (24%), identified missed patterns (19%), offered multiple perspectives (17%), connected to real-world applications (16%), and helped organize thoughts (13%). One student articulated career value: *"I got a better understanding of what the results actually meant... I find this super useful, because I like making models on my own time, so this helps me in my hopefully future career."*

4.3. Student Perceptions of Learning Enhancement versus Dependency

4.3.1. Overall Perceptions

When asked directly whether AI assistance enhanced learning or created dependency, students expressed nuanced views. Across all assignments, 36.0% perceived AI as primarily enhancing learning, 25.3% reported balanced perspectives acknowledging both benefits and concerns, while only 1.3% expressed strong dependency concerns. Notably, 37.3% of responses fell into the "Other" category, which, upon examination, consisted primarily of additional positive responses not directly addressing the dependency question. Despite some concerns, an overwhelming 81% of students indicated they intended to continue using AI assistance for future ML work.

The low explicit dependency rate (1.3%) contrasts with the 10.7% confidence decline rate, suggesting students may experience reduced confidence without attributing it to dependency. The proportion reporting "balanced perspectives" decreased from 28.6% (A1) to 8.3% (A5), indicating students developed more definitive opinions with experience—either becoming confident in AI's benefits or more concerned about dependency rather than remaining ambivalent.

4.3.2. Implications

The intention-perception gap is significant: while 26.6% acknowledged dependency concerns, 81% planned continued usage. This suggests that immediate practical benefits outweigh longer-term skill development concerns in students' decision-making.

5. Discussion

This study examined how AI assistance affects undergraduate students' ability to interpret ML results in an authentic classroom context. We discuss our findings in relation to existing literature and explore pedagogical implications.

5.1. Interpreting AI's Impact on Learning

Our moderate effect size (Cohen's d=0.47) aligns with prior observations that structured AI assistance enhances learning when properly integrated [16]. However, the substantial individual variation, 48% showing no change and 11% declining, contradicts assumptions that AI uniformly helps all learners [2]. This variability suggests AI effectiveness depends on factors beyond simple access, including students' prior knowledge, learning strategies, and ability to critically evaluate AI responses.

The Assignment 4 anomaly provides crucial insight into when AI assistance is most valuable. The only statistically significant improvement occurred with Support Vector Machines and Neural Networks, concepts complex enough to genuinely challenge students but structured enough for effective AI explanation. This extends existing frameworks [2] by demonstrating that AI effectiveness depends not just on *how* tools are used but on *what content* they're applied to. Simpler concepts may not require AI scaffolding, while mid-to-high complexity concepts benefit substantially.

Differential effects across understanding dimensions reveal AI's strengths and limitations. Strong improvements in causal reasoning and metrics interpretation indicate AI excels at technical explanation, addressing the interpretive deficits identified in ML students [23]. AI's ability to construct explanatory narratives linking model mechanics to observable outcomes fills a critical pedagogical gap. However, weaker improvements in real-world applications align with observations that connecting technical results to practical contexts requires domain expertise that general-purpose AI tools lack [23]. This limitation has important implications: educators cannot rely on AI alone for contextual understanding and must provide direct instruction for domain-specific applications.

The natural baseline confidence growth across the semester provides evidence against harmful dependency. If AI created reliance, students' pre-AI confidence would stagnate or decline. Instead, genuine skill development occurred alongside AI usage, supporting suggestions that AI can scaffold learning without preventing independent capability when instructional design requires independent engagement before AI assistance [17]. Our three-phase structure appears effective at balancing support with skill development.

5.2. Understanding Interaction Patterns and Challenges

The dominance of ChatGPT despite a lack of correlation with outcomes suggests that tool choice matters less than interaction strategies, a finding with practical implications for educators debating which platforms to recommend or restrict. Rather than mandating specific tools, the focus should be on teaching effective AI interaction applicable across platforms.

Students' challenges reveal fundamental tensions. Complaints about information overload and unfamiliar terminology highlight a mismatch: AI optimizes for comprehensive technical accuracy while novice learners need focused guidance calibrated to their understanding level. This echoes observations about cognitive load management [33], excessive information triggers superficial rather than deep processing.

The 18% reporting AI-induced doubt represents a concerning pattern potentially undermining metacognitive development that has been identified as essential for engineering expertise [29]. Students viewing AI as authoritative rather than collaborative suggest a need for explicit instruction in critical AI evaluation, questioning, verifying, and sometimes rejecting AI suggestions rather than accepting them uncritically.

