Case Study: AI assistant heuristic evaluation

Jessica Wood – Sr. UX Researcher

Product background:

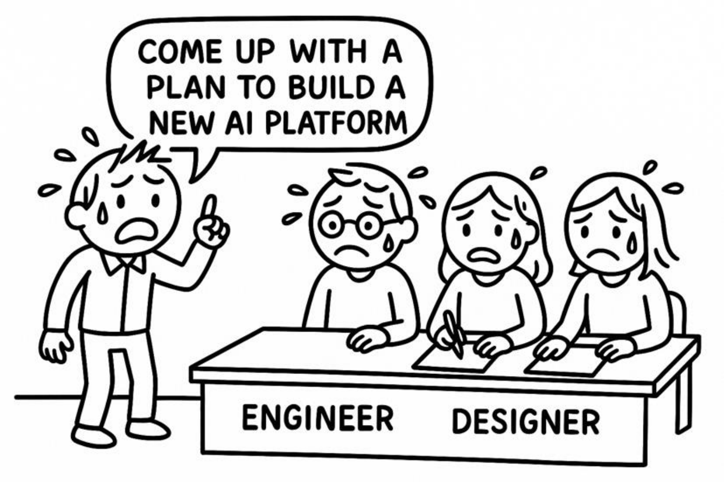
Company X set out to launch an AI-powered learning platform for an upcoming tech conference. The platform was designed to reduce friction in the upskilling process. By adapting to each user’s goals, skill level, and time constraints, the platform aimed to deliver personalized recommendations to empower users to gain certifications, build new skills, and solve technical challenges more efficiently and with greater clarity

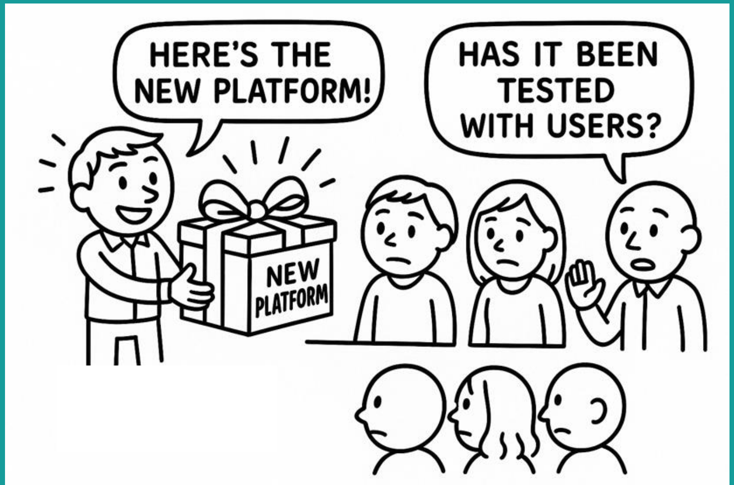
In a nutshell...

- An AI-powered learning platform was built and prepared for launch, but core features had been developed without any real user validation

- With little time remaining before launch, a senior stakeholder raised concerns about potential usability risks during the live demo

- The product manager for the new platform reached out to me and asked what could be done to gather feedback from users to ensure a successful launch experience





Challenges:

Stakeholder Expectations

- Stakeholders and Product Managers expected both deep and actionable insights, which didn’t align with the compressed timeline
- “Actionable” meant different things to different people. Some wanted strategic insights (e.g., unclear value prop), others wanted tactical changes (e.g., button visibility/placement)

Feasibility

- Limited timeframe in which to provide insights prevented in-depth user research to explore strategic questions, usability pain points, and foundational understanding

Product Understanding

- Because the product was still evolving and the team wasn’t always aligned on what was finalized versus experimental, it was challenging to know where to focus research time and resources

Methodology:

Heuristic evaluation/Expert review

To navigate these challenges, I scoped the work as a rapid heuristic evaluation. This method allowed me to provide quick, actionable insights by focusing on the most severe usability risks and prioritizing issues based on user experience impact rather than engineering effort

Research questions

- Which tenets and traps are being violated across the three experiences?
- How do these traps impact the user experience?
- What are the considerations and/or changes that should be made ahead of the product launch and beyond?

