Poster#: 3

AI for organization: Designing for Cognitive Resilience using the Controlled Exposure Framework

Qingzheng Gao, Shreya Chavan; PI: Prof. Akram Bayat

Northeastern University
Khoury College of Computer Sciences

AI FOR ORGANIZATIONS GRAND CHALLENGE

Research-driven ideas on how AI can supercharge collaboration, coordination, and decision-making in real-world organizations.

HAI Stanford University Human-Centered Artificial Intelligence × Google DeepMind

MOTIVATION

Organizations are adopting AI faster than employees can adapt, causing mistrust, reduced autonomy, and emerging cognitive dependency.

INTERVIEWS & QUESTIONNAIRES

We conducted interviews and questionnaires on role and industry, AI tools used, **data concerns, trust,** skill impacts, changes in thinking, organizational **guidance**, factors shaping comfort or **resistance**, key AI principles, and additional experiences.

USER RESEARCH INSIGHTS

INTERVIEW FINDINGS:
- Ambivalence: efficiency gains vs. worry about **privacy** & **skills**.
- Emerging **cognitive dependency** on AI for judgment.
- **Policy gap:** unclear rules → self-censorship or AI avoidance.

QUESTIONNAIRE FINDINGS:
- Employees fear **privacy loss, surveillance, and loss of autonomy.**
- Organizations rarely explain **data use, monitoring, or override paths**, undermining trust.
- Heavy reliance on AI risks **skill atrophy** and **over-dependence** in high-stakes, low-frequency situations.

KEY INSIGHT: Trust depends more on **transparency, control, and recourse** than on raw accuracy.

PROBLEM: We lack controlled exposure guidelines that balance AI efficiency with human judgment, skill retention, and trust.

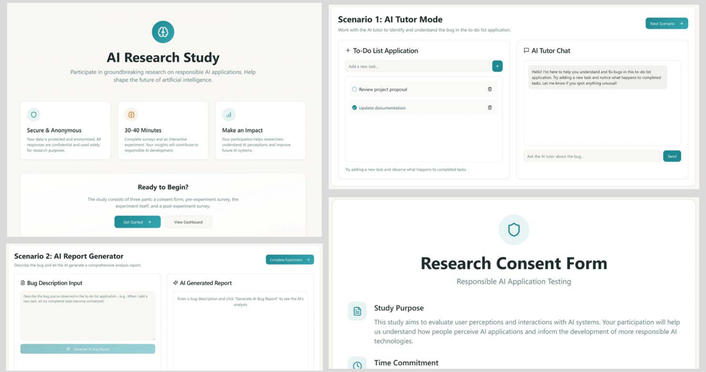
PROBLEM DEFINITION

4 Research Questions:
1) What drives employee **trust** or **fear** toward workplace AI?
2) How do **privacy** and **transparency** affect AI acceptance?
3) How do **communication** and **governance** influence trust?
4) How does **organizational support** shape long-term AI use?

EXPERIMENT DESIGN

Controlled Experiment: Between-subjects study with CS students reviewing a crisis AI spec.
- **Condition A (Control): Manual Authoring Baseline** – draft review from scratch (no AI text).
- **Condition B (Experimental): Co-Generation & Verification** – start from flawed AI draft + reflection prompt; verify and revise.

Compare **flaw detection** and **unaided quiz performance**.

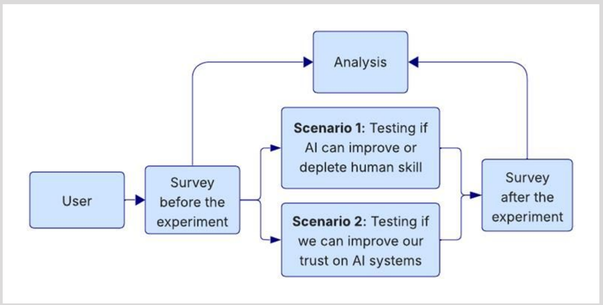


IRB COMPLIANCE

IRB compliance was a core design constraint, utilizing a realistic but fictional **Crisis Resilience Triage Engine (CRTE)** scenario with minimal deception, privacy safeguards, and clear risk communication. We developed the full IRB package: AI systems in HSR form, protocol, **information sheet**, instructions and prompts, task script, and pre/post-task surveys.

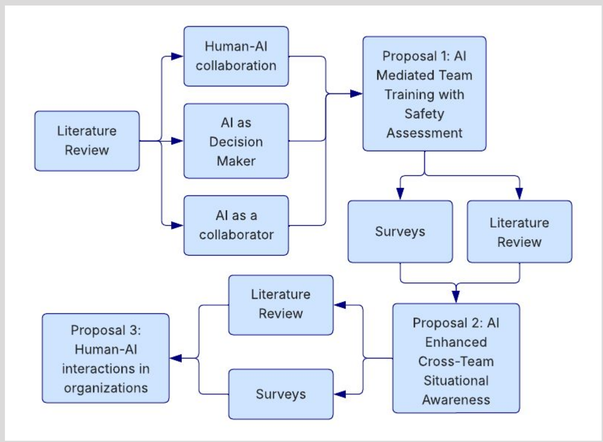
Trust hinges on Transparency and Control, NOT just Accuracy. Organizations must design for Recourse.

EXPERIMENT PROCEDURE



NEXT STEPS

- Complete the prototype for conducting experiment
- Conduct the experiment in a controlled setting
- Publish the findings of the experiment



REFERENCES

S. Karny, A. Baez, and P. Pataranutaporn, “Neural Transparency: Mechanistic Interpretability Interfaces for Anticipating Model Behaviors for Personalized AI,” Oct. 31, 2025, doi: 10.48550/arXiv.2511.00230.
D. Rice, and F. Bartram, "AI and Workplace Wellness: Trust and Transparency in Practice," Oct 2, 2025.