

WORKSHOP

# AI as a UX Research *Assistant.*

What helps, what hurts, and what still needs you.

---

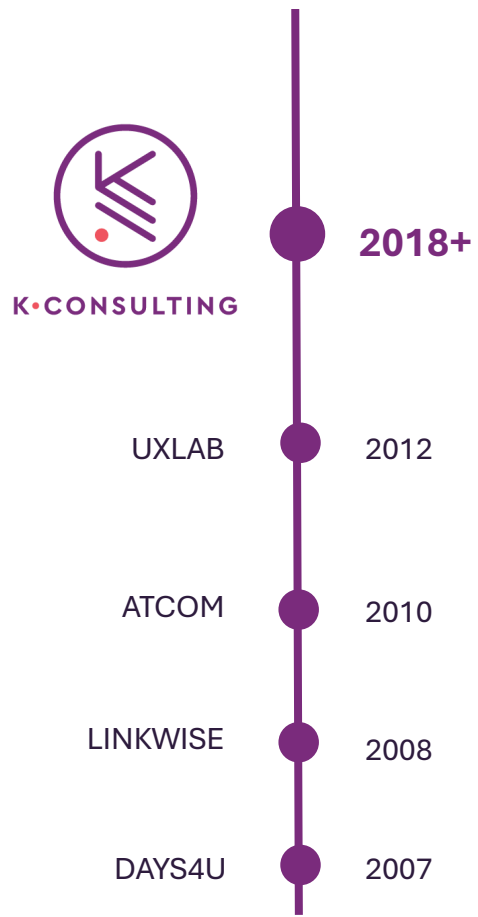
**Katerina Karagianni**  
*UX & CX Consultant & Researcher*



WHO AM I

# UX Consultant & Researcher

*15+ years of experience*



## AGENDA

# What we'll *cover*.

- 01** Intro
- 02** What LLMs actually do
- 03** The AI-enabled UXR workflow
- 04** From notes to insights: using AI responsibly
- 05** Cognitive offloading & the future of expertise

SECTION ONE

# *01*

---

**Intro**

## INTRO

# Would you *trust* this AI summary?

“Participants consistently expressed frustration with the onboarding experience, describing the process as overly complex and emotionally draining. Most users indicated that unclear navigation reduced trust in the platform early in the journey.

Users also showed a strong preference for personalization features, suggesting a clear opportunity to improve long-term engagement through AI-driven onboarding and adaptive experiences.”

## INTRO

# Let's look *closer*.

Overstated  
certainty

Unsupported  
causality

Too generic?

“Participants consistently expressed frustration with the onboarding experience, describing the process as overly complex and emotionally draining. Most users indicated that unclear navigation reduced trust in the platform early in the journey.

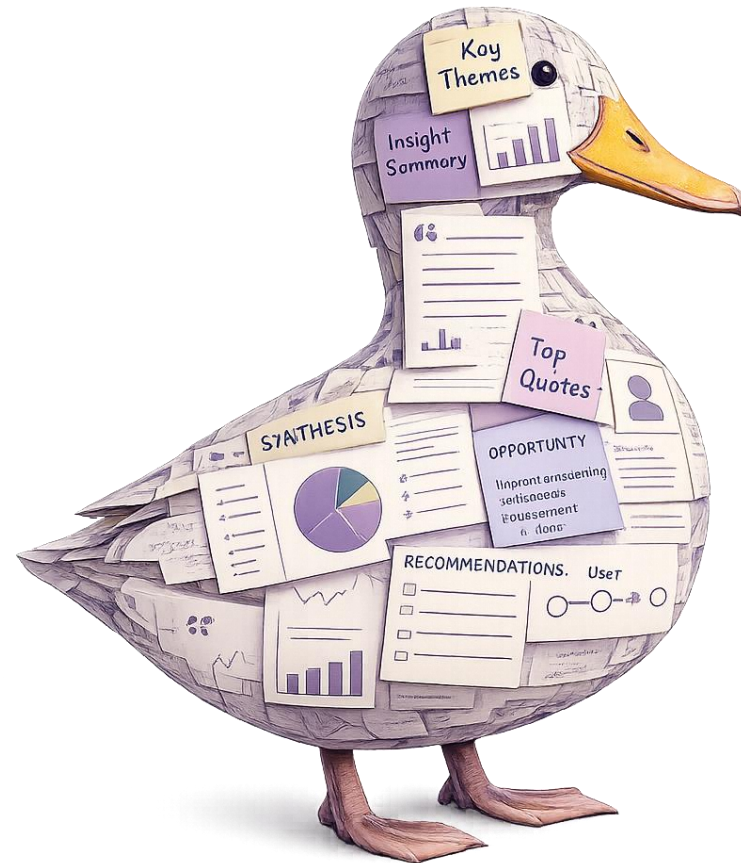
Too  
coherent?

Users also showed a strong preference for personalization features, suggesting a clear opportunity to improve long-term engagement through AI-driven onboarding and adaptive experiences.”

Premature  
solutioning

Evidence  
trail?

# If it *looks* like research...



...we should still be skeptical.

INTRO

# Fluency is not *evidence*.

LLMs produce language that sounds:

