Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Navigating the Human-Tech Nexus: Hybrid Intelligence as a Driver of Innovation and Trust in Software Engineering Decision-Making

ORKUN KOCAK

orkun.kocak@tedu.edu.tr

TED University

Şahin İdil

Scientific and Technological Research Council of Turkey

Murat Aydoğan

Fırat University

Article

Keywords: AI in engineering education, human–AI collaboration, software engineering decision-making, trust in automation, hybrid intelligence, scenario-based evaluation

Posted Date: August 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9615155/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Artificial intelligence (AI) is increasingly influencing software engineering workflows, playing a significant role in algorithm design, architectural decision-making, secure system development, and performance optimization. As AI-assisted tools have become more prevalent in professional practice, understanding their impact on decision-making processes in computer engineering education has become essential. This study presents an AI-supported research platform designed to examine how computer engineering students and professionals evaluate human-generated, AI-generated, and human–AI hybrid solutions across common software development scenarios.AI is increasingly embedded in software engineering workflows, yet the educational impact of AI-assisted decision support remains insufficiently understood. This study extends an existing scenario-based manuscript by incorporating an empirical analysis of a survey dataset collected through an AI-supported research platform. The platform exposed participants to realistic software engineering tasks and recorded preferences for AI-only, human-only, and human–AI hybrid solutions together with ratings of accuracy, feasibility, creativity, and trust. The analyzed dataset comprised 87 evaluations from 18 participants. Descriptive results showed that hybrid solutions were selected most frequently **(52 of 87 evaluations, 59.8%)**, followed by AI (20.7%) and human solutions (19.5%). Human solutions achieved the highest mean scores for accuracy (M = 4.35) and trust (M = 4.47), whereas hybrid solutions combined similarly strong accuracy (M = 4.33) with the highest creativity score (M = 4.17). AI-only solutions remained viable but were rated lower overall, especially in accuracy and trust. Although inferential tests were limited by sample size and repeated-response structure, the pattern consistently favored hybrid decision-making as the most balanced configuration. The findings suggest that software engineering education should not treat AI as a direct substitute for human judgment; rather, it should cultivate learners' ability to critique, calibrate, and integrate AI assistance within accountable human-centered workflows.

1. Introduction

Software engineering has seen massive changes in the way developers work due to the significant advancements in Artificial Intelligence (AI) over the last decade.

Machine learning algorithms and AI-assisted tool development capabilities have moved beyond simply automating repetitive tasks to generate, evaluate, and recommend complex engineering solutions. As a result of the growth in AI supported tool development, there has been a significant shift in how decisions are being made in the software development community. Traditional software development workflows were based on human-centered decision making; however, with the introduction of AI, decision making now involves collaborative human-AI decision making.

Decision-making is a key skill in computer engineering education. Computer engineers are regularly called upon to analyze constraints associated with a specific problem, determine viable alternatives among potential design strategies, make compromises between competing requirements including performance, security, scalability and maintainability, and explain why they chose particular design

elements in a dynamic and uncertain environment. For decades, traditional engineering education has placed emphasis on developing analytical reasoning skills, technical knowledge, and domain expertise to form the basis for making good decisions. However, the advent of AI systems that can create technically reasonable solutions raises questions about the applicability of the above-mentioned foundational principles of decision-making, and also creates new cognitive and pedagogical dynamics.

From an educational perspective, the use of AI supported tools presents both opportunities and challenges. In addition to supporting the learning process through faster problem exploration, AI can introduce students to possible alternative solutions, and facilitate the reflective process. Conversely, over-reliance on AI generated output could lead to decreased levels of critical thinking, conceal accountability for decisions, and fundamentally change the nature of establishing trust in engineering decisions. The relevance of these issues to computer engineering is paramount because decision-making quality can directly impact the reliability, security, and ethics of the systems developed.

Prior research in AI-based engineering education has largely addressed intelligent tutoring systems, automated grading, code completion tools, and personalized learning environments (Alanazi et al., 2025; Prather et al., 2023; Kazemitabaar et al., 2023). While these studies show improved efficiencies and learner engagement, most fail to examine the decision-making process in depth (Lim et al., 2025; Wood and Moss, 2024). There is very little empirical data related to how learners and practitioners perceive and evaluate AI-generated solutions versus those created by humans, nor how hybrid approaches affect confidence, creativity, and trust in the decision-making process (Nguyen et al., 2025; Wood and Moss, 2024). Furthermore, many studies utilize abstract tasks or problems of narrow definition, which does not accurately capture the complexities of actual software engineering decision-making in practice (Lim et al., 2025).

Similarly, current research in software engineering is placing increased emphasis on human-AI collaboration as opposed to replacing human decision-making with AI decision-making. Hybrid models, which combine human judgment and AI generated insight to facilitate decision-making, are being proposed as a potentially powerful method for addressing complex decision-making environments. However, the success of human-AI collaboration will depend on how individuals interpret AI generated output, incorporate AI output into their own decision-making process, and assign accountability for final decisions. Until researchers conduct systematic empirical investigations of human-AI collaboration, assumptions regarding its benefits are largely speculative.

Research Gap and Positioning of the Present Study

The present study is important for two interconnected reasons. First, it addresses a clearly identified gap in the literature: despite the rapid integration of AI tools into software engineering practice and education, empirical evidence comparing how human-generated, AI-generated, and hybrid solutions are evaluated under realistic, multi-criteria conditions remains scarce (Lim et al., 2025; Wood and Moss, 2024). Most existing studies focus on performance outcomes or tool usability rather than on the evaluative judgment processes that engineers exercise when choosing between competing solutions.

This study directly targets that gap by capturing not only which solution type participants prefer, but also how they rate it across four dimensions — accuracy, feasibility, creativity, and trust — within authentic software engineering scenarios. Second, the study makes a methodological contribution: by combining solution-type preferences, multi-criteria ratings, demographic variables, and task completion time within a single platform, it enables a more comprehensive view of decision-making behavior than is typically available in the literature (Nguyen et al., 2025; Kazemitabaar et al., 2023).

A second area of importance that remains relatively unexamined is the role of contextual and demographic factors in AI assisted decision-making. Learner's prior experience, educational background, and degree of specialization within their domain may all impact how they perceive AI generated suggestions, how much they trust them, and how quickly they arrive at a decision. In addition, decision time can be used as an indicator of the level of cognitive effort and confidence involved in arriving at a decision (Yan et al., 2025), although decision time is not commonly measured in studies of AI-based engineering education.

This study addresses the lack of empirical research in these areas by introducing an AI-supported research platform that allows for the examination of decision-making in computer engineering contexts. The platform provides the capability to compare human generated, AI generated, and hybrid solutions for multiple realistic software engineering scenarios. The participants evaluate each solution based on multiple criteria (i.e., accuracy, feasibility, creativity, and trust) while providing demographic information and recording the time it took to complete each task to allow for a comprehensive evaluation.