Join forces

Reviewed findings with a second Senior UX researcher to ensure alignment on high-priority issues and strengthen confidence in the recommendations

Thinking ahead

After aligning with the team that an expert review would help address immediate usability concerns, I clarified that this would be Phase 1 and proposed a follow-up product walkthrough with 12 users post-launch to:

- Validate the heuristic findings
- Uncover additional insights, including pain points, mental model mismatches, and potential trust issues

Key findings:

Severity 1 blockers prevent task completion or disrupt core workflows
Users are unable to complete essential actions such as saving chat history, editing or saving personalized learning plans, or creating an account directly from the AI interface. These gaps could lead to task abandonment and undermine the platform's effectiveness

Inconsistent UI patterns create friction and user confusion
Irregular use of icons, duplicated prompts, and hidden menu items violate basic usability principles and result in a fragmented experience. Users may struggle to navigate, undo actions, or develop familiarity with core features

Trust and efficiency are undermined by friction points
Missing features like copy buttons, complex feedback flows, and invisible reset options create unnecessary effort. These pain points may reduce confidence, increase cognitive load, and lead users to disengage or turn to alternative solutions

Weak or missing value propositions reduce engagement
Users may not understand what the AI Assistant is, how it benefits them, or why it's worth their time. Without a clear value proposition, the platform risks underuse, even when functionality exists

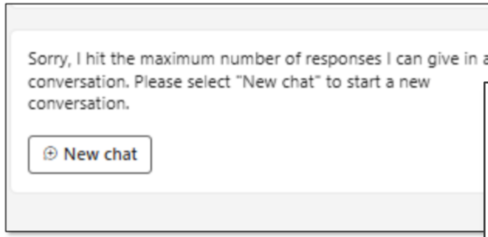
Report examples:

FINDING 10: Sev 1

When a user reaches 10 turns and starts a 'New chat,' previous chats disappear with no way to retrieve them, potentially causing dissatisfaction and reluctance to continue using the feature

- The inability to see view previous chats or have a way to save them violates the Forgiving tenet and triggers the Irreversible Action trap.
- Without being able to save their chats, it causes users to repeat their work which violates the Efficiency tenet and Unnecessary steps trap.
- This lack of retrieval also violates the Protective tenet, and the users fall into the Data Loss trap as their previous chats simply disappear. This may lead to negative emotions, poor brand image, or task abandonment
- Recently users in a previous study expressed wanting to learn from their past engagements and be able to view this history in an easy to find location.

AI Assistant Prompt Sidecar



Sorry, I hit the maximum number of responses I can give in a conversation. Please select "New chat" to start a new conversation.

Considerations

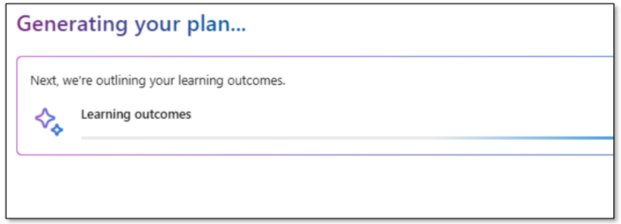
- Consider enabling users to save their chat history or have the system automatically save it for later retrieval.

FINDING 19: Sev 2

The system takes over a minute to generate a personalized plan, which could frustrate users who expect a faster process based on their experiences with other AI systems

- The system can take over a minute to generate a plan, violating the Responsive tenet and triggering the Slow or No Response trap.
- Users may expect the AI personalized learning plan to work as quickly as other forms of AI e.g., ChatGPT. This may be even more frustrating for those who did not provide a lot of detail or background information.