coherent • persuasive • authoritative • plausible

BUT

**Plausibility is not the same as truth**

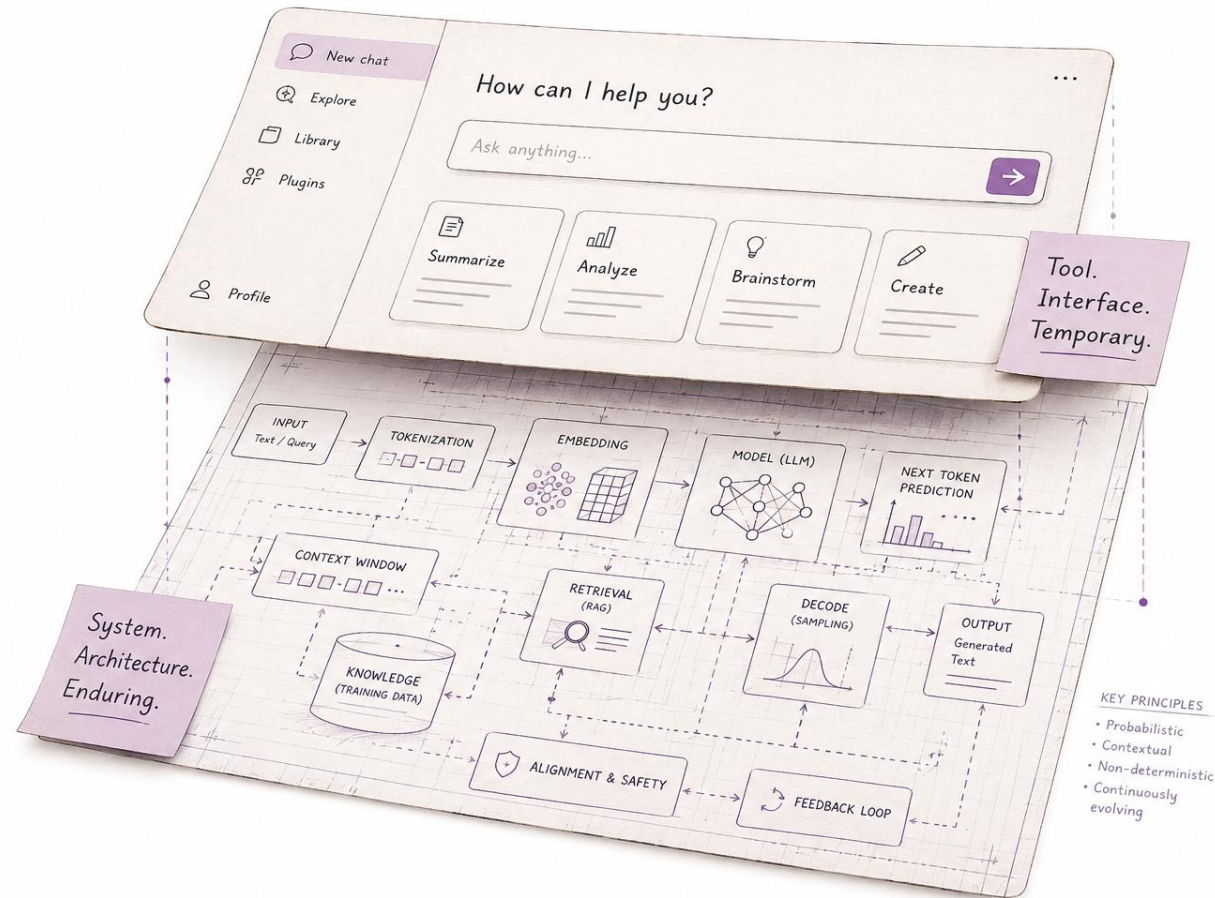
## OUR RESPONSIBILITY



As researchers, we are trained to **question methods, evidence, validity, and bias.**

To use AI meaningfully in UXR, we need to understand **what these systems *actually* are.**

# Tool literacy vs *AI literacy*



This workshop is not about tools, but about understanding the **system beneath the interface**.

SECTION TWO

# 02

---

**What LLMs actually do**

# The mental model *problem*



LLMs do **not work** the way many people intuitively imagine.

# Predictive language at *massive scale*



LLMs are much closer to an **autocomplete system** on an unimaginable scale, than to a human thinker reasoning through ideas.

WHAT LLMS ACTUALLY DO

# *Probabilistic* vs **Deterministic** systems

PROBABILISTIC

---



**Same input, *many plausible outputs.***

VS

DETERMINISTIC

---



**Same input, *same output.***

# One prompt.

“I mean... it works, mostly. Once you know where everything is, it’s fine. But the first time was frustrating. I kept worrying I’d click the wrong thing. I actually liked that there were many options, but it also felt overwhelming at first.

My colleague had no problem with it though — she said it was intuitive immediately.

I probably wouldn’t stop using it because of this, but I did avoid some features because I wasn’t confident I understood them.”

Summarize the key user insights from this interview excerpt.

# One prompt. *Multiple analyses.*



1. Learnability, not usability. The product is "fine" once internalized; friction concentrates at first use. That points to an onboarding or discoverability gap, not a core IA or interaction failure.
2. Fear of irreversible action. "Worried I'd click the wrong thing" suggests weak affordances, unclear consequences, or missing undo/confirmation patterns. Worth probing what specifically triggered the worry, since the source matters (destructive actions look the same as safe ones? Unclear labels? No preview?).
3. Ambivalence about feature density. She values the optionality but the entry experience overwhelms. Standard power vs. progressive disclosure tradeoff. The product isn't necessarily too feature-rich; the surfacing strategy may be wrong.
4. Self-reported feature avoidance. This is the most consequential line in the excerpt and the one most likely to be underweighted. She says she won't stop using the product, but she has already shrunk her usage surface to the parts she trusts. That is reduced value extraction, which over time correlates with lower stickiness and lower willingness to pay or advocate, even if it doesn't show up as churn.



## 1. High Learning Curve vs. Long-Term Utility

The interface is **functional but not "walk-up-and-use" friendly**. While the user eventually became comfortable ("once you know where everything is, it's fine"), the initial barrier to entry was high enough to cause significant friction.

## 2. The "Paradox of Choice"

The product suffers from a classic case of over-choice. While the user appreciated the depth of functionality, the sheer volume of options led to **cognitive overload**. Instead of feeling empowered by the features, the user felt overwhelmed.

## 3. Low Error Confidence

A major pain point is the **fear of accidental consequences**. The user lacked "psychological safety"—the feeling that they could explore without breaking something. This led to:

- **Feature Avoidance:** Intentionally ignoring advanced tools due to a lack of confidence.
- **Anxiety-Driven Navigation:** Hesitation and worry during the first-time use.

## 4. Subjective Intuitiveness

The contrast between the user and their colleague highlights that **"intuitive" is relative**. Success with the current UI may depend heavily on a user's prior mental models or technical background, meaning the design isn't universally accessible to all segments of the target audience.



Key user insights from the excerpt:

- **Initial usability friction:** The first-time experience created confusion and hesitation. The user felt uncertain about navigation and feared making mistakes.
- **Learnability improves over time:** The product becomes manageable once the user builds familiarity and learns where things are.
- **Feature richness is double-edged:** The user values having many options, but the volume of choices increased cognitive load during onboarding.
- **Confidence gap affects behavior:** Uncertainty did not cause abandonment, but it reduced feature adoption. The user self-limited usage to avoid errors.
- **Perceived intuitiveness varies by user:** Another user found the interface immediately intuitive, suggesting the experience may depend on prior mental models, digital fluency, expectations, or working style.
- **Retention risk is moderate, not critical:** The friction was not severe enough to stop usage entirely, but it may suppress engagement, efficiency, and discovery of advanced functionality.
- **Possible UX implication:** The issue may not be the existence of many features itself, but insufficient guidance, hierarchy, onboarding, affordances, or progressive disclosure for first-time users.

THE THROUGHLINE

The output is  
*plausible.*

That doesn't mean  
*it's true.*



WHAT LLMS ACTUALLY DO

# Approach AI like a *research method.*

WITH

*Curiosity. Skepticism. Methodological awareness.*

SECTION THREE

# 03

---

**The AI-Enabled Workflow**

# The *six stages* of UXR

01



**DISCOVER**

02



**PLAN**

03



**CAPTURE**

04



**ANALYSE**

05



**DECIDE**

06



**ORGANIZE**

# The *six stages* of UXR

01



DISCOVER

02



PLAN

03



CAPTURE

04



ANALYSE

05



DECIDE

06



ORGANIZE



# *Discover:* Make sense of the landscape

## WHAT

Understand the problem space before deciding what to research or test



## HOW

Review existing knowledge, scan the market, look at competitors, audit the current experience, surface early pain-point signals



## WHAT AI CAN DO FOR US

*Synthesize large volumes of information, compare sources, surface patterns, organize assumptions, and identify gaps*

**AI helps us get oriented faster, but Discovery outputs are still signals, not validated truth.**

# Web & market *scanning*

Answer Links Images Share Download Comet 10 sources

What are the current UX patterns used by major European ecommerce retailers to communicate delivery costs, delivery timing, returns, and loyalty benefits during checkout? Cite sources where possible.