Although prior work documents the growing presence of AI in programming education and professional software development, several gaps remain. First, relatively few studies directly compare human-generated, AI-generated, and hybrid human–AI-generated solutions using the same standardized scenarios and evaluation criteria, which limits interpretability across studies (Kazemitabaar et al. 2023; Prather et al. 2023). Second, many evaluations rely on narrow tasks that under-represent the multi-constraint nature of real software engineering decisions, where feasibility, maintainability, security, and risk must be balanced (Boehm 1988). Third, demographic and behavioral dimensions—such as experience level (student vs. professional) and time-on-task—are not consistently integrated into analyses, despite evidence that AI tool effects may differ across user groups and contexts (Peng et al. 2023; Alanazi et al. 2025).

The present study addresses these gaps by adopting a scenario-based comparative design in which participants evaluate the same set of core computer engineering scenarios and rate solution alternatives across multiple dimensions—accuracy, feasibility, creativity, and trust—while also providing demographic information and behavioral measures (e.g., completion time). In doing so, the study extends existing research on AI in programming education and human–AI collaboration by offering a structured empirical approach to understanding how AI reshapes decision-making in computer engineering contexts.

2. Related Work

2.1 Artificial Intelligence in Engineering Education

AI has been used in engineering education for decades, first appearing prominently through intelligent tutoring systems (ITS) and adaptive learning environments. Early ITS research demonstrated that AI-driven instructional support could improve learning and provide structured guidance by modeling learner knowledge and delivering targeted feedback (Anderson et al. 1995). Follow-up work also showed the value of tutorial dialogue and self-explanation support for strengthening conceptual understanding, emphasizing how interactive AI systems can scaffold deeper reasoning rather than only providing answers (Aleven et al. 2003).

More recently, machine learning—particularly large language models (LLMs)—has expanded the role of AI in computer engineering curricula, enabling tools that support programming, automated feedback, and code generation. Controlled studies suggest that AI code generators can increase task completion and performance for novice programmers, while also raising new questions about how learners interpret, validate, and learn from generated code (Kazemitabaar et al. 2023). Observational and interview-based work further indicates that novice programmers can perceive substantial usability benefits from AI pair-programming tools (e.g., GitHub Copilot), yet often face metacognitive challenges such as over-trusting generated code, incomplete understanding, and reduced deliberate reasoning (Prather et al. 2023). Evidence syntheses likewise report that AI tools can improve performance metrics across programming courses, but results vary by learning context, task type, and learner characteristics—highlighting the need for careful study design and multi-dimensional assessment (Alanazi et al. 2025). Complementing these findings, educational research has begun to examine how informal use of LLMs influences learning outcomes over longer durations, suggesting that the net impact may depend on how these tools are integrated into instruction and assessment (Jošt et al. 2024).

2.2 Human–AI Decision-Making in Software Engineering Workflows

Decision-making is central to software engineering, encompassing choices about algorithms, architectures, security strategies, databases, and performance optimization. Classic software engineering literature emphasizes that effective development involves iterative, risk-aware decision cycles in which early choices substantially shape downstream cost, feasibility, and system quality (Boehm 1988). In educational settings, developing this competence is essential because real-world software engineering problems rarely have a single correct answer; instead, they involve trade-offs among competing constraints such as maintainability, scalability, security, performance, and time-to-delivery.

Engineering education studies show that learners often struggle with ill-structured problems—problems that involve uncertain information, multiple viable solution paths, and evaluative judgment rather than procedural execution (Jonassen 1997). Scenario-based and problem-based learning approaches have therefore been proposed to strengthen decision-making by immersing learners in realistic contexts that

require justification and reflection (Strobel and van Barneveld 2009). However, while these approaches develop evaluative judgment in principle, limited work has explored how emerging AI assistants reshape the decision processes students use when comparing alternative solution strategies in authentic software engineering scenarios.

2.3 Human–AI Collaboration in Engineering Tasks

Human–AI collaboration is increasingly framed as a practical mode of work in software engineering, where AI functions as a "partner" that proposes alternatives, accelerates iteration, and provides on-demand guidance. Empirical evidence from developer productivity studies suggests that AI pair-programming tools can speed completion of programming tasks in controlled experiments (Peng et al. 2023). In educational contexts, AI assistance can change how novices approach problem solving—shifting attention from producing code from scratch toward evaluating, modifying, and integrating generated solutions (Kazemitabaar et al. 2023; Prather et al. 2023). This shift is consequential for decision-making because evaluation becomes a primary skill: learners must assess whether an AI output is correct, safe, maintainable, and appropriate for the constraints of the scenario.

From a research standpoint, human–AI collaboration also raises methodological challenges. If learners can rapidly obtain plausible candidate solutions, then traditional measures of "correctness" alone may fail to capture meaningful differences in decision competence. Instead, research increasingly calls for study designs that capture multi-criteria judgment and reasoning quality—such as feasibility under constraints, the creativity or novelty of trade-offs, and the participant's confidence and justification behavior—particularly when AI outputs are persuasive but occasionally unreliable (Menzies and Zimmermann 2013; Prather et al. 2023). These concerns motivate comparative evaluation approaches that examine not only what choice participants make, but also how they justify and time their decisions when solutions originate from human, AI, or hybrid sources.

2.4 Trust and Evaluation of AI-assisted Decisions

Trust plays a critical role in whether people rely on automation appropriately. Too little trust can lead to disuse, while too much trust can create overreliance and automation bias—especially when the automated aid is perceived as authoritative (Dzindolet et al. 2003; Lee and See 2004). Human–AI trust is shaped by factors such as perceived accuracy, transparency, prior experience, and contextual risk (Choung et al., 2023). These issues become particularly salient in education because students must decide when to accept AI-generated suggestions, when to challenge them, and how to verify correctness and safety.

Explainability and interpretability are frequently proposed as mechanisms to support better trust calibration. Work on explainable AI argues that explanations should align with human expectations and support understanding of why a system produced an output, as well as when it may fail (Miller 2019). At the same time, interpretability research emphasizes that "explainability" is not a single property but a context-dependent goal that requires rigorous evaluation and appropriate stakeholder-oriented measures (Doshi-Velez and Kim 2017). Within programming and software engineering tasks, these

insights suggest that learners' trust in AI solutions is likely intertwined with their ability to evaluate evidence, understand rationale, and identify limitations—yet empirical studies that explicitly measure trust alongside other engineering criteria (e.g., feasibility and creativity) remain relatively scarce.