5.3. Dependency Perceptions and Long-term Concerns

While only 1.3% expressed strong dependency concerns and 36% perceived learning enhancement, 81% intended future usage despite some concerns, revealing a pragmatic prioritization of immediate benefits over longer-term skill development considerations. Students may not recognize dependency until attempting tasks without AI support. The evolution from 28.6% expressing balanced perspectives initially to only 8.3% finally indicates increasingly confident opinions, though whether these accurately reflect AI's actual effects remains uncertain. Students may confuse feeling informed by AI explanations with genuine independent understanding, a distinction requiring objective performance measures to validate.

5.4. Pedagogical Implications

Our findings suggest evidence-based recommendations for AI integration. First, structured phasing is essential, requiring independent work before AI assistance ensures engagement with core concepts. Second, explicit instruction matters; students need guidance on formulating queries, critically evaluating responses, and recognizing when AI exceeds their understanding. Third, complexity matching is crucial; AI appears most valuable for mid-to-high complexity concepts; educators should strategically deploy it where students genuinely need scaffolding. Fourth, dimension-specific strategies are needed, leveraging AI for causal reasoning and technical explanation while providing direct instruction for real-world contextualization. Finally, monitor individual responses, with nearly half showing no change and 10% decline, identify students not benefiting, and provide alternative support.

6. Conclusion

This exploratory study provides preliminary empirical evidence that AI assistance produces moderate but potentially meaningful improvements in undergraduate students' confidence interpreting machine learning results (9.4% average increase, Cohen's $d = 0.47$), though benefits vary substantially by individual students and concept complexity. Support Vector Machines and Neural Networks showed the strongest observed outcomes with statistically significant improvements and zero confidence declines, suggesting a complexity "sweet spot" for AI-assisted learning within this sample. AI particularly supports higher-order skills, causal reasoning, metrics interpretation, and limitation identification, while struggling with domain-specific contextualization. Students' baseline confidence increased naturally across the semester, providing evidence that AI scaffolding complements rather than replaces genuine skill development when structured appropriately.

For engineering educators, these preliminary findings suggest implementing structured phasing requiring independent work before AI assistance, providing explicit instruction in effective AI usage and critical evaluation, strategically deploying AI for mid-to-high complexity concepts, and monitoring individual student responses. Given the modest sample size ($n = 19$), these results should be interpreted as exploratory and require replication with larger, more diverse student populations before broader generalization. Future research should incorporate objective performance measures, employ control groups, and conduct long-term follow-up to assess transfer to contexts without AI support. As AI tools become ubiquitous in engineering education and practice, the path forward requires moving beyond binary debates toward nuanced, evidence-based integration that leverages AI's strengths while maintaining intellectual rigor and genuine skill development.

7. References

[1] M. Sullivan, A. R. Kelly, and P. McLaughlan, "ChatGPT in higher education: Considerations for academic integrity and student learning," *J. Appl. Learn. Teach.*, vol. 6, no. 1, 2023.

[2] E. Kasneci *et al.*, "ChatGPT for good? On opportunities and challenges of large language models for education," *Learn. Individ. Differ.*, vol. 103, p. 102274, 2023.

[3] C. K. Y. Chan and W. Hu, "Students' voices on generative AI: Perceptions, benefits, and challenges in higher education," *Int. J. Educ. Technol. High. Educ.*, vol. 20, no. 1, p. 43, 2023.

[4] D. Baidoo-Anu and L. Owusu Ansah, "Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning," *J. AI*, vol. 7, no. 1, pp. 52–62, 2023.

[5] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," ACM Comput. Surv., vol. 51, no. 5, pp. 1–42, Art. no. 93, Aug. 2018, doi: 10.1145/3236009.

[6] L. W. Anderson *et al.*, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York, NY, USA: Longman, 2001.

[7] 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 *Proc. 2023 ACM Conf. Int. Comput. Educ. Res.*, vol. 1, 2023, pp. 106–121.

[8] R. Arum, M. Calderon Leon, X. Li, and J. Lopes, "ChatGPT early adoption in higher education: Variation in student usage, instructional support, and educational equity," *AERA Open*, vol. 11, no. 1, 2025.

[9] C. Song and Y. Song, "Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students," *Front. Psychol.*, vol. 14, p. 1260843, 2023.

[10] J. Finnie-Ansley, P. Denny, B. A. Becker, A. Luxton-Reilly, and J. Prather, "The robots are coming: Exploring the implications of OpenAI Codex on introductory programming," in *Proc. 24th Australas. Comput. Educ. Conf.*, 2022, pp. 10–19.

[11] J. Rudolph, S. Tan, and S. Tan, "War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education," *J. Appl. Learn. Teach.*, vol. 6, no. 1, 2023.

[12] D. R. Cotton, P. A. Cotton, and J. R. Shipway, "Chatting and cheating: Ensuring academic integrity in the era of ChatGPT," *Innov. Educ. Teach. Int.*, vol. 61, no. 2, pp. 228–239, 2023.