Major European ecommerce retailers like Zalando, ASOS, and Otto use transparent, concise UX patterns during checkout to build trust and reduce abandonment, emphasizing upfront pricing and clear summaries per EU guidelines on dark patterns. These patterns prioritize progress indicators, localized costs, and easy access to policy links. [prisync · 1](#)

### Delivery Costs

Costs appear early in the cart or checkout with full breakdowns including taxes, shipping fees, and estimates—no surprises at payment. Retailers display all options (e.g., standard, express) with prices and estimated dates upfront, often using progress bars or summaries like "€5.99, 2-3 days". Free shipping thresholds are highlighted near the buy button. [linkedin · 2](#)

### Delivery Timing

Specific "delivery date" phrasing replaces vague "speed" terms, shown per option with visual timelines or calendars. Milestones like "arrives by May 15" integrate into order summaries, with choices for in-store pickup or express. EU sites localize for regions, avoiding hidden fees. [centra · 2](#)

### Returns

Policies link prominently in footers and checkout summaries as bullet-point overviews (e.g., "Free returns within 30 days") with FAQs and return tools. Free returns are explicitly stated if offered, with clear refund timelines and status trackers post-purchase. Processes simplify to one-click initiation, compliant with EU transparency rules. [baymard · 2](#)

### Loyalty Benefits

Perks auto-apply for logged-in users (e.g., discounts, free shipping) via personalized summaries like "VIP: -10% applied". Progress toward rewards shows in side panels or order totals, tying to wallets or points without forced sign-ups. Guest checkout remains available to avoid friction. [linkedin · 2](#)

10 sources

### Follow-ups

How does Zalando's checkout specifically communicate delivery costs and returns

What UX patterns does ASOS use for loyalty benefits during checkout

Ask a follow-up

+ Search Model ↓ ↑

10 sources

Sources for What are the current UX patterns used by major European ecommerce retailers to communicate delivery costs, delivery timing,...

- [prisync](#)  
**8 Payment & Checkout UX Practices in E-commerce - Prisync**  
E-Commerce payments should provide the easiest usability and experience for your customers. Follow this post to learn how.
- [baymard](#)  
**35 Data-Driven E-Commerce Best Practices - Baymard**  
These specific guidelines are sourced directly from Baymard Institute's 71,000+ hours of large-scale UX research that go beyond UX basics.
- [linkedin](#)  
**The Impact Of Checkout Experience On Brand Loyalty - LinkedIn**  
Explore top LinkedIn retail & merchandising content from experienced professionals.
- [linkedin](#)  
**101 E-Commerce UX Guidelines - LinkedIn**  
Great user experience in e-commerce isn't just about clean aesthetics or clever copy — it's about removing friction, earning trust, and enabling users to accomplish their goals with clarity and confidence. As digital competition intensifies and consumer...
- [uxpatterns](#)  
**Checkout Flow - UX Patterns**  
Learn how to implement checkout flows. Discover best practices for payment forms, order review, and conversion optimization.
- [baymard](#)  
**120 Order Returns' Design Examples - Baymard Institute**  
Our testing reveals a long array of potential pitfalls in the online returns experience, such as initiating the return on the e-commerce site, moving through the return flow, printing the shipping label, returns costs, returns shipping methods, in-store returns, and...
- [centra](#)  
**5 pillars of a luxury checkout experience for fashion brands - Centra**  
Struggling with supply chain issues in fashion wholesale? Discover the 5 biggest challenges fashion brands face and how Centra can simplify operations and protect margins.
- [dbsoftware](#)  
**Ecommerce UX Best Practices: A Practitioner's Guide for 2026**  
Ecommerce UX design best practices from real builds. Navigation, checkout, mobile, and AI personalization are backed by verified data from 450K+ user platforms.
- [instinctools](#)  
**Ecommerce UX Best Practices: User Experience Design Guide**  
Ecommerce UX best practices you need to take heed of in 2026 — based on 10K hours of UX research. No fluff, just actionable insights.



# AI-powered *academic research*

The screenshot shows the Elicit website homepage. At the top, there is a navigation bar with the Elicit logo and links for Solutions, Industries, Resources, Customers, and Pricing. On the right side of the navigation bar are 'Sign in' and 'Sign up' buttons. The main heading is 'AI for Scientific Research', followed by the tagline 'Elicit helps researchers be 10x more evidence-based'. A prominent 'Try now' button with a right-pointing arrow is centered below the tagline. Below this is a search input field containing the text 'What are the cognitive effects of theanine?'. Underneath the search field, there is a note: 'Weak research question. Try these for inspiration:' followed by five buttons: 'Definition', 'Mechanism', 'Dosage', 'Onset', and 'Comparison'. At the bottom of the page, it states 'TRUSTED BY OVER 5 MILLION RESEARCHERS, INCLUDING AT' and lists logos for B BRAUN, NASA, Stanford, Takeda, and Unilever. A link 'READ OUR CUSTOMER STORIES' is located at the very bottom.

The screenshot displays the Consensus AI research interface. On the left is a sidebar with navigation options: 'New Thread', 'Home', 'My Library', and 'History'. Below these are 'Recents' with a 'Default Options User...' entry. The main area features the Consensus logo and the text 'Research starts here'. A search bar is labeled 'Ask the research...' and contains 'Corpus' and 'Deep +' dropdown menus, along with a 'Filter' button. Below the search bar are three tool buttons: 'Get a Yes / No answer', 'Quick TL;DR', and 'Find promising research questions'. At the bottom of the sidebar, there is a 'Tools' section with 'Paper search' and 'Citation Graph' options, followed by 'Demos', 'Contact', and a user profile for 'katerina@workwit...'. A message indicator shows '14 / 15 Pro messages left' and an 'Upgrade' button. The footer of the interface includes the slogan 'The new standard for academic research' and a help icon.



## Research starts here

What empirical evidence shows that default options influence user decision-making, opt-in rates, or product choice in digital or service contexts?



Corpus ▾



Deep +



Filter



What empirical evidence shows that default options influence user decision-making, opt-in rates, or product choice in digital or service contexts?

- empirical studies of default options influencing user decision-making, opt-in rates, or prod... 553.2K
- field experiments and observational evidence on default settings, preselected options, and ... 422.8K
- +1 more

### Default Options and User Decisions in Digital & Service Contexts

Across many online and service settings, preselected options (defaults) systematically shift what people choose, how often they opt in, and which products they select.

#### Overall Strength of Evidence

A large meta-analysis of 58 default studies (n≈73,675) finds a **substantial average default effect** (d≈0.68), translating to about a 27 percentage-point change in choosing the default in binary decisions. Defaults are especially effective in consumer domains and when perceived as endorsement or status quo.