AI Personalized Learning Plan: Plan Generation



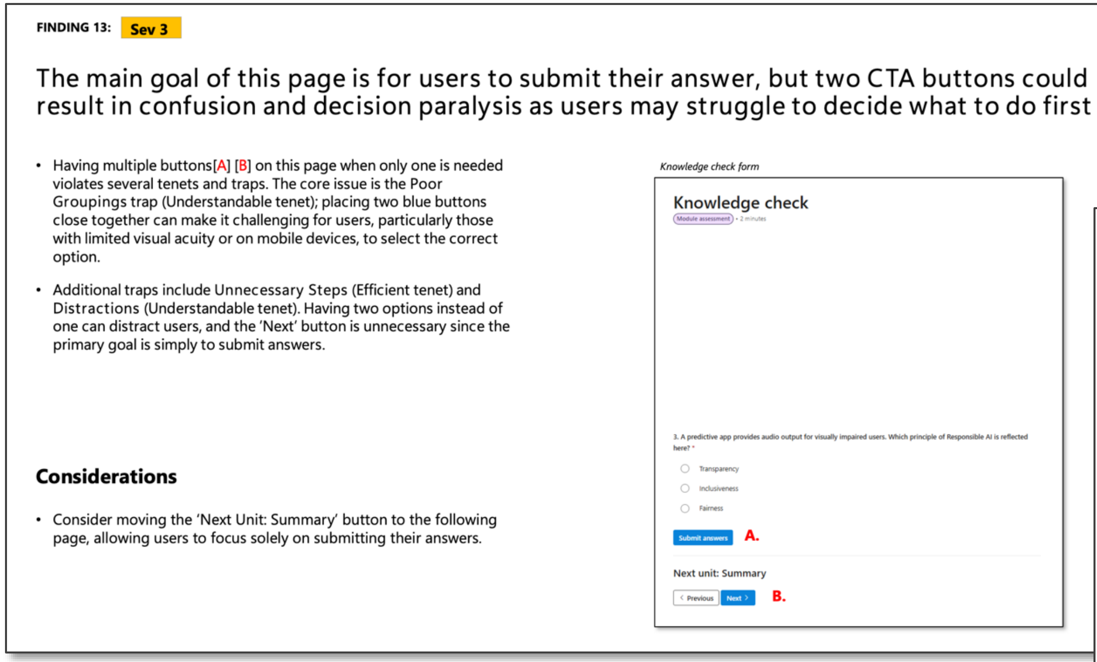
Generating your plan...

Next, we're outlining your learning outcomes.

Considerations

- Consider collaborating with the content strategy team to develop language that includes time-based updates, helping users know what to expect throughout the process.

Report examples:



FINDING 13: Sev 3

The main goal of this page is for users to submit their answer, but two CTA buttons could result in confusion and decision paralysis as users may struggle to decide what to do first

- Having multiple buttons[A] [B] on this page when only one is needed violates several tenets and traps. The core issue is the Poor Groupings trap (Understandable tenet); placing two blue buttons close together can make it challenging for users, particularly those with limited visual acuity or on mobile devices, to select the correct option.
- Additional traps include Unnecessary Steps (Efficient tenet) and Distractions (Understandable tenet). Having two options instead of one can distract users, and the 'Next' button is unnecessary since the primary goal is simply to submit answers.

Knowledge check form

Knowledge check

Module assessment • 2 minutes

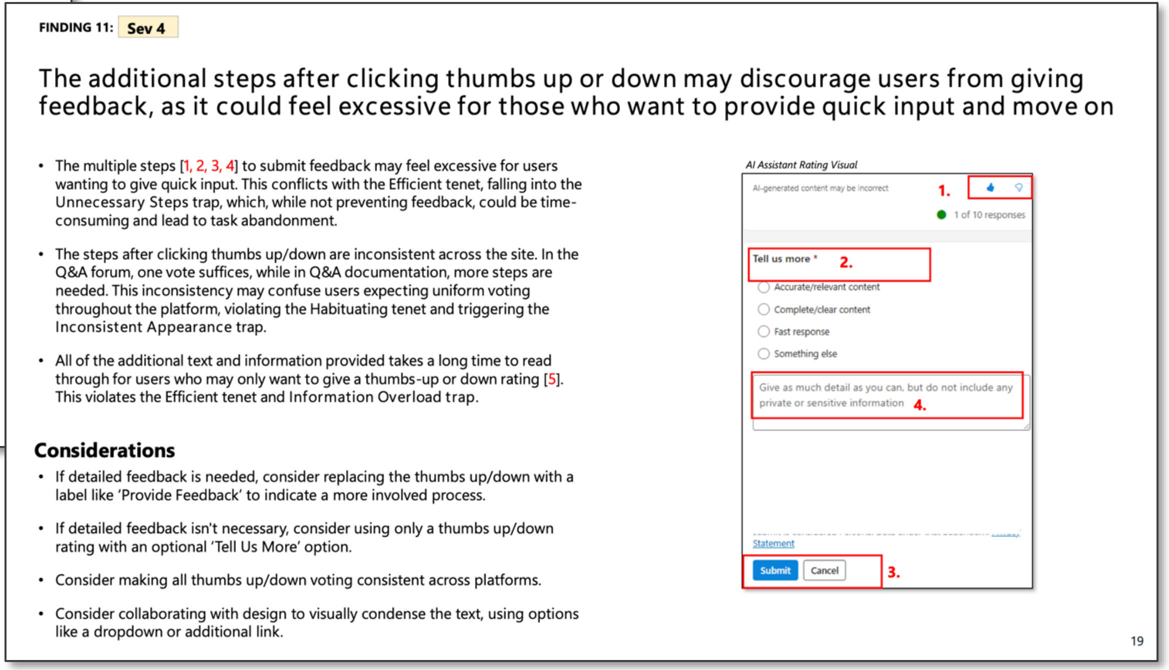
3. A predictive app provides audio output for visually impaired users. Which principle of Responsible AI is reflected here? *

