

Beyond Adoption: How Adaptive AI Feedback Shapes Trust and Design Improvement

Ting Liao, Assistant Professor

Human-AI Interaction (HAI) Design Lab
Department of Systems Engineering
Stevens Institute of Technology

AI4SE & SE4AI Workshop 2026

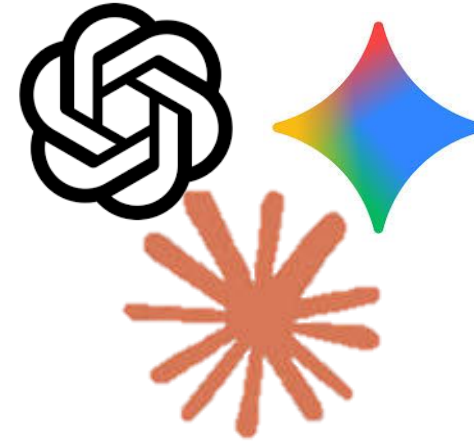
Calibrating Trust Through AI's Waves



Stanford Driving Simulator Audio Engine



Voice Agents
(Amazon Alexa)



Foundational Models

2016
Autonomous driving
Physical simulator
Assistive

2026
Large language models
Virtual interaction
Collaborative

How do human users form trust in AI, similarly or differently than they do for other humans?

How can we, as engineers and design researchers, facilitate the process for calibrated trust by formulating (new) design principles and mental models?

Unpacking Trust from a System Perspective

- SE4AI approach to understand trust in human-AI interaction
 - Trust is a dynamic, socio-technical construct influenced by system performance, interface design, agent behaviors, and individual user characteristics [1,2].
- Modular architecture to test design variants
 - It requires a testbed to assess the impact of chosen factors on user experience and ideation outcomes.
- AI in a proactive, co-creating role
 - Attention is now being paid to how AI systems can augment human creativity as cognitive collaborators, particularly in early-stage design [3-6].



Building AIDA: An Evolutionary Approach

Method: System Development

- The development of the **Artificial Intelligence Design Assistant (AIDA)** follows an evolutionary approach.

Phase 1: Baseline Model Development

Utilized a fine-tuned RoBERTa model trained on 672 design ideas for sustainable washing machines from a previous ideation study [7].

Phase 2: Integration of Modern LLMs

Integrated GPT-4o (OpenAI) and LLaMA3 (Meta) as foundational models for rich, multi-turn conversational understanding.

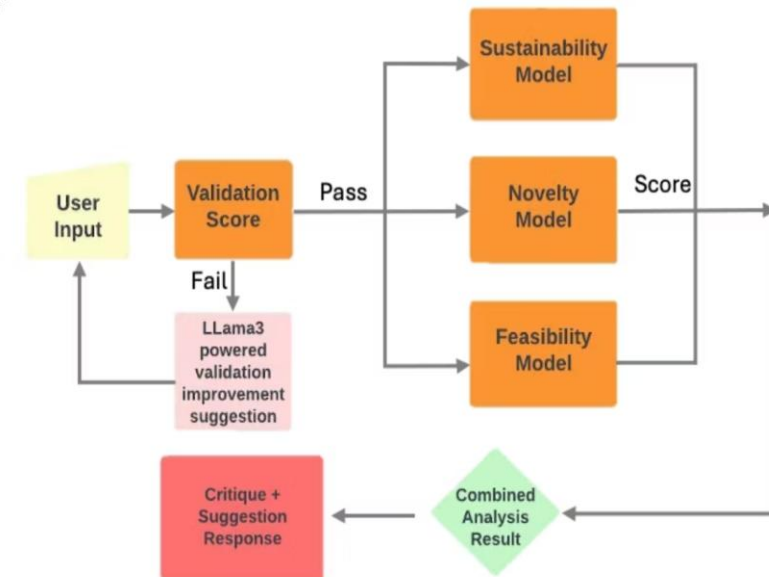
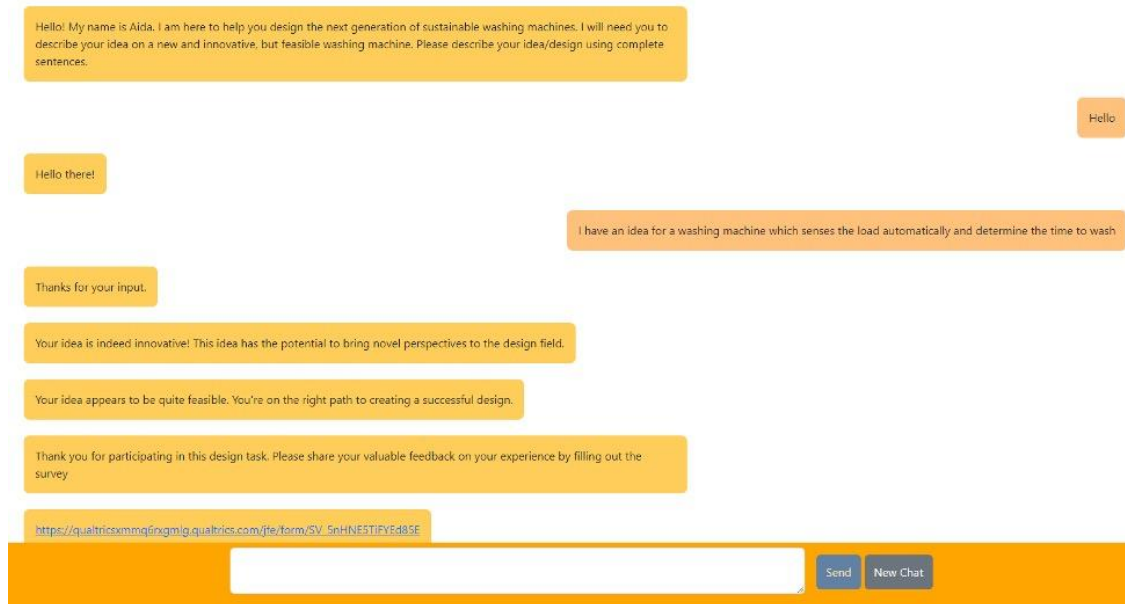
Phase 3: Adaptive Conversation Management