3. Method

3.1 Research design

This study employed a scenario-based comparative design to analyze software engineering decision-making processes. The six scenarios (S1–S6) included in the platform were structured to represent critical technical trade-offs in the modern software development cycle (architectural choices, performance optimization, and security protocols). Each scenario was designed as an 'ill-structured' problem, requiring participants to make multi-criteria evaluations such as technical accuracy and feasibility.

The study used a scenario-based comparative design implemented through an AI-supported survey platform. Participants reviewed software engineering cases and selected the solution type they considered most appropriate among AI, human, and hybrid alternatives. They then rated the selected solution on four dimensions: accuracy, feasibility, creativity, and trust.

3.2 Dataset and participants

The dataset contained 87 response records generated by 18 participants, each of whom completed multiple scenario evaluations. The average participant age was 25.28 years (SD = 2.65; range = 21–31). Demographic information included education level, engineering field, and years of experience. Because each participant responded to multiple scenarios, the 87 records represent repeated evaluations rather than independent single-shot observations; all analyses account for this repeated-measures structure.

The dataset consists of 87 repeated assessment records obtained from 18 participants. The fact that 66.7% of the participants hold a graduate degree and 55.6% specialize directly in computer engineering enhances the technical depth of the collected data. The data dependency (nested data structure) resulting from each participant responding to multiple scenarios was controlled during the analysis phase using repeated-measure protocols and non-parametric tests.

Table 1
summarises the demographic profile of the 18 participants.

Category	Group	n	%
Education level	Master's	12	66.7%
Education level	Bachelor's	5	27.8%
Education level	Doctorate	1	5.6%
Engineering field	Computer engineering / CS	10	55.6%
Engineering field	Other engineering	8	44.4%
Experience (years)	1	7	38.9%
Experience (years)	2	2	11.1%
Experience (years)	3	3	16.7%
Experience (years)	4	5	27.8%
Experience (years)	5	1	5.6%

Table 1. *Participant profile based on unique users (n = 18).*

Participants selected the solution type they deemed most appropriate for each scenario (AI, Human, or Hybrid) via the AI-powered platform and then rated this solution on a 5-point Likert scale across four psychometric dimensions (accuracy, feasibility, creativity, and trust). To examine the relationship between decision-making speed and cognitive effort, the platform automatically recorded the time spent on each task (completion time) in minutes.

3.3 Analytical approach

The expanded analysis emphasized descriptive statistics to align with the sample size and repeated-response structure. Solution-type frequencies were calculated across all evaluations, followed by mean and standard deviation estimates for the four rating dimensions. Scenario-level distributions and completion-time patterns were also examined. To provide an exploratory inferential layer, chi-square tests were used for categorical distributions and Kruskal–Wallis tests were used for rating comparisons across solution types. These tests are reported cautiously as supportive evidence rather than definitive causal claims.

3.4 Ethical Approval

and Informed Consent

This study was approved by the Social and Humanities Sciences Research Ethics Committee of Fırat University (Decision No. 19, Session 2026/8, dated 27 April 2026). All participants provided voluntary informed consent prior to participation. No personally identifiable information was collected. All

procedures were performed in accordance with the ethical standards of the institutional research committee and the principles of the Declaration of Helsinki. Data collection was conducted at Fırat University, Elazığ, Türkiye."

4. Results

4.1 Overall preference distribution

Across all 87 evaluations, hybrid solutions were preferred most frequently (52, 59.8%), while AI-only (18, 20.7%) and human-only (17, 19.5%) selections were comparatively less common. This pattern indicates that respondents generally favored a collaborative decision-making model over purely automated or purely human alternatives.

Table 2
presents the overall frequency distribution of solution type preferences across all 87 evaluations.

Solution type	Responses (N)	Share (%)
AI	18	20.7
Human	17	19.5
Hybrid	52	59.8

Table 2. *Distribution of selected solution types across all evaluations.*

To visualize this distribution, Fig. 2 presents a bar chart of absolute selection frequencies across the three solution categories.

Having established the overall preference pattern, the following section examines how participants rated their chosen solution type across the four evaluation dimensions.

4.2 Comparative evaluation scores

Table 3
presents descriptive statistics for the four rating dimensions by solution type.

Solution type	n	Accuracy (M ± SD)	Feasibility (M ± SD)	Creativity (M ± SD)	Trust (M ± SD)
AI	18	3.94 ± 0.80	4.00 ± 0.84	4.00 ± 0.77	4.11 ± 0.68
Human	17	4.35 ± 0.61	4.29 ± 0.69	4.00 ± 0.94	4.47 ± 0.51
Hybrid	52	4.33 ± 0.58	4.25 ± 0.76	4.17 ± 0.68	4.29 ± 0.67

Table 3. *Descriptive statistics for solution ratings by selected solution type.*

Figure 3 visualizes these mean ratings side by side, making the relative strengths and weaknesses of each solution type immediately apparent.

Human solutions produced the highest average scores for accuracy (M = 4.35) and trust (M = 4.47). Hybrid solutions achieved nearly identical accuracy (M = 4.33) and the highest creativity score (M = 4.17). AI-only solutions remained positively rated overall, but their mean values were lower across the four criteria. Taken together, the descriptive profile suggests that respondents viewed hybrid solutions as the strongest overall compromise between technical adequacy and collaborative value.

Beyond these aggregate ratings, it is important to examine whether the preference for hybrid solutions was consistent across individual scenarios or driven by a subset of cases.

4.3 Scenario-level distribution

Table 4
shows the distribution of solution type selections across the six scenarios.

Scenario	AI	Human	Hybrid	Total
1	3	3	12	18
2	5	2	10	17
3	4	4	9	17
4	2	4	11	17
5	4	3	10	17
6	0	1	0	1

Table 4. *Scenario-level distribution of selected solution types.*

Figure 4 *translates these counts into proportional shares within each scenario, making cross-scenario comparisons more interpretable.*

4.4 Completion time patterns

Completion time was highly skewed due to several unusually long responses. Table 5 reports both mean and median values; medians are more informative given this skew."

Table 5
Completion time by selected solution type in minutes.

Solution type	Mean minutes	Median minutes
AI	11.95	1.68
Human	16.99	3.66
Hybrid	4.53	1.76

The box plots in Fig. 5 illustrate these distributions in greater detail, revealing the extent and direction of the skew within each solution type group.

4.5 Exploratory inferential analysis

Table 6
summarises the results of the exploratory Kruskal–Wallis comparisons across solution types.

Outcome	Kruskal–Wallis	p-value	Effect size
Accuracy	H = 3.74	p = 0.154	ε^2=0.021
Feasibility	H = 1.63	p = 0.443	ε^2=0.000
Creativity	H = 0.67	p = 0.716	ε^2=0.000
Trust	H = 2.42	p = 0.298	ε^2=0.005
Completion time	H = 4.10	p = 0.129	ε^2=0.025

Table 6
Exploratory Kruskal–Wallis comparisons across solution types. To complement the rating comparisons, chi-square tests were conducted to examine whether categorical demographic factors — scenario, education level, engineering field, and experience — were associated with solution type preference.

