

Performance, Suggestion Adoption, and Trust in Iterative AI-Assisted Ideation

Performance, Adoption, and Trust in AI-Assisted Ideation

Jiacheng Sun; Ting Liao*

Department of Systems Engineering, Stevens Institute of Technology, 1 Castle Point Terrace, Hoboken, NJ 07030, USA, tliao@stevens.edu, ORCID: 0000-0003-1786-8706

AI-assisted ideation is increasingly used in early-stage design, but little is known about how repeated interaction shapes performance, suggestion adoption, and trust. This study analyzes 219 ideation sessions from 73 users of the AI Design Assistant (AIDA) using mixed-effects models on turn-level and session-level records. Performance improved gradually across turns rather than through discrete gains tied to specific AI responses. Adoption coincided with local upturns but did not predict sustained improvement or trust. Trust was associated with affective features of AI responses and moderated by user input length and lexical diversity. Performance, adoption, and trust are better treated as related but distinct processes.

CCS CONCEPTS • Human-centered computing • Human computer interaction (HCI) • HCI design and evaluation methods • User studies

Additional Keywords and Phrases: AI-assisted ideation, design ideation, human-AI collaboration, trust, longitudinal analysis

1 INTRODUCTION

Large language models support early-stage design ideation [2, 15]. Since ideation develops across repeated turns [11], endpoint evaluations miss how collaboration changes over time [13, 14]. We ask how performance evolves (RQ1), whether gains are gradual or discrete (RQ2), whether adoption predicts improvement [1, 16] (RQ3), and how user input moderates associations between AI-response features and trust [6, 8] (RQ4).

2 METHOD

Data came from AIDA, a web-based chatbot supporting iterative design ideation [9, 10]. As illustrated in Figure 1, a session comprises repeated turns, each with a user input, AI response, and composite score. The hierarchical dataset included 73 users, 219 sessions, and 537 valid turns.

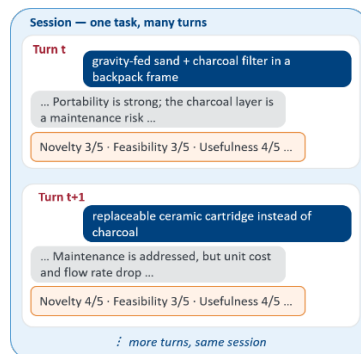


Figure 1: Example of one turn in the AIDA ideation workflow.

Measures were composite score, turn-to-turn change, and a 75th-percentile upturn indicator [5]. BERT similarity measured adoption [4]. AI-response features were length, structuredness, GoEmotions empathy [3], and VADER sentiment

[7]; user features were word count and lexical diversity. Trust and usefulness came from post-interaction ratings. Standardized models included user random intercepts..

3 RESULTS

Performance development (RQ1). Turn index positively predicted score across 537 turns ($\beta = .222, p < .005$). Means rose from 2.81 at turn 1 to 3.37 at turn 2 and 3.46 at turn 3, then flattened as observations declined.

Gradual versus discrete change (RQ2). Across 300 adjacent-score turns, no AI-response feature predicted score change or any of 113 upturns. Only user word count predicted change (coef. = .187, $p < .005$), linking short-term improvement to user elaboration.

Adoption (RQ3). Adoption did not predict next-turn change (135 turns) but did predict upturns (318 turns, coef. = .339, $p < .05$). At the session level, neither adoption nor improvement predicted trust; empathy did (coef. = .153, $p < .005$).

Subjective evaluation (RQ4). Empathy increased trust (coef. = .183, $p < .005$), while structuredness decreased it (coef. = -.211, $p < .05$). Sentiment interacted with word count and lexical diversity (both $p < .05$), shifting from positive for shorter, less diverse inputs to negative for longer, more diverse inputs. Only empathy predicted usefulness (coef. = .113, $p < .005$).

4 DISCUSSION AND CONCLUSION

Improvement did not come from a few especially effective AI responses. It accumulated as users developed and revised their ideas across turns, consistent with process-oriented accounts of human-AI co-creation [11, 14]. The pattern suggests that improvement is carried by the cumulative work of revision rather than by any single well-formed suggestion, and that evaluations built around one AI output may understate what iterative support contributes.

Adoption behaved as a distinct process: higher semantic uptake coincided with local upturns but did not predict next-turn improvement or session-level trust. It should not be used as a proxy for effectiveness [1, 16]. Because uptake predicted local upturns without predicting sustained gains, it appears to mark moments when a suggestion happened to fit the current design state rather than a general mechanism of improvement. Optimizing systems for uptake alone risks improving a signal that does not carry through to outcomes.