Focused on conversation design and full session memory, enabling fully adaptive interactions tailored to the ongoing context.

Building AIDA: Interface & Evaluation Framework

Method: System Development

- AIDA prompts design problems and assesses user-generated ideas on validity, feasibility, novelty, and sustainability.



Modeling System Aesthetics, Performance, and Behavior

- RQ1. How do AI system design attributes influence user trust in creative design tasks?
- RQ2. Which design attributes are relatively more important in shaping trust?
- RQ3. How does trust dynamically evolve with collaborative performance?

Version	Aesthetics	Feedback	Behavior
V1	Humanoid Avatar	Detailed Feedback	Empathetic Response
V2	Abstract Icon	Detailed Feedback	Non-Empathetic Response
V3	Humanoid Avatar	Scores and Overview	Non-Empathetic Response
V4	Abstract Icon	Scores and Overview	Empathetic Response

A Four-step Experiment, Looped for Four Design Versions

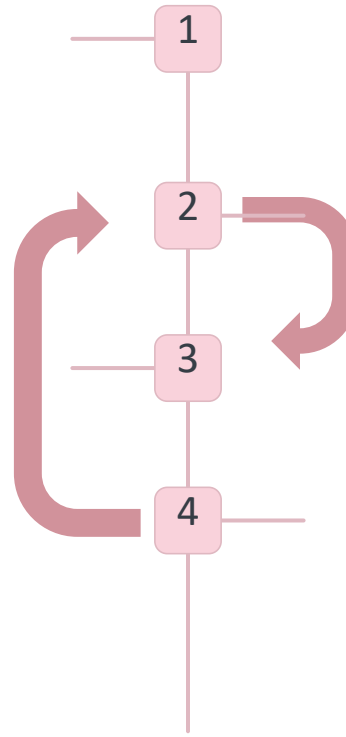
Experimental Procedure

Step I: Pre-interaction Survey

Assessed demographics, initial attitudes, and expertise.

Step III: Design Iteration

Participants generated and revised concepts based on critiques.



Step II: Intro. of Prompt

Introduced a different design prompt.

Step IV: Post-Int. Survey

Filled out a survey about the chatbot they just used.



Exposing the Handoff to Human Judgment

Results and Data Preparation

- **AIDA's role**
 - evaluates, critiques, suggests
- **User's role**
 - decides what to carry forward
- **Data**
 - 73 users · 219 sessions · 531 scored turns

Session — one task, many turns

Turn t

Initial idea text ...

... Feedbacks ...

Novelty 3/5 · Feasibility
3/5 · Usefulness 4/5 ...

Turn t+1

Revised idea text ...

... Feedbacks...

Novelty 4/5 · Feasibility
3/5 · Usefulness 4/5 ...