Factor	Chi-square	df	p-value	Cramér's V
Scenario	χ^2=6.93	10	p = 0.732	V = 0.200
Education level	χ^2=8.08	4	p = 0.089	V = 0.215
Engineering field	χ^2=3.85	2	p = 0.146	V = 0.210
Experience years	χ^2=11.90	8	p = 0.156	V = 0.262

Table 7 presents the chi-square test results for categorical associations with solution type.

Table 7. Exploratory chi-square tests for categorical association with solution type.

None of the exploratory tests reached conventional significance, yet the descriptive pattern remained consistent throughout. The following section interprets these findings in relation to the four research questions.

5. Discussion

5.1 RQ1: How Does AI-Based Decision Support Influence Human Judgment in Engineering Workflows?

The most salient finding is the strong preference for hybrid solutions (59.8%), indicating that participants neither replaced human judgment with AI nor rejected it. Instead, they favored configurations where AI generates candidate solutions while human oversight remains central. This aligns with research framing AI as a collaborative partner rather than a substitute (Passerini et al., 2025; Gomez et al., 2025). Technically, this reflects the generative nature of LLMs, which produce plausible outputs without guaranteeing correctness. Moreover, these systems are known to occasionally generate hallucinated outputs—responses that are fluent but factually incorrect or unsupported. As a result, their primary value lies in expanding the solution space rather than finalizing decisions. Participants appear to recognize this, using AI for augmentation while retaining human validation and contextual adaptation.

Importantly, this suggests bounded reliance rather than automation bias. Unlike prior findings of overreliance, participants showed calibrated use of AI, likely supported by the scenario-based design encouraging comparison. This indicates a shift from generation-focused to evaluation-driven decision-making. However, this preference may partly reflect risk-averse behavior or selection bias, as participants evaluated only chosen solutions. Future studies should include controlled comparisons. Educationally, these results highlight the need to develop evaluative intelligence—enabling students to critically assess and integrate AI outputs—shifting the engineer's role toward decision orchestration (Kazemitabaar et al., 2023; Prather et al., 2023).

5.2 RQ2: How Does AI Assistance Affect Trust Formation and Trust Calibration?

Human-generated solutions received higher trust (M = 4.47) than AI-only solutions (M = 4.11), suggesting that reliability and accountability remain tied to human agency. This aligns with trust-in-automation literature emphasizing predictability and interpretability (Lee and See, 2004; Li et al., 2024). From a technical perspective, this trust gap can be explained by the probabilistic nature of LLMs, which generate outputs without providing transparent reasoning processes. In addition, these systems may produce hallucinated content—outputs that appear coherent and convincing but are factually incorrect or unverifiable. This creates epistemic uncertainty, particularly in engineering contexts where correctness and traceability are critical, and reduces user confidence in AI-only solutions. Despite this, participants did not exhibit strong automation bias. Instead, they demonstrated calibrated trust, moderating reliance on AI while still benefiting from it in hybrid settings. The scenario-based design likely supported this

behavior by encouraging active comparison rather than passive acceptance. However, lower trust in AI may also reflect limited familiarity or the absence of explainability features. Prior work shows that even imperfect explanations can improve trust, suggesting that the observed gap may be influenced by system design rather than inherent limitations alone. These findings imply that trust calibration should be treated as a core learning objective. Research on AI-powered educational tools further confirms that student trust perceptions critically shape technology adoption decisions (Farhat et al., 2025). Students must learn when to trust, question, or reject AI outputs through structured tasks involving justification, failure analysis, and reflection.

5.3 RQ3: What is the Impact of AI Integration on Decision Quality Compared to Human-only Processes?

Hybrid solutions achieved accuracy comparable to human solutions (M = 4.33 vs. 4.35) while outperforming AI-only solutions, and yielded the highest creativity scores (M = 4.17). This supports the hybrid intelligence framework, where human–AI collaboration leverages complementary strengths (Vaccaro et al., 2024; Alipour et al., 2025). Technically, LLMs excel at generating diverse solutions, while humans provide contextual validation and constraint handling. Hybrid performance reflects this division of cognitive labor, where AI expands options and humans evaluate them. Completion times reinforce this: human-only decisions took longer (median = 3.66 min), while hybrid solutions maintained quality with less time (1.76 min), indicating a shift from generation to evaluation (Vasconcelos et al., 2023). Higher creativity suggests a synergy effect, where AI-generated novelty combined with human filtering enables combinatorial innovation. However, results may be influenced by presentation effects or selection bias. Controlled comparisons are needed. Overall, hybrid decision-making emerges not as a compromise but as a distinct and potentially superior mode of engineering judgment.

5.4 RQ4: How Do AI Autonomy and Transparency Moderate Human Judgment, Trust, and Decision Quality?

Although autonomy and transparency were not directly manipulated, the preference for hybrid solutions suggests users favor human-in-the-loop systems over full automation, consistent with prior work (Gomez et al., 2025; Alipour et al., 2025). From an autonomy perspective, participants preferred bounded autonomy, where AI supports but does not replace decision-making. This aligns with adaptive autonomy frameworks. Transparency provides further insight. Lower trust in AI-only solutions likely reflects perceived opacity, as LLMs do not expose reasoning processes. This limits users' ability to evaluate outputs. Prior research shows that transparency mechanisms—such as explanations or confidence indicators—can improve trust (Vasconcelos et al., 2023; Gottschlich et al., 2025). Thus, reduced trust may reflect lack of interpretability rather than capability. However, these findings remain inferential, as autonomy and transparency were not experimentally controlled. Future studies should isolate these variables. Overall, effective human–AI collaboration depends not only on performance, but on how control and understanding are shared.

5.5 Implications for Curriculum Design

The findings suggest a fundamental shift in engineering education, where AI is not an optional tool but an integral part of decision-making. A key implication is the development of evaluative intelligence. As shown in RQ1 and RQ3, students must assess and refine AI outputs rather than generate solutions from scratch. This challenges traditional assessment models and calls for evaluating reasoning, justification, and integration. Trust calibration (RQ2) should also be a core objective. Students must learn to adjust reliance on AI based on context and limitations through structured activities such as reflection and failure analysis. Additionally, learning environments should reflect human-in-the-loop interaction (RQ4), allowing students to engage with AI under varying levels of autonomy and transparency. More broadly, the engineer's role is shifting from solution generator to decision orchestrator, requiring both technical AI literacy and metacognitive skills. Finally, curricula should adopt scenario-based, interaction-centered designs aligned with human–AI collaboration research (Gomez et al., 2025; Alipour et al., 2025).