#### Default Effects in Digital & Service Environments

Context / Outcome	Default manipulation	Main effect on behavior	Citations
Online charity platform	Donation and platform "co-donation" amounts defaulted	Strong shift of individual contributions toward default amounts; aggregate charity giving unchanged, platform support higher with higher defaults	2 3
Residential electricity pricing	Opt-in vs opt-out to time-varying tariffs	Huge increase in enrollment under opt-out; "passive" defaulted customers still reduced consumption when on dynamic pricing	4
App onboarding &	Wording that implies	Dramatic increases in	5

How do transparency disclosures and informed consent requirements moderate default option in...TAB

+ Corpus Deep Filter

#### 1 When and why defaults influence decisions: a meta-analysis of default effects

KEY TAKEAWAY · Defaults have a significant influence on decisions, but their effectiveness varies, with consumer domains being more effective and environmental domains less effective, and endorsement or endowment being more effective.

META-ANALYSIS HIGHLY CITED

2019 · 483 citations · J. Jachimowicz et al. PDF

Behavioural Public Policy

#### 2 Defaults and Donations: Evidence from a Field Experiment

KEY TAKEAWAY · Defaults strongly affect individual behavior, but aggregate donation levels remain unaffected, while co-donations increase in the default amount.

HIGHLY CITED

2019 · 87 citations · Steffen Altmann et al. PDF

Review of Economics and Statistics

#### 3 Defaults and Donations: Evidence from a Field Experiment

KEY TAKEAWAY · Defaults on charitable donation websites strongly impact individual behavior, but do not affect aggregate donations, as they influence some people to donate more and others to donate less.

HIGHLY CITED

2019 · 109 citations · Steffen Altmann et al. PDF

SSRN Electronic Journal

#### 4 Default Effects And Follow-On Behaviour: Evidence From An Electricity Pricing Program

KEY TAKEAWAY · Default effects in electricity pricing programs lead to reduced participation and subsequent consumption reductions in passive consumers.

RCT

2021 · 59 citations · M. Fowlie et al. PDF

The Review of Economic Studies

#### 5 Nudging App Adoption: Choice Architecture Facilitates Consumer Uptake of Mobile Apps

KEY TAKEAWAY · Integrating multiple feature decisions into a single choice, changing colors to match habitual patterns, and wording options as if enabling the app is the default

# Existing *knowledge* retrieval

The screenshot displays the NotebookLM interface. The top bar shows the notebook title "The Impact of Olfactory Cues on Consumer Attitudes". On the left, a "Sources" panel lists five research papers, all of which are checked. The main "Chat" window contains a user query: "What are the key themes appearing across these research papers on the sense of smell and consumer behaviour in the retail context?". The AI response is structured as follows:

The research papers identify several key themes regarding the influence of olfactory cues on consumer behavior, focusing on how scents interact with the human mind, other sensory stimuli, and the retail environment to shape attitudes and purchasing decisions.

**The Role of Congruence and Fit**

A central theme across the studies is **congruence**, or the "fit" between a scent and the advertised product or environment. Research indicates that when a scent is highly relevant to the product (e.g., a floral scent for a travel ad featuring wildflowers), it can improve product evaluations and recall <sup>1</sup> -. Conversely, **poorly fitting scents**—such as a pine scent used for a floral-themed advertisement—can significantly lower consumer attitudes toward both the ad and the brand, particularly among individuals highly motivated to process information <sup>5</sup> -. Interestingly, while scent-product congruence is generally beneficial, some research suggests that **incongruence between different sensory cues** (like scent and color) can actually enhance consumer response by creating surprise and increasing attention <sup>9</sup> -.

**Psychological and Biological Mechanisms**

The sources highlight that smell is uniquely powerful because of its **direct link to the limbic system**, the part of the brain responsible for emotional responses and memory <sup>14</sup> -. Key mechanisms through which scent influences consumers include:

- **Mood Shift:** Pleasant scents evoke positive mood states, which then color the consumer's judgment of products or environments <sup>19</sup> -.
- **Hedonic Transfer:** Attitudes may be influenced through a direct "sharing" of the scent's pleasantness with the associated object, even without a noticeable shift in the consumer's mood <sup>22</sup> -.
- **Processing Fluency:** The **complexity of a scent** affects how easily it is processed <sup>28</sup> <sup>29</sup> . **Simple, single-dimension scents** (like lemon) are easier for the brain to process than complex blends, leading to increased cognitive ease and, subsequently, higher actual spending in retail settings <sup>30</sup> -.

**Impact on Consumer Behavior and Sales**

Olfactory cues are shown to have a measurable impact on **approach behavior** and purchasing <sup>1</sup> -.

- **Approach vs. Avoidance:** Pleasant ambient scents encourage consumers to stay longer in a store, visit more frequently, and interact more with the environment <sup>18</sup> -.
- **Retail Sales:** Studies demonstrate that the presence of an appropriate, simple scent can

# The *six stages* of UXR

01



DISCOVER

02



PLAN

03



CAPTURE

04



ANALYSE

05



DECIDE

06



ORGANIZE



# *Plan:* Turn ambiguity into a research plan

## WHAT

Define what we need to learn, from whom, and through which method



## HOW

Set objectives, research questions, hypotheses, participant criteria, method choice, guides, tasks, and screeners



## WHAT AI CAN DO FOR US

*Drafts first versions, creates alternatives, checks for bias, compares methods, and helps stress-test the research plan*

**AI helps structure the plan, but the researcher still owns the method and the evidence standard.**

# Use AI as a *sparring partner*



INDICATIVE PROMPT FOR  
METHOD SELECTION

Given these research objectives,  
compare 3 possible study designs:

1. moderated usability testing
2. unmoderated task-based testing
3. customer interviews

For each, explain what it can answer,  
what it cannot answer, risks, sample  
needs, and when you would choose it.