: More turns, same session



Measuring Three Processes Separately

Analysis: Metrics of Interest

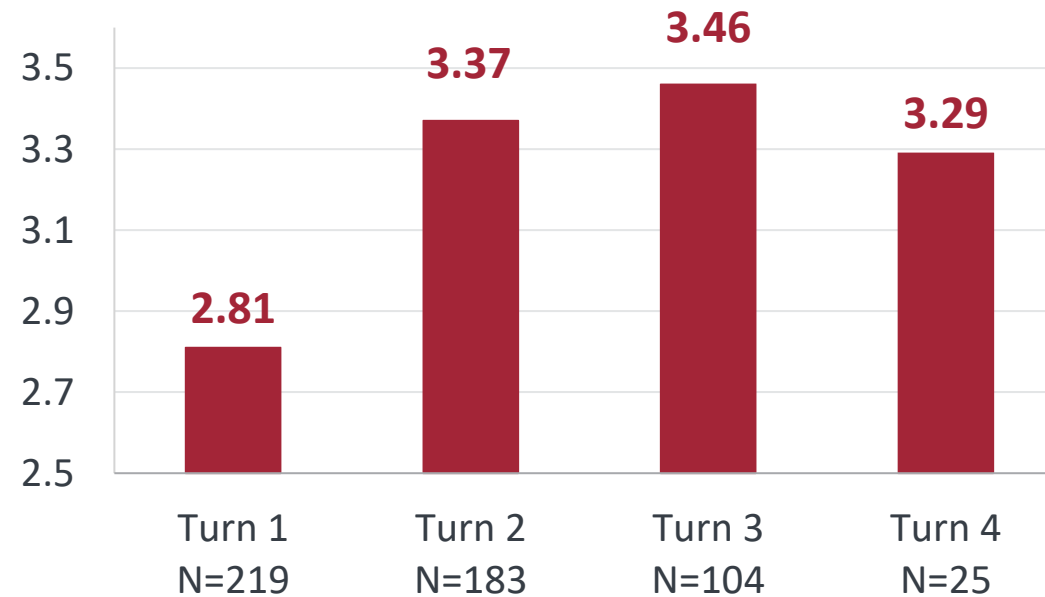
Performance is evaluated turn by turn, while adoption and trust are assessed independently.

- **Performance — the outcome, scored every turn**
 - Composite score of the user's design state logged after each AI response; turn-to-turn change, upturn = top quartile [8,9]
- **Adoption proxy — the behavior, not the outcome**
 - Semantic similarity between the user's next input and the previous AI response [10,11]
- **Trust — the perception, self-reported**
 - Post-interaction ratings, averaged within a session [12–16]
- Related but distinct: uptake and trust can be high while the score stays flat



Analysis | Tracking Performance Gains Across Turns

- Mean score rose 2.81 → 3.46 by Turn 3; linear mixed-effects model, turn index $\beta = .222$, $p < .005$, random intercept for user.



Every participant completed at least two turns per session; any turns beyond that were optional.

Analysis | Linking Adoption to Performance Change

Do gradual performance gains reflect adoption of the AI's advice?

- Method — measuring uptake, then asking what it predicts
 - Mixed-effects models with a random intercept per user: adoption $\rightarrow \Delta\text{score}_{t+1}$, and adoption $\rightarrow \text{upturn}$
- Result — adoption coincides with local progress, but not next-turn gain
 - Upturn: coef = .339, $p < .05$; Bayesian posterior $M = .432$, $SD = .173$, $z = 2.50$
 - Δscore_{t+1} : adoption n.s.; only baseline score $-.152$ and AI sentiment $-.187$, both $p < .005$

Analysis | Testing If Trust Tracks Performance Change

If not adoption, does trust explain the gains?

- Method — model the user's own rating at both turn and session levels
 - Empathy = modeled probability the reply conveys caring; structuredness = reply laid out in lists or sections
- Result — trust followed how the AI wrote, not what improved
 - Session level: mean adoption and total improvement both n.s.; AI empathy .153, $p < .005$
 - Turn level: empathy .183 ($p < .005$) was associated with higher trust, structuredness $-.211$ ($p < .05$) was associated with lower trust



Analysis | Linking AI Tone to Trust Across User Styles

Does a warmer tone increase trust for everyone?