6. Conclusion

This study strengthens the existing manuscript by integrating empirical results, summary tables, and visual evidence from the latest survey dataset. The results indicate that hybrid human–AI decision-making is the dominant preference across the evaluated software engineering scenarios, while human-only solutions remain the most trusted and hybrid solutions are viewed as especially creative. Together, these findings suggest that the future of software engineering education should focus on calibrated collaboration with AI rather than simple substitution or prohibition.

Limitations and Future Work

The current dataset is modest in size and contains repeated evaluations from the same users across multiple scenarios. As a result, inferential results should be interpreted cautiously and future studies should recruit a larger participant pool with clearer separation between student and professional subgroups.

The present analysis is primarily descriptive because the available data records the chosen solution type rather than full comparative ratings for all three alternatives within each scenario. Future platform versions could record side-by-side ratings of every alternative, confidence scores, qualitative justifications, and transparency manipulations to enable stronger causal inferences.

Additional future directions include mixed-effects modeling for repeated measures, scenario-specific analysis of justification quality, and experimental manipulation of AI explainability, autonomy, and source disclosure.

Due to the sample size and the fact that the data did not follow a normal distribution, the Kruskal-Wallis test was used for group comparisons, while Chi-square tests were used to examine relationships between categorical variables. The analyses did not focus solely on statistical significance ($p < .05$); instead, effect size coefficients (ε^2 and Cramér's V) were carefully reported to understand the impact of solution types on variance.

This study adopts an exploratory approach by nature. In the current data collection model, having participants rate only the solution type they selected may carry a risk of selection bias. Future studies aim to employ designs in which each participant evaluates all alternatives side-by-side through a blind review process and qualitative justification data is also collected.

Declarations

Author Contribution

O.K. developed the conceptual framework of the study, designed the AI-powered research platform, wrote the main article text, and prepared Figures 1-3. Ş.İ. contributed to the theoretical framework, relevant studies, and article revisions. M.A. collected the data and performed the statistical analyses. All authors reviewed and approved the final article.

Data Availability

The dataset analyzed in this study consists of 87 survey response records collected from 18 participants via an AI-supported research platform. The data are not publicly available due to the privacy and confidentiality of participant information, in accordance with the ethics approval granted by the Social and Humanities Sciences Research Ethics Committee of Fırat University (Decision No. 19, Session 2026/8). Anonymized data may be made available from the corresponding author upon reasonable request.

References

1. Alanazi M, Soh B, Samra H, Li A (2025) The influence of artificial intelligence tools on learning outcomes in computer programming: A systematic review and meta-analysis. Computers 14(5):185. https://doi.org/10.3390/computers14050185
2. Aleven V, Koedinger KR, Popescu O (2003) A tutorial dialogue system to support self-explanation: Evaluation and open questions. In U. Hoppe, F. Verdejo, & J. Kay (Eds.), Proceedings of the 11th International Conference on Artificial Intelligence in Education (pp. 39–46). IOS Press
3. Alipour M, Khalid S, Alqarni M, Hussain S (2025) Enhancing intuitive decision-making through human–AI collaboration: A review. Informatics 12(4):135. https://doi.org/10.3390/informatics12040135
4. Anderson JR, Corbett AT, Koedinger KR, Pelletier R (1995) Cognitive tutors: Lessons learned. J Learn Sci 4(2):167–207. https://doi.org/10.1207/s15327809jls0402_2
5. Boehm BW (1988) A spiral model of software development and enhancement. Computer 21(5):61–72. https://doi.org/10.1109/2.59

6. Choung H, David P, Ross A (2023) Trust in AI and its role in the acceptance of AI technologies. Int J Human–Computer Interact 39(9):1727–1739. https://doi.org/10.1080/10447318.2022.2050543
7. Doshi-Velez F, Kim B (2017) Towards a rigorous science of interpretable machine learning. arXiv:1702.08608. https://arxiv.org/abs/1702.08608
8. Dzindolet MT, Peterson SA, Pomranky RA, Pierce LG, Beck HP (2003) The role of trust in automation reliance. Int J Human–Computer Stud 58(6):697–718. https://doi.org/10.1016/S1071-5819(03)00038-7
9. Farhat F, Elhussain M, Jabeur N, Hasan MK (2025) The critical role of trust in adopting AI-powered educational technology for learning: An instrument for measuring student perceptions. Computers Hum Behavior: Artif Hum 3:100114. https://doi.org/10.1016/j.chbah.2025.100114
10. Gomez C, Teso S, Stumpf S, Leavy S, O'Sullivan B (2025) Human-AI collaboration is not very collaborative yet: A taxonomy of interaction patterns in AI-assisted decision making from a systematic review. Front Comput Sci 6:1521066. https://doi.org/10.3389/fcomp.2024.1521066
11. Gottschlich C, Hopf J, Krol B (2025) Exploring automation bias in human–AI collaboration: A review and implications for explainable AI. AI & Society. https://doi.org/10.1007/s00146-025-02422-7
12. Jonassen DH (1997) Instructional design models for well-structured and ill-structured problem-solving learning outcomes. Education Tech Research Dev 45(1):65–94. https://doi.org/10.1007/BF02299613
13. Jošt G, Taneski V, Karakatič S (2024) The impact of large language models on programming education and student learning outcomes. Appl Sci 14(10):4115. https://doi.org/10.3390/app14104115
14. Kazemitabaar M, Chow J, Ma CKT, Ericson BJ, Weintrop D, Grossman T (2023) Studying the effect of AI code generators on supporting novice learners in introductory programming. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). https://doi.org/10.1145/3544548.3580919
15. Lee JD, See KA (2004) Trust in automation: Designing for appropriate reliance. Hum Factors 46(1):50–80. https://doi.org/10.1518/hfes.46.1.50_30392
16. Li X, Wu Y, Huang R, Luan H (2024) Developing trustworthy artificial intelligence: Insights from research on interpersonal, human-automation, and human-AI trust. Front Psychol 15:1382693. https://doi.org/10.3389/fpsyg.2024.1382693
17. Lim WM, Gunasekara A, Pallant JL, Pallant JI, Pechenkina E (2025) Artificial intelligence innovations, challenges and emerging trends in engineering education. Int J STEM Educ. https://doi.org/10.1007/s44217-026-01137-1
18. Menzies T, Zimmermann T (2013) Software analytics: So what? IEEE Softw 30(4):31–37. https://doi.org/10.1109/MS.2013.86
19. Miller T (2019) Explanation in artificial intelligence: Insights from the social sciences. Artif Intell 267:1–38. https://doi.org/10.1016/j.artint.2018.07.007