Trust followed a separate logic, associated with how the AI communicated, particularly empathic cues, rather than with adoption or improvement [12]. Affective cues carry no fixed value: positive sentiment supported trust for short, less varied inputs but reversed when users elaborated at length, suggesting trust depends on the fit between AI tone and context. This did not extend to perceived usefulness, so the two are not interchangeable. For system design, affective framing is therefore better treated as context-sensitive than as uniformly beneficial, and trust and usefulness warrant separate instrumentation in evaluation.

Limitations include few late-turn observations, reliance on the AIDA framework rather than human ratings, and mainly student participants. Even so, performance, adoption, and evaluation are best treated as distinct processes and studied longitudinally.

ACKNOWLEDGMENTS

The authors used ChatGPT to support language editing, manuscript formatting, and revision of author-written text. All research design, data analysis, interpretation, references, and final manuscript content were checked and approved by the authors.

REFERENCES

- [1] Ashkinaze, J. et al. 2025. How AI Ideas Affect the Creativity, Diversity, and Evolution of Human Ideas: Evidence From a Large, Dynamic Experiment. *Proceedings of the ACM Collective Intelligence Conference* (New York, NY, USA, Aug. 2025), 198–213.
- [2] Chen, P. et al. 2025. CoExploreDS: Framing and Advancing Collaborative Design Space Exploration Between Human and AI. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama Japan, Apr. 2025), 1–20.
- [3] Demszky, D. et al. 2020. GoEmotions: A Dataset of Fine-Grained Emotions. arXiv.
- [4] Devlin, J. et al. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv.
- [5] Eton, D.T. et al. 2020. Known-groups validity and responsiveness to change of the Patient Experience with Treatment and Self-management (PETS vs. 2.0): a patient-reported measure of treatment burden. *Quality of Life Research*. 29, 11 (Nov. 2020), 3143–3154. <https://doi.org/10.1007/s11136-020-02546-x>.
- [6] Henrique, B.M. and Santos, E. 2024. Trust in artificial intelligence: Literature review and main path analysis. *Computers in Human Behavior: Artificial Humans*. 2, 1 (Jan. 2024), 100043. <https://doi.org/10.1016/j.chbah.2024.100043>.
- [7] Hutto, C. and Gilbert, E. 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*. 8, 1 (May 2014), 216–225. <https://doi.org/10.1609/icwsm.v8i1.14550>.
- [8] Klingbeil, A. et al. 2024. Trust and reliance on AI — An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior*. 160, (Nov. 2024), 108352. <https://doi.org/10.1016/j.chb.2024.108352>.
- [9] Mathews Renji, N. and Liao, T. 2025. Designing Trust: How System-Level Design Attributes Shape AI-Assisted Design Ideation. *Volume 4: 22nd International Conference on Design Education (DEC); 30th Design for Manufacturing and the Life Cycle Conference (DFMLC); 37th International Conference on Design Theory and Methodology (DTM)* (Anaheim, California, USA, Aug. 2025), V004T06A042.
- [10] Renji, N.M. et al. 2024. Augmenting Engineering Design With AI: Introducing the AI Design Assistant (AIDA). *Volume 6: 36th International Conference on Design Theory and Methodology (DTM)* (Washington, DC, USA, Aug. 2024), V006T06A016.
- [11] Shen, H. et al. 2025. IdeationWeb: Tracking the Evolution of Design Ideas in Human-AI Co-Creation. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, Apr. 2025), 1–19.
- [12] Sun, Y. and Wang, T. 2026. Be Friendly, Not Friends: How LLM Sycophancy Shapes User Trust.
- [13] Vaccaro, M. et al. 2024. When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*. 8, 12 (Oct. 2024), 2293–2303. <https://doi.org/10.1038/s41562-024-02024-1>.
- [14] Walton, S.P. et al. 2026. From Metrics to Meaning: Time to Rethink Evaluation in Human–AI Collaborative Design. *ACM Trans. Interact. Intell. Syst.* 16, 1 (Jan. 2026), 3:1–3:31. <https://doi.org/10.1145/3773292>.
- [15] Wang, N. et al. 2025. Exploring creativity in human–AI co-creation: a comparative study across design experience. *Frontiers in Computer Science*. 7, (Sept. 2025). <https://doi.org/10.3389/fcomp.2025.1672735>.
- [16] Yang, K. et al. 2024. Ink and Algorithm: Exploring Temporal Dynamics in Human-AI Collaborative Writing. arXiv.