[13] M. Perkins, J. Roe, D. Postma, J. McGaughran, and D. Hickerson, "Detection of GPT-4 generated text in higher education: Combining academic judgement and software to identify generative AI tool misuse," *J. Acad. Ethics*, vol. 22, no. 1, pp. 89–113, Mar. 2024, doi: 10.1007/s10805-023-09492-6.

[14] M. Firat, "What ChatGPT means for universities: Perceptions of scholars and students," *J. Appl. Learn. Teach.*, vol. 6, no. 1, 2023.

[15] B. A. Becker *et al.*, "Programming is hard—or at least it used to be: Educational opportunities and challenges of AI code generation," in *Proc. 54th ACM Tech. Symp. Comput. Sci. Educ.*, vol. 1, 2023, pp. 500–506.

[16] H. Aljuaid, "The impact of artificial intelligence tools on academic writing instruction in higher education: A systematic review," *Arab World Engl. J.*, Special Issue: ChatGPT, pp. 26–55, 2024.

[17] J. Prather *et al.*, "The widening gap: The benefits and harms of generative AI for novice programmers," in *Proc. 2024 ACM Conf. Int. Comput. Educ. Res.*, vol. 1, 2024, pp. 469–486.

[18] C. K. Lo, "What is the impact of ChatGPT on education? A rapid review of the literature," *Educ. Sci.*, vol. 13, no. 4, p. 410, 2023.

[19] Y. K. Dwivedi *et al.*, "Opinion paper: 'So what if ChatGPT wrote it?' Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," *Int. J. Inf. Manage.*, vol. 71, p. 102642, 2023.

[20] M. Baez, F. Daniel, and F. Casati, "Conversational web interaction: Proposal of a dialog-based natural language interaction paradigm for the web," in *Int. Conf. Web Eng.*, 2019, pp. 94–110.

[21] D. Danks and A. J. London, "Algorithmic bias in autonomous systems," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, 2017, pp. 4691–4697.

[22] Z. C. Lipton, "The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery," *Queue*, vol. 16, no. 3, pp. 31–57, 2018.

[23] K. Holstein, B. M. McLaren, and V. Aleven, "Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity," *J. Learn. Anal.*, vol. 6, no. 2, pp. 27–52, 2019.

[24] A. B. Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, 2020.

[25] M. Prince, "Does active learning work? A review of the research," *J. Eng. Educ.*, vol. 93, no. 3, pp. 223–231, 2004.

[26] J. Prather *et al.*, "What do we think we think we are doing? Metacognition and self-regulation in programming," in *Proc. 2020 ACM Conf. Int. Comput. Educ. Res.*, 2020, pp. 2–13.

[27] B. J. Zimmerman, "Becoming a self-regulated learner: An overview," *Theory Pract.*, vol. 41, no. 2, pp. 64–70, 2002.

[28] D. A. Schön, *The Reflective Practitioner: How Professionals Think in Action*. New York, NY, USA: Basic Books, 1984.

[29] J. Turns, B. Sattler, K. Yasuhara, J. Borgford-Parnell, and C. J. Atman, "Integrating reflection into engineering education," in *Proc. 2014 ASEE Annu. Conf. Expo.*, 2014.

[30] D. Boud, R. Keogh, and D. Walker, *Reflection: Turning Experience into Learning*. London, UK: Routledge, 2013.

[31] J. Dewey, *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. Boston, MA, USA: DC Heath, 1933.

[32] G. Schraw and R. S. Dennison, "Assessing metacognitive awareness," *Contemp. Educ. Psychol.*, vol. 19, no. 4, pp. 460–475, 1994.

[33] D. Kahneman, *Thinking, Fast and Slow*. New York, NY, USA: Farrar, Straus and Giroux, 2011.

[34] J. Langan-Fox and S. Grant, "The Thematic Apperception Test: Toward a standard measure of the big three motives," *J. Pers. Assess.*, vol. 87, no. 3, pp. 277–291, 2006.

[35] J. Kruger and D. Dunning, "Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments," *J. Pers. Soc. Psychol.*, vol. 77, no. 6, pp. 1121–1134, 1999.

[36] S. Bayne, "What's the matter with 'technology-enhanced learning'?" *Learn. Media Technol.*, vol. 40, no. 1, pp. 5–20, 2015.

[37] W. Holmes, M. Bialik, and C. Fadel, *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Boston, MA, USA: Center for Curriculum Redesign, 2019.

[38] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education—where are the educators?" *Int. J. Educ. Technol. High. Educ.*, vol. 16, no. 1, pp. 1–27, 2019.

[39] A. Ng and K. Soo, *Machine Learning Yearning: Technical Strategy for AI Engineers*. 2018.