20. Nguyen Q, Lim CP, Nguyen TT (2025) The impact of AI-assisted pair programming on student motivation, programming anxiety, collaborative learning, and programming performance. Int J STEM Educ 12:18. https://doi.org/10.1186/s40594-025-00537-3
21. Passerini A, Gema A, Minervini P, Sayin B, Tentori K (2025) Fostering effective hybrid human-LLM reasoning and decision making. Front Artif Intell 7:1464690. https://doi.org/10.3389/frai.2024.1464690
22. Peng S, Kalliamvakou E, Cihon P, Demirer M (2023) The impact of AI on developer productivity: Evidence from GitHub Copilot. arXiv:2302.06590. https://arxiv.org/abs/2302.06590
23. Prather J, Reeves BN, Denny P, Becker BA, Leinonen J, Luxton-Reilly A, Powell G, Finnie-Ansley J, Santos EA (2023) It's weird that it knows what I want: Usability and interactions with Copilot for novice programmers. ACM Trans Computer–Human Interact 31(1) Article 4. https://doi.org/10.1145/3617367
24. Strobel J, van Barneveld A (2009) When is PBL more effective? A meta-synthesis of meta-analyses comparing PBL to conventional classrooms. Interdisciplinary J Problem-Based Learn 3(1). https://doi.org/10.7771/1541-5015.1046
25. Vaccaro M, Almaatouq A, Malone T (2024) When combinations of humans and AI are useful: A systematic review and meta-analysis. Nat Hum Behav 8(12):2293–2303. https://doi.org/10.1038/s41562-024-02024-1
26. Vasconcelos H, Jörke M, Grunde-McLaughlin M, Gerstenberg T, Bernstein MS, Krishna R (2023) Explanations can reduce overreliance on AI systems during decision-making. Proceedings of the ACM on Human-Computer Interaction, 7(CSCW1), Article 129. https://doi.org/10.1145/3579605
27. Wood D, Moss SH (2024) Evaluating the impact of students' generative AI use in educational contexts. J Res Innovative Teach Learn 17(2):152–167. https://doi.org/10.1108/JRIT-06-2024-0151
28. Yan L, Greiff S, Teuber Z, Gašević D (2025) Effects of generative artificial intelligence on cognitive effort and task performance. Trials 26:185. https://doi.org/10.1186/s13063-025-08950-3

Figures

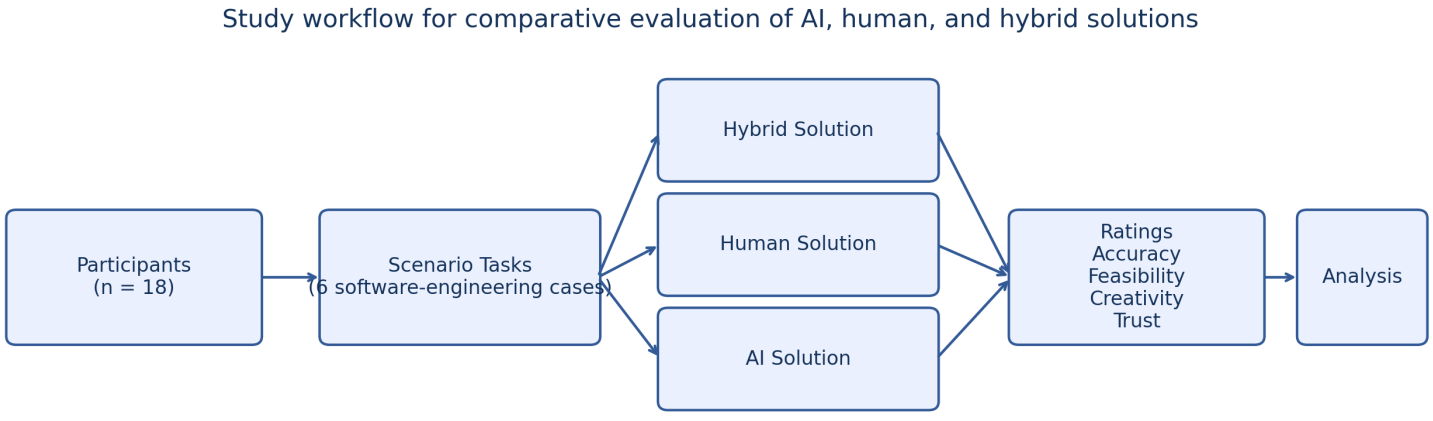


Figure 1

Workflow of the scenario-based evaluation process used in the study.