- Method — let the user's own input moderate the AI's tone
 - User input described by word count and lexical diversity (unique words / total words); same model refit for perceived usefulness
- Result — the effect of tone flips with the user
 - AI sentiment $\times$ user word count = $-.305$, $p < .05$
 - Conditional effect: $+.394$ for terse users (-1 SD) $\rightarrow -.042$ at the mean $\rightarrow -.477$ for verbose users ($+1$ SD)
 - Usefulness: empathy stays positive, moderation n.s. — trust and usefulness are not interchangeable



Trust and Adoption Are Not Proxies for Improvement

Conclusions

- Performance improved primarily in the early turns
- Adoption came with local upturns, but not with next-turn gains
- Trust followed how the AI wrote, and the tone effect flipped with the user
- Associational mixed-effects analyses; not causal estimates



From Performance Gains to Calibrated Reliance

Future Implications

- Calibration, not maximum trust
 - Make the contribution visible and the uncertainty legible, so reliance can track evidence
- Measure human judgment, not only the model
 - Evaluate what a system does to the design outcomes people produce with it
- At community scale, the same tension remains
 - Capability vs. judgment; automation vs. augmentation





This research was supported
by NSF Grant 2301846 and
NSF CAREER Grant 2543591.



THANK YOU



Jiacheng Sun



Ting Liao*
tliao@stevens

Stevens Institute of Technology
1 Castle Point Terrace, Hoboken, NJ 07030

9/22/2026

References

1. Rotter, J. B., "Interpersonal Trust, Trustworthiness, and Gullibility," *American Psychologist* 35, 1–7 (1980). DOI: 10.1037/0003-066X.35.1.1
2. Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J. & Parasuraman, R., "A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction," *Human Factors: The Journal of the Human Factors and Ergonomics Society* 53, 517–527 (2011). DOI: 10.1177/0018720811417254
3. Zhang, G., Raina, A., Cagan, J. & McComb, C., "A cautionary tale about the impact of AI on human design teams," *Des Stud* 72, (2021). DOI: 10.1016/j.destud.2021.100990
4. Liao, Q. V., Gruen, D. & Miller, S., "Questioning the AI: Informing Design Practices for Explainable AI User Experiences," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* 1–15 (ACM, 2020). DOI: 10.1145/3313831.3376590
5. Gyory, J. T., Song, B., Cagan, J. & McComb, C., "Communication in AI-assisted teams during an interdisciplinary drone design problem," *Proceedings of the Design Society* 1, 651–660 (2021). DOI: 10.1017/pds.2021.65
6. Bertelsen, L. & Short, A.-R., "State of the Art: A Review of AI Art Generation Methods for Rigorous Design," in *Volume 3B: 49th Design Automation Conference (DAC)* (American Society of Mechanical Engineers, 2023). DOI: 10.1115/DETC2023-116833
7. Mathews Renji, N., and Liao, T., 2025, "Designing Trust: How System-Level Design Attributes Shape AI-Assisted Design Ideation," *ASME IDETC*, p. V004T06A042.
8. Minet, A., Wentzel, D., Raff, S., and Garbas, J., 2024, "Design Thinking in Physical and Virtual Environments," *Technological Forecasting and Social Change*, 207, p. 123596.
9. Mayer, S., and Schwemmler, M., 2025, "The Impact of Design Thinking and Its Underlying Theoretical Mechanisms: A Review of the Literature," *Creativity and Innovation Management*, 34(1), pp. 78–110.
10. Ashkinaze, J., Mendelsohn, J., Qiwei, L., Budak, C., and Gilbert, E., 2025, "How AI Ideas Affect the Creativity, Diversity, and Evolution of Human Ideas," *ACM Collective Intelligence Conference*, pp. 198–213.
11. Beck, J., Eckman, S., Kern, C., and Kreuter, F., 2025, "Bias in the Loop: How Humans Evaluate AI-Generated Suggestions," *arXiv:2509.08514*.



References

12. Schemmer, M., Kuehl, N., Benz, C., Bartos, A., and Satzger, G., 2023, “Appropriate Reliance on AI Advice: Conceptualization and the Effect of Explanations,” IUI ’23.
13. Guo, Z., Wu, Y., Hartline, J. D., and Hullman, J., 2024, “A Decision Theoretic Framework for Measuring AI Reliance,” ACM FAccT, pp. 221–236.
14. Klingbeil, A., Grützner, C., and Schreck, P., 2024, “Trust and Reliance on AI — An Experimental Study on the Extent and Costs of Overreliance on AI,” Computers in Human Behavior, 160, p. 108352.
15. Ng, S. W. T., and Zhang, R., 2025, “Trust in AI Chatbots: A Systematic Review,” Telematics and Informatics, 97(C), p. 102240.
16. Bach, T. A., Khan, A., Hallock, H., Beltrão, G., and Sousa, S., 2024, “A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective,” International Journal of Human–Computer Interaction, 40(5), pp. 1251–1266.