○ Transparency
○ Inclusiveness
○ Fairness

Next unit: Summary

Considerations

- Consider moving the 'Next Unit: Summary' button to the following page, allowing users to focus solely on submitting their answers.



FINDING 11: Sev 4

The additional steps after clicking thumbs up or down may discourage users from giving feedback, as it could feel excessive for those who want to provide quick input and move on

- The multiple steps [1, 2, 3, 4] to submit feedback may feel excessive for users wanting to give quick input. This conflicts with the Efficient tenet, falling into the Unnecessary Steps trap, which, while not preventing feedback, could be time-consuming and lead to task abandonment.
- The steps after clicking thumbs up/down are inconsistent across the site. In the Q&A forum, one vote suffices, while in Q&A documentation, more steps are needed. This inconsistency may confuse users expecting uniform voting throughout the platform, violating the Habituating tenet and triggering the Inconsistent Appearance trap.
- All of the additional text and information provided takes a long time to read through for users who may only want to give a thumbs-up or down rating [5]. This violates the Efficient tenet and Information Overload trap.

AI Assistant Rating Visual

AI-generated content may be incorrect

● 1 of 10 responses

Tell us more *

○ Accurate/relevant content
○ Complete/clear content
○ Fast response
○ Something else

Give as much detail as you can, but do not include any private or sensitive information

Statement

Considerations

- If detailed feedback is needed, consider replacing the thumbs up/down with a label like 'Provide Feedback' to indicate a more involved process.
- If detailed feedback isn't necessary, consider using only a thumbs up/down rating with an optional 'Tell Us More' option.
- Consider making all thumbs up/down voting consistent across platforms.
- Consider collaborating with design to visually condense the text, using options like a dropdown or additional link.

Wrap up:

Expect Candor

Not all stakeholders were familiar with heuristic evaluations or their value in identifying usability risks. Anticipating this, I began the presentation with a brief explanation of Tenets and Traps, the severity rating scale, and noted that I had validated key findings with another senior researcher to strengthen credibility

Boldly Go

I recommended a method that fit the time constraints and allowed us to move quickly. While I didn't anticipate pushback at the time, I stood by the approach and helped guide the team through unfamiliar territory when questions arose

During the review, I welcomed thoughtful challenges, especially from PMs and engineers who asked:
"Isn't this just your opinion?" or "Let's just wait until we have real data since this isn't user validation"

Framing of results

- Heuristic evaluations are predictive, not definitive. They rely on well-established UX principles that help identify likely friction points before users interact with the product
- I acknowledged the limitations. This method does not replace direct user research, but it is a low-effort, high-value approach that helps teams prioritize what to test, fix, or monitor after launch.
- I reiterated that heuristic evaluation is meant to complement user data, not to substitute it

Measuring success

The evaluation helped the team make confident tradeoffs. Out of 21 issues found: Three Sev 1 and 10 low-effort issues (Sev 2-4) were addressed pre-launch, while more complex items requiring additional engineering were flagged for post-launch