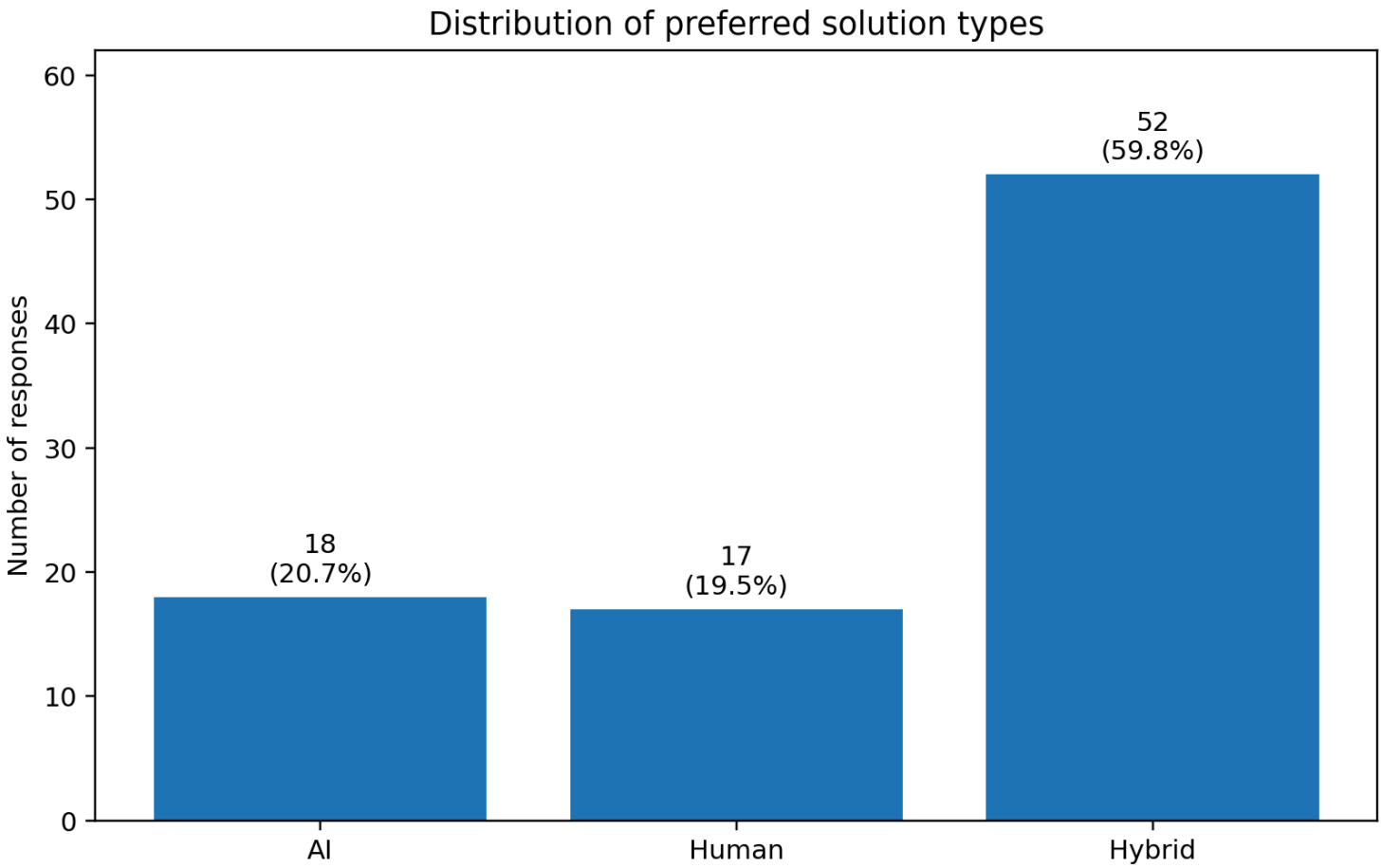


Figure 2

Frequency of AI, human, and hybrid selections across the dataset.

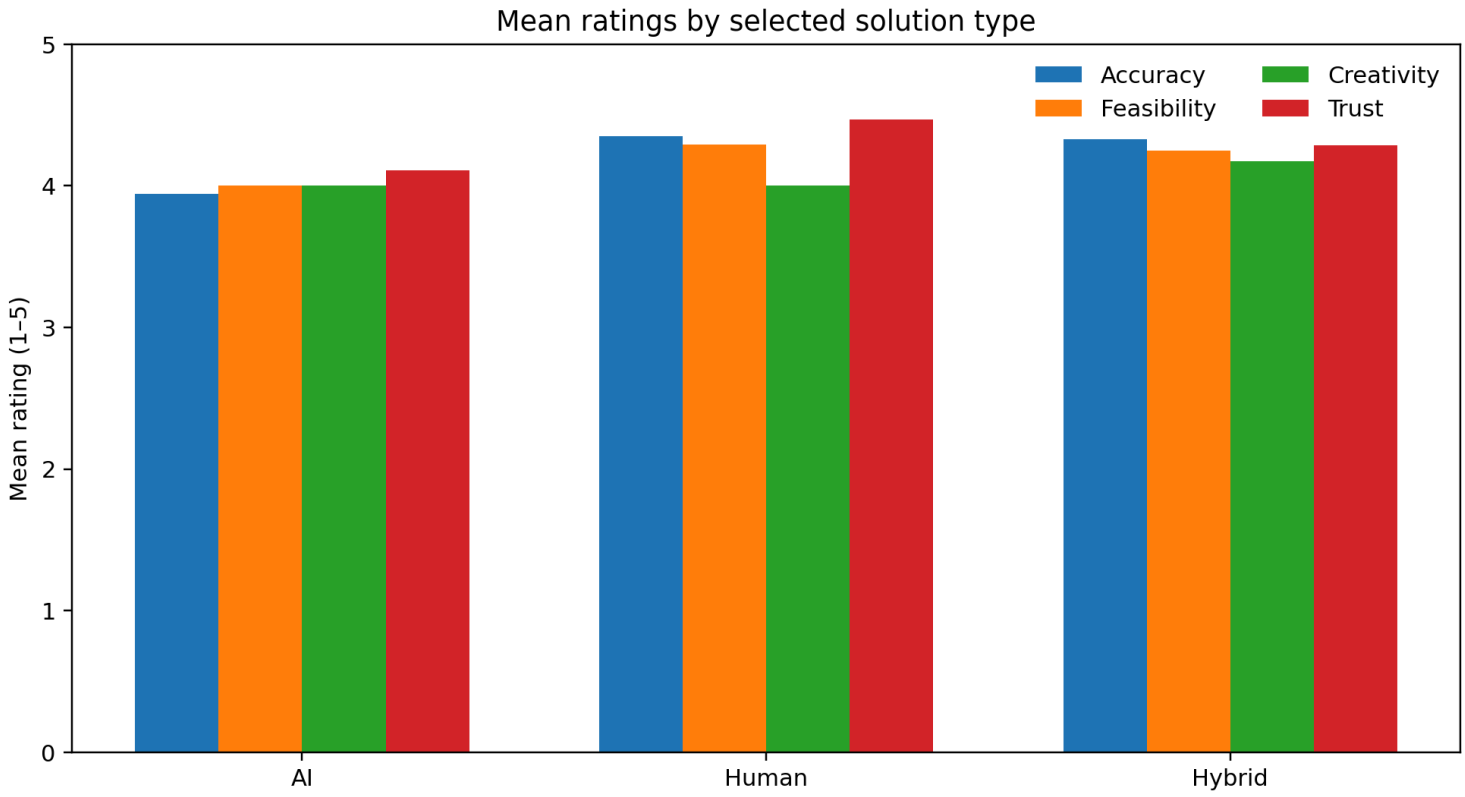


Figure 3

Mean ratings for accuracy, feasibility, creativity, and trust by selected solution type.