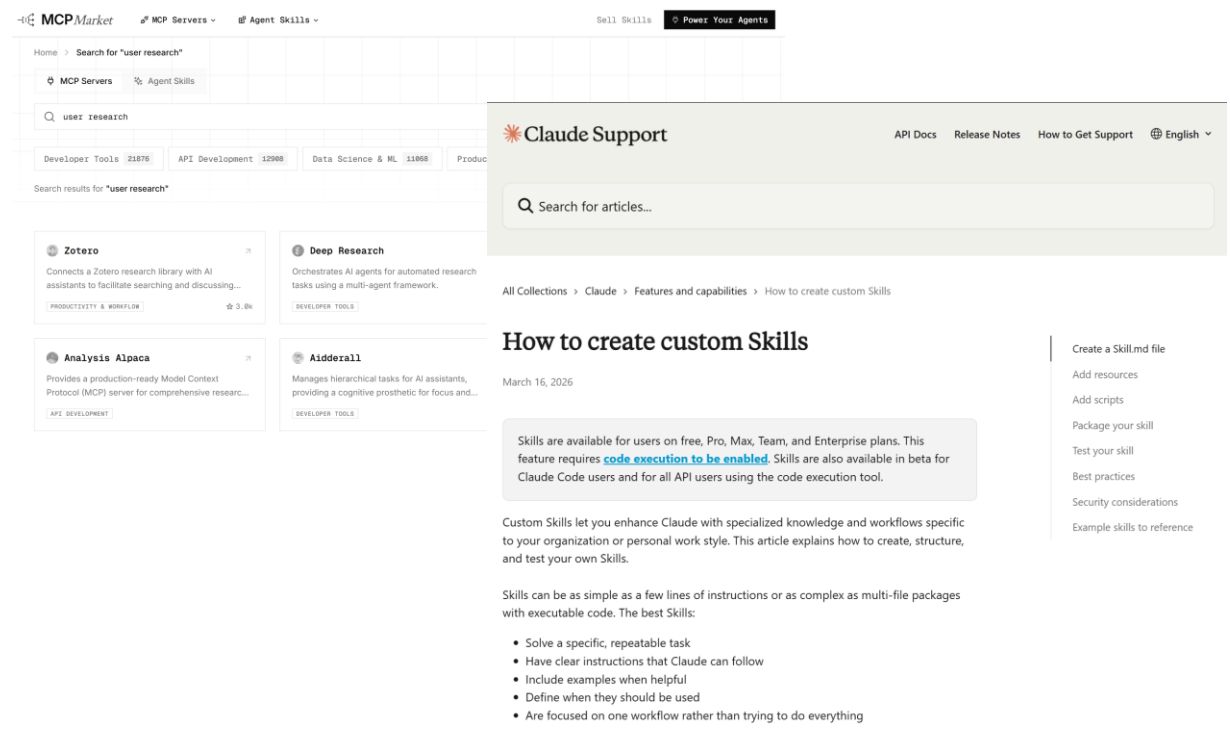
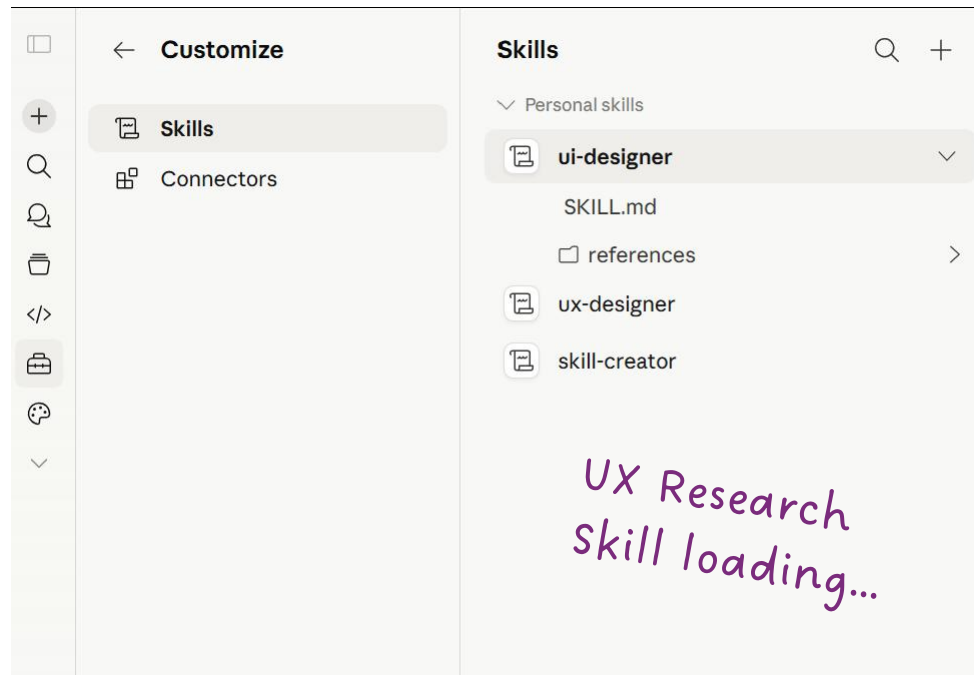
INDICATIVE PROMPT FOR  
USER TESTING GUIDE

Create a moderated usability testing guide for this  
checkout flow. Include:

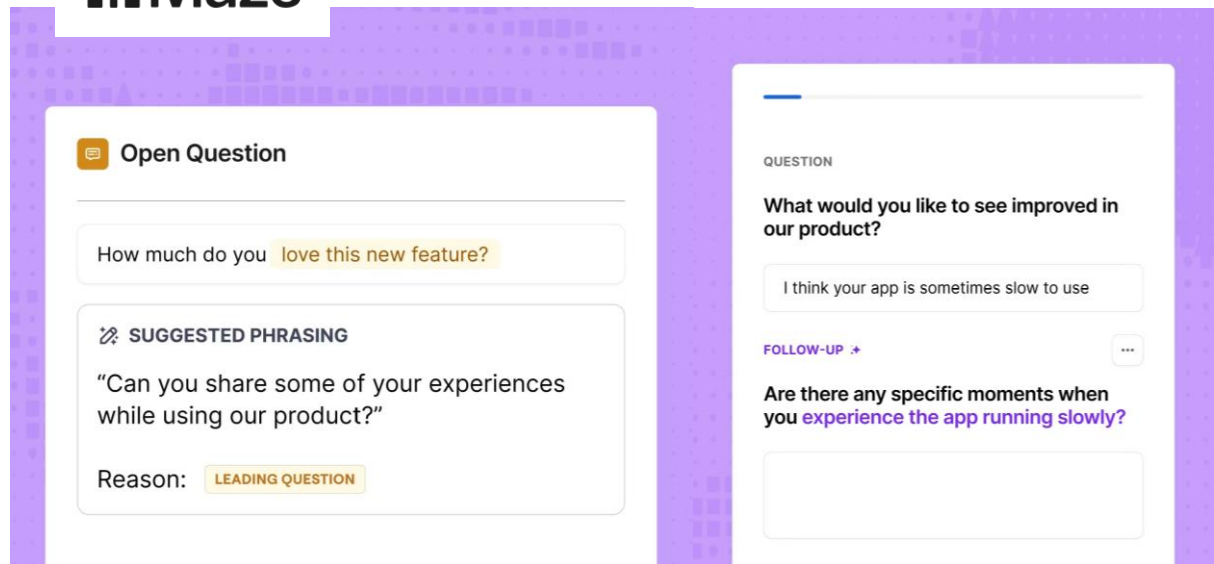
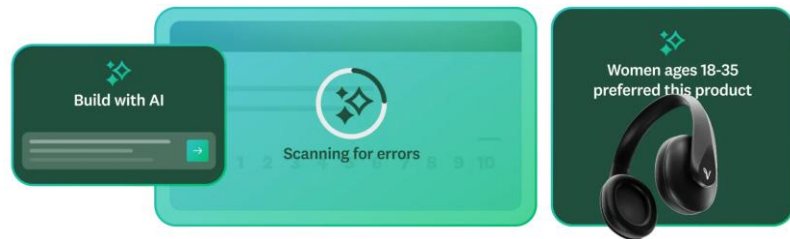
- warm-up questions
- task scenarios
- follow-up questions about delivery cost, delivery timing, payment options, and loyalty benefits
- closing questions

Then review the guide and flag any leading, vague, or  
double-barrelled questions.

# Claude *Skills* are your friend



# Most established tools include *AI-enhanced* features



# Experiment with *AI-native* tools

Listen Solutions Features Customers Resources Careers

Sign in [Demo](#)

## The AI-first research platform

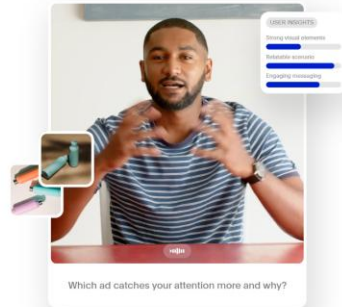
Replace manual research methods with AI-moderated customer interviews.

Announcing our Series B with \$100M raised to date • [Read more](#)

### Understand what your customers want, and why. Fast.

Listen's AI researcher finds your participants, conducts in-depth interviews, and delivers actionable insights in hours, not weeks.

[Book a Demo](#) [Try for Free](#)



### AI-moderated interviews

Probe deeper with an AI interviewer that conducts personalized interviews with each customer, at scale.

### Actionable results, instantly

Generate key takeaways, personas, and reveal top themes. Automatically.



### Recruit participants

Use our pool of millions of participants (B2B and B2C), integrate with your provider, or recruit your own.

### Any language

Translate and transcribe automatically between more than 100 languages.


### Test any stimuli

Videos, images or Figma prototypes - we handle all types of stimuli.

### Video, audio, or text

Engage with customers through video recordings, audio chats, or text.

# Experiment with *AI-native* tools



I want to understand more about the B2B software buying journey, specifically for Training and LMS (Learning Management System) software. How do B2B buyers search for and choose software? Would they trust an AI-powered recommendation agent on a software directory site versus traditional self-service search methods (filters, lists, reviews).

# The *six stages* of UXR

01



DISCOVER

02



PLAN

03



CAPTURE

04



ANALYSE

05



DECIDE

06



ORGANIZE



# *Capture:* Collect raw evidence

## WHAT

Capture real user input, behaviours, and session evidence



## HOW

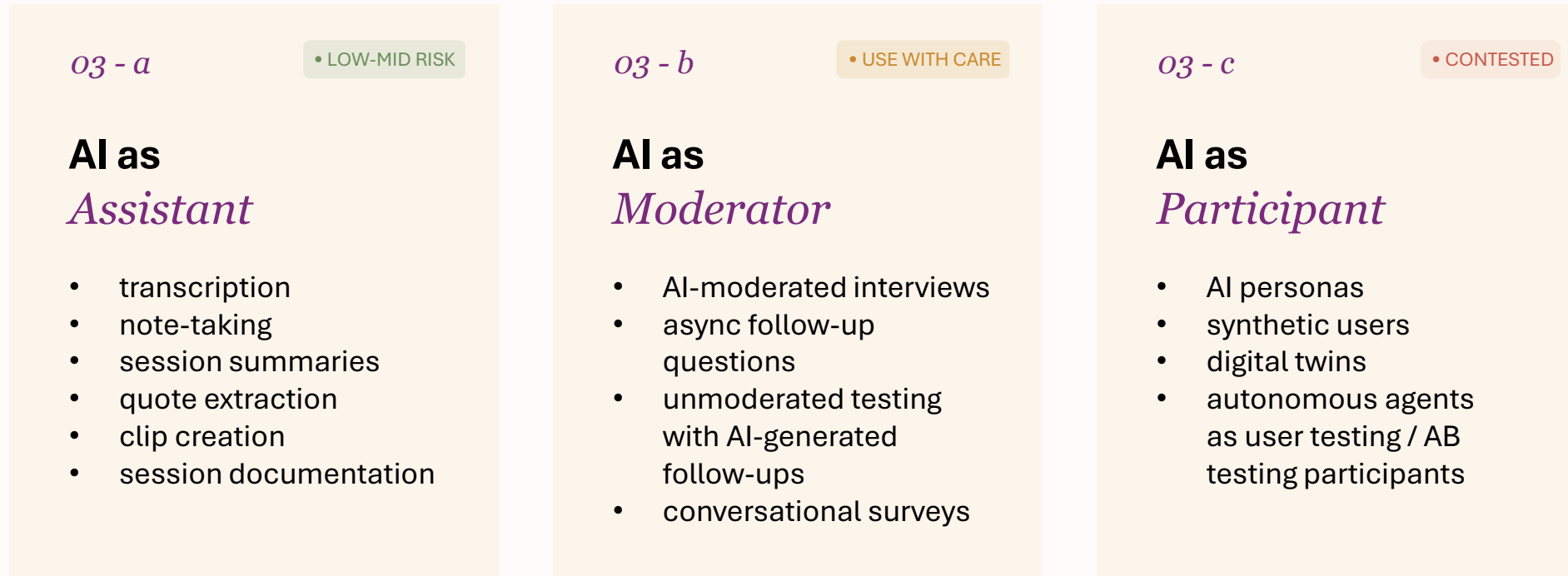
Recruit participants, run sessions, collect responses, record behaviour, transcribe and document what happened



## WHAT AI CAN DO FOR US

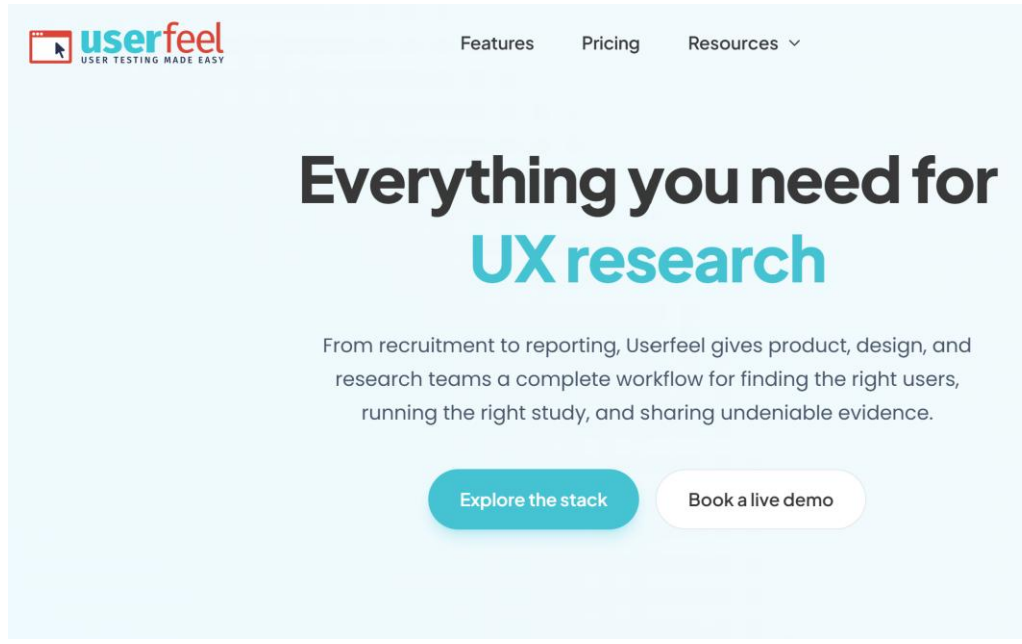
*AI assists, moderates, or simulates — with increasing risk.*

# 3 levels of AI involvement

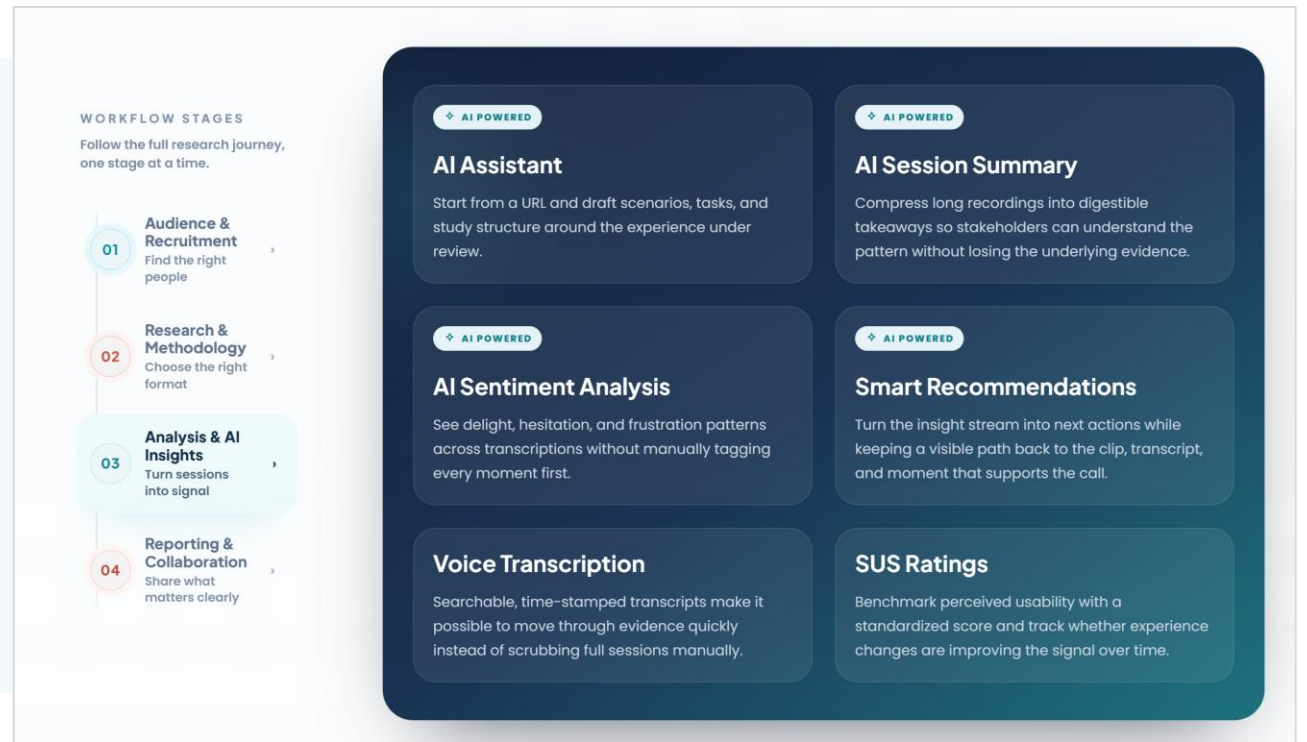


The more AI moves from supporting evidence to replacing its source, the higher the risk.

# Established tools include *AI-powered* features

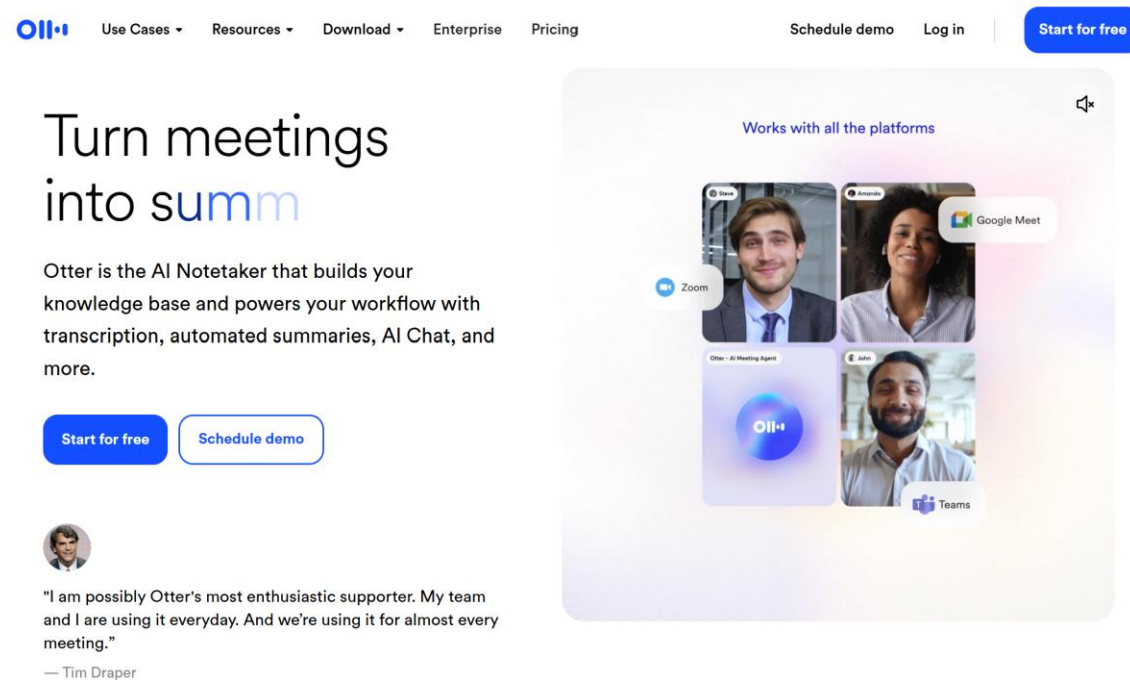


The image shows the top portion of the Userfeel website. At the top left is the Userfeel logo with the tagline "USER TESTING MADE EASY". To the right are navigation links for "Features", "Pricing", and "Resources". The main heading reads "Everything you need for UX research". Below this is a paragraph: "From recruitment to reporting, Userfeel gives product, design, and research teams a complete workflow for finding the right users, running the right study, and sharing undeniable evidence." At the bottom are two buttons: "Explore the stack" and "Book a live demo".

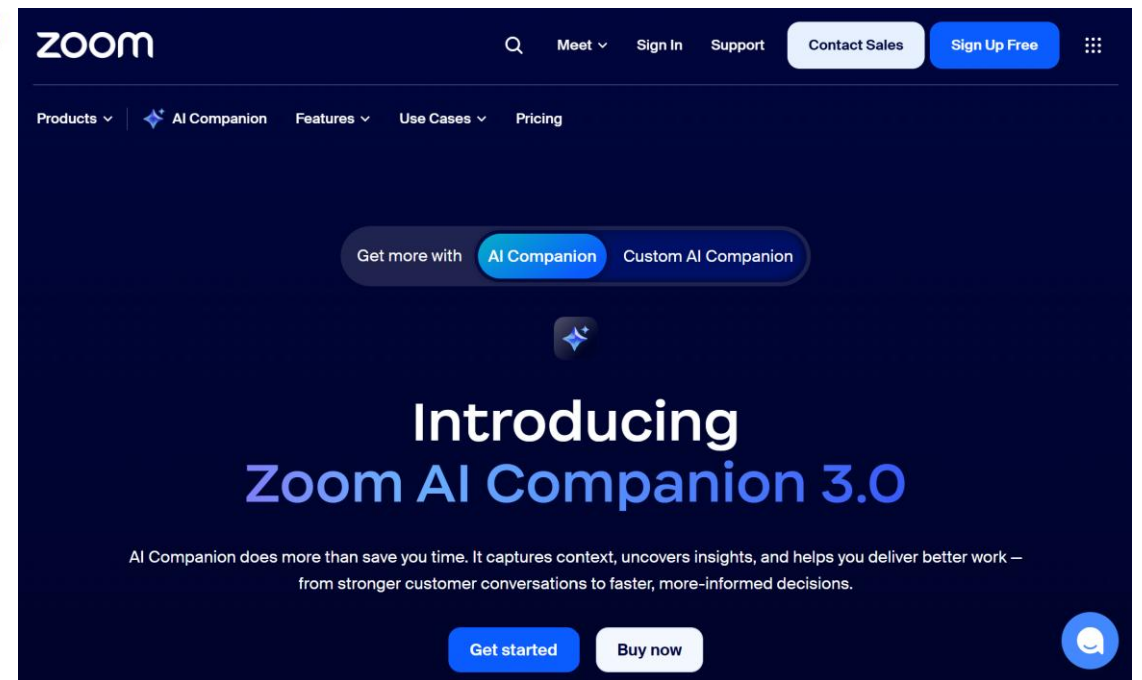


This section details the AI-powered workflow stages and features. On the left, a vertical timeline lists four stages: 01 Audience & Recruitment (Find the right people), 02 Research & Methodology (Choose the right format), 03 Analysis & AI Insights (Turn sessions into signal), and 04 Reporting & Collaboration (Share what matters clearly). The right side features a grid of six AI-powered features, each with a brief description: AI Assistant (Start from a URL and draft scenarios, tasks, and study structure around the experience under review), AI Session Summary (Compress long recordings into digestible takeaways so stakeholders can understand the pattern without losing the underlying evidence), AI Sentiment Analysis (See delight, hesitation, and frustration patterns across transcriptions without manually tagging every moment first), Smart Recommendations (Turn the insight stream into next actions while keeping a visible path back to the clip, transcript, and moment that supports the call), Voice Transcription (Searchable, time-stamped transcripts make it possible to move through evidence quickly instead of scrubbing full sessions manually), and SUS Ratings (Benchmark perceived usability with a standardized score and track whether experience changes are improving the signal over time).

# *AI note-taking*: capturing the session in real time

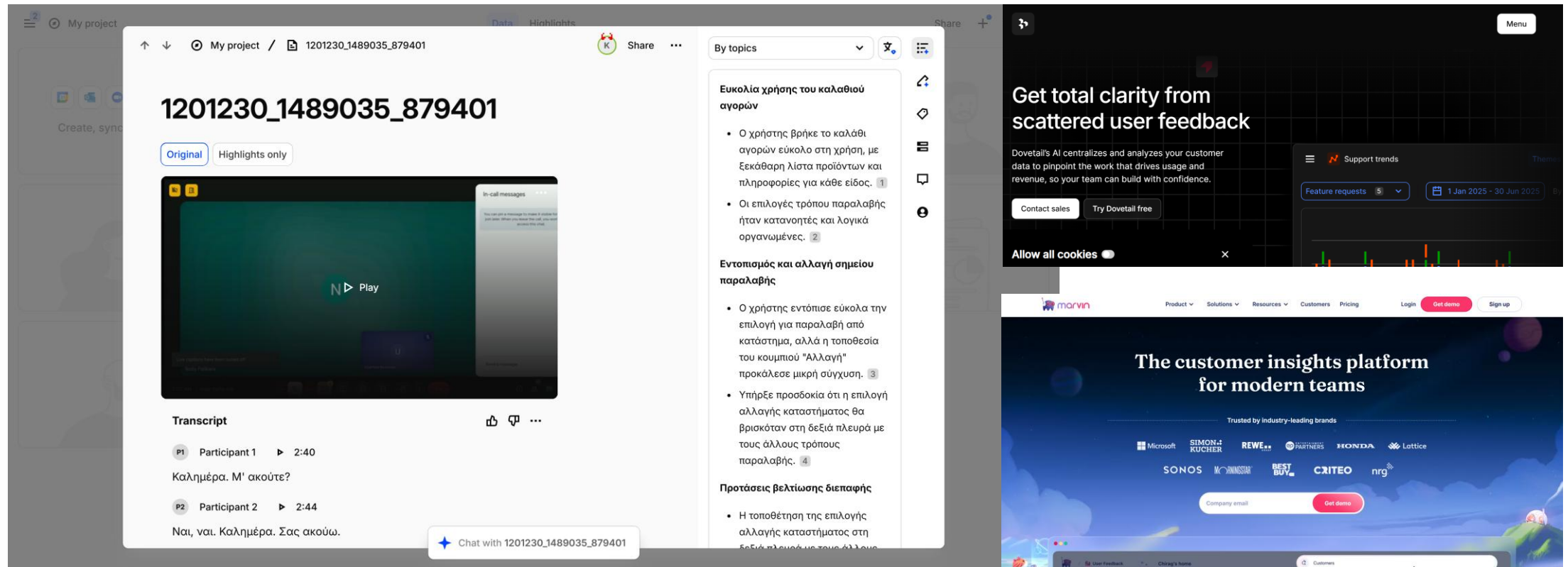


The screenshot shows the Otter.ai website homepage. At the top, there is a navigation bar with links for 'Use Cases', 'Resources', 'Download', 'Enterprise', and 'Pricing'. On the right side of the navigation bar are 'Schedule demo', 'Log in', and a blue 'Start for free' button. The main content area features the headline 'Turn meetings into summ' (partially visible) and a sub-headline 'Otter is the AI Notetaker that builds your knowledge base and powers your workflow with transcription, automated summaries, AI Chat, and more.' Below this text are two buttons: 'Start for free' and 'Schedule demo'. A testimonial section follows, featuring a small profile picture of Tim Draper and a quote: 'I am possibly Otter's most enthusiastic supporter. My team and I are using it everyday. And we're using it for almost every meeting.' The testimonial is attributed to '— Tim Draper'.



The screenshot shows the Zoom AI Companion 3.0 product page. The top navigation bar includes the Zoom logo, a search icon, and links for 'Meet', 'Sign In', 'Support', 'Contact Sales', and 'Sign Up Free'. Below the navigation bar, there are links for 'Products', 'AI Companion', 'Features', 'Use Cases', and 'Pricing'. A prominent call to action reads 'Get more with AI Companion Custom AI Companion'. The main headline is 'Introducing Zoom AI Companion 3.0'. Below the headline, a sub-headline states: 'AI Companion does more than save you time. It captures context, uncovers insights, and helps you deliver better work – from stronger customer conversations to faster, more-informed decisions.' At the bottom of the page, there are two buttons: 'Get started' and 'Buy now', along with a chat icon in the bottom right corner.

# From recordings to *evidence*



# *AI moderation:* Feedback from real participants

Listen Solutions Features Customers Resources Careers

