

Can AI Really Do UX Research?

Three Studies Exploring AI's Strengths, Limitations,
and Need for Human Expertise



Presented by Theresa Wilkinson



Can AI Really Do UX Research?



Agenda

1. Project Overview
2. Research Approach
3. Study 1: Can AI Design a Good UX Study?
4. Study 2: Can AI Analyze UX Evaluation Results?
5. Study 3: Quantitative Analysis
6. Study 3: Human Verification
7. What We Learned Across the Three Studies
8. How the Research Evolved
9. Implications for UX Researchers
10. Final Takeaway

Project Overview

- Evaluated ChatGPT, Claude, Gemini, and Copilot across three UX research studies.
- Each study examined a different application of AI within the UX research process.
- The same inputs and prompts were provided to all four models within each study.
- The studies progressed from research planning to interpretation and quantitative data analysis.
- The goal was to understand where AI can support UX research and where human expertise remains essential.

Research Approach

Study 1 — Can AI Design a Good UX Study?

Evaluated AI-generated research questions, participant criteria, sample sizes, and usability tasks.

Study 2 — Can AI Analyze UX Evaluation Results?

Evaluated how AI synthesized evaluation results into conclusions, implications, limitations, and recommendations.

Study 3 — Can AI Accurately Analyze UX Research Data?

Evaluated whether AI could accurately analyze the same participant-level UX research dataset.

Overall Takeaway

The studies progressively increased the complexity of the work assigned to AI.

Study 1: Can AI Design a Good UX Study?

- All four AI models produced usable starting points for UX research planning.
- Core UX research components were consistently identified.
- The largest differences involved research questions, participant selection, sample size, and usability tasks.
- Sample-size recommendations ranged from 8 to 48 participants and one to six rounds of testing.
- No single AI generated a complete research plan.
- Human expertise was needed to evaluate completeness, practicality, accessibility, and project constraints.

Overall Takeaway

AI was useful for accelerating research planning, but the plans required human review and refinement.

Study 2: Can AI Analyze UX Evaluation Results?

- All four models identified the primary workflows and important participant characteristics.
- Agreement was stronger around what should be tested than how the research should be conducted.
- Sample size, research-question coverage, task coverage, methodology, and supporting rationale varied substantially.
- Accessibility was consistently acknowledged but inconsistently incorporated into actual usability tasks.
- Different models demonstrated different strengths; no single model was strongest across every area.
- Human judgment remained necessary to balance methodological rigor with practical project constraints.

Overall Takeaway

- Comparing multiple AI outputs revealed strengths and gaps that would not necessarily be apparent when using a single model.

Study 3: Quantitative Analysis

- All four models identified at least some important usability findings.
- Quantitative errors, missed findings, and unsupported or inaccurate interpretations were identified.
- Different models produced different types of errors.
- Time-on-task calculations included incorrect or substantially inflated values.
- Incorrect calculations sometimes affected the interpretation of whether participants struggled with a task.
- Reported values had to be checked individually against the source data.

Overall Takeaway

AI-generated quantitative findings could not be assumed to be accurate.

Study 3: Human Verification

- Calculations had to be checked against the original dataset.
- Reported findings had to be reviewed for accuracy.
- Missed findings had to be identified manually. Interpretations had to be checked to determine whether they were supported by the data.
- Poor report organization increased verification time.
- At times, performing the analysis directly would have been faster than verifying the AI output.

Overall Takeaway

The researcher remained the quality control throughout the analysis.

What We Learned Across the Three Studies

- AI consistently produced useful starting points.
- AI performance varied depending on the research task.
- Different models demonstrated different strengths and weaknesses.
- More complex analytical tasks exposed greater limitations.
- Comparing multiple models helped identify gaps and inconsistencies.
- Human expertise became increasingly important as the work moved from drafting to quantitative analysis.

Overall Takeaway

AI was most valuable as a research assistant rather than as a replacement for an experienced UX researcher.

How the Research Evolved

- The project was designed as a series of focused studies rather than one broad evaluation of AI.
- Findings from each study informed what needed to be examined next.
- Prompts became more specific as limitations in AI responses became apparent.
- The evaluation approach evolved from reviewing generated deliverables to verifying AI analysis directly against source data.
- The final study placed greater emphasis on quantitative accuracy, statistical reasoning, and verification.

Overall Takeaway

The methodology evolved as the research revealed new questions about what AI could—and could not—reliably do.

Implications for UX Researchers

- AI can accelerate portions of UX research planning and analysis.
- AI-generated work should be treated as a starting point rather than a finished deliverable.
- Researchers need sufficient expertise to recognize missing information, methodological problems, calculation errors, and unsupported conclusions.
- Quantitative AI analysis should be verified against the source data.
- Multiple AI models can provide different perspectives, but using several models does not eliminate the need for human verification.
- AI provides the greatest value when combined with the experienced researchers' judgment.

Final Takeaway

- Across all three studies, AI demonstrated meaningful potential to support UX research.
- Its strongest value was in accelerating initial work, identifying core research components, and providing useful starting points.
- Its limitations became more significant as the work required deeper analysis, quantitative accuracy, statistical reasoning, and evidence-based interpretation.
- **AI can support UX research, but experienced researchers remain essential for determining whether the output is complete, accurate, practical, and supported by the evidence.**