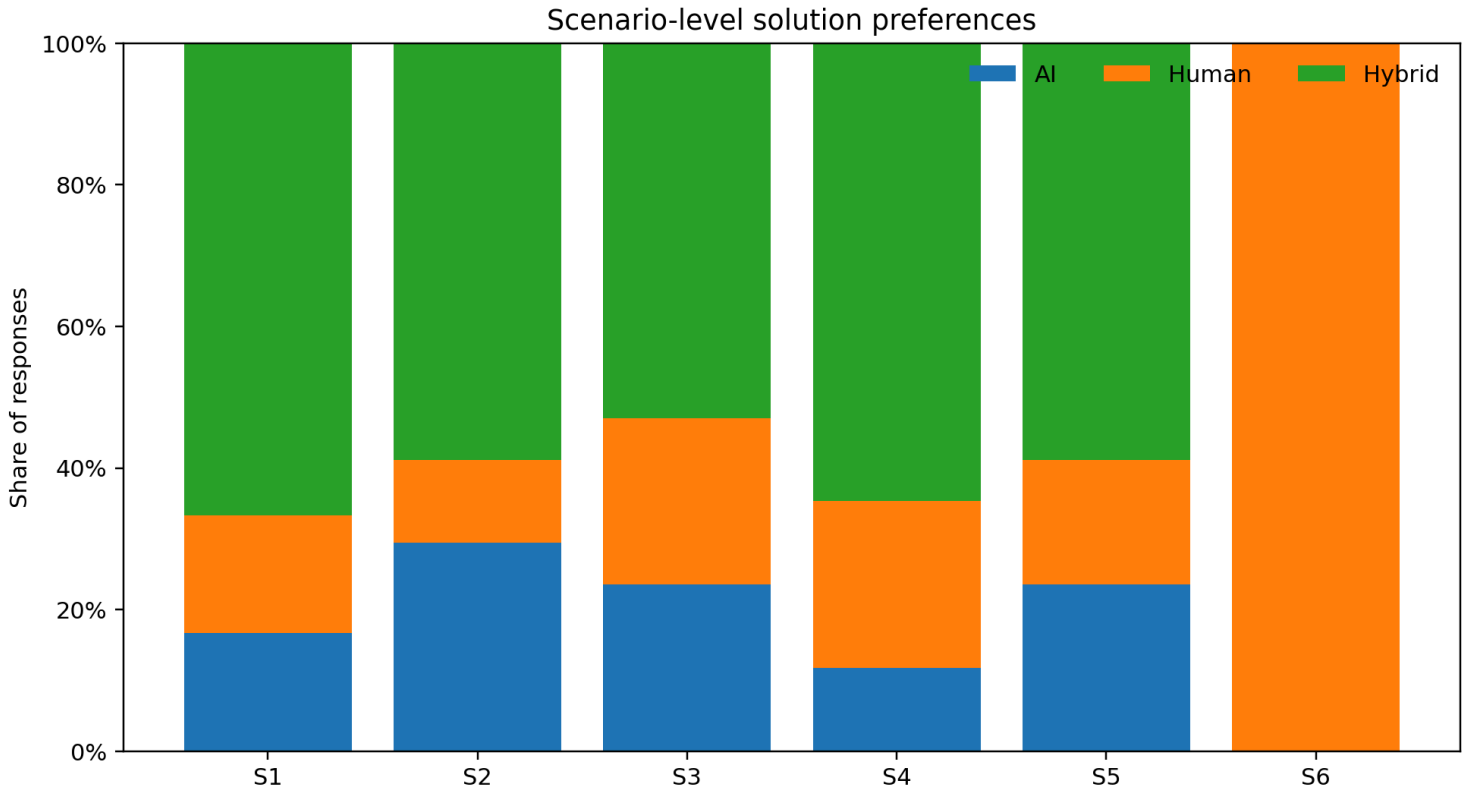


Figure 4

Relative share of AI, human, and hybrid selections within each scenario

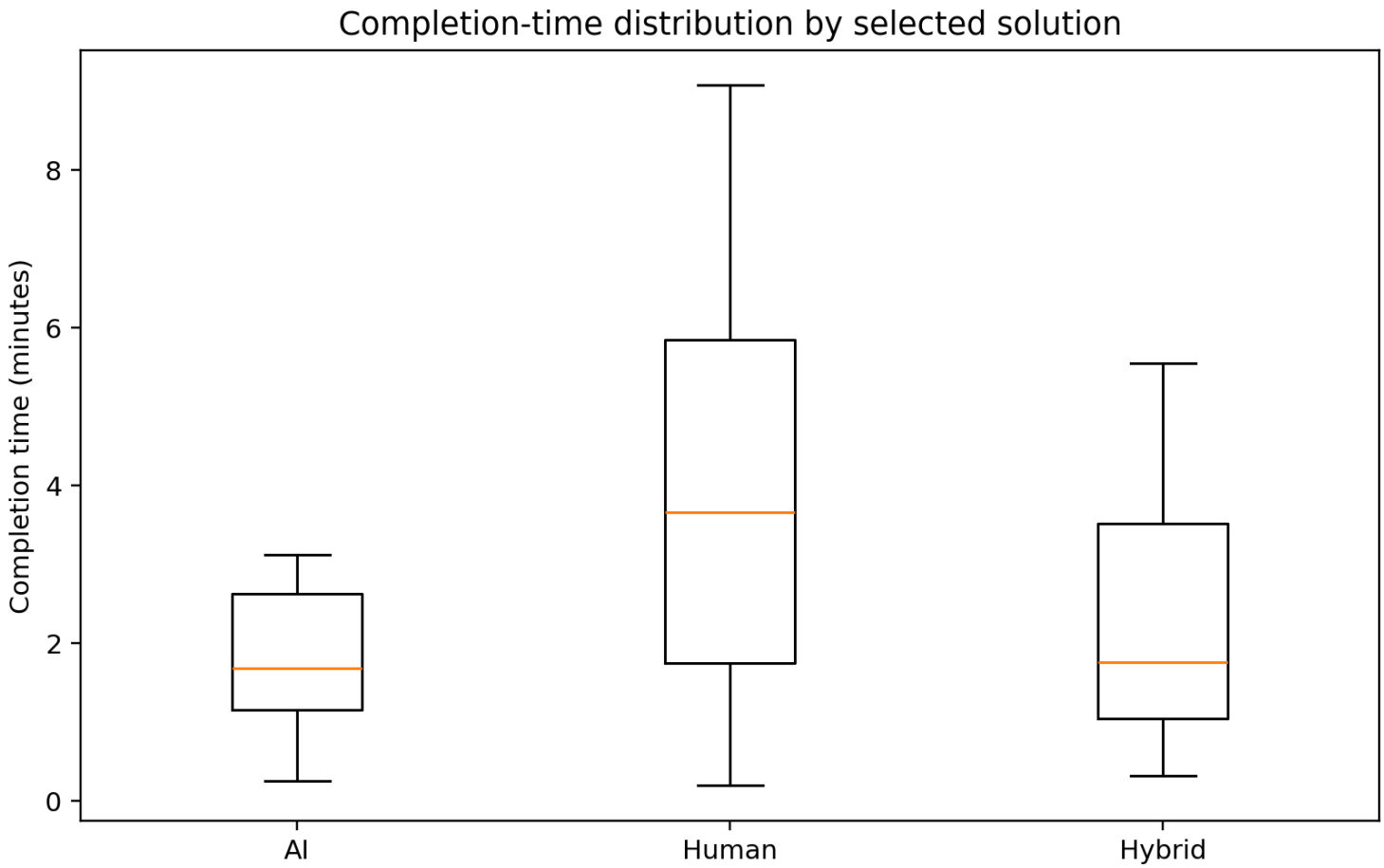


Figure 5

Completion-time distributions by selected solution type (outliers suppressed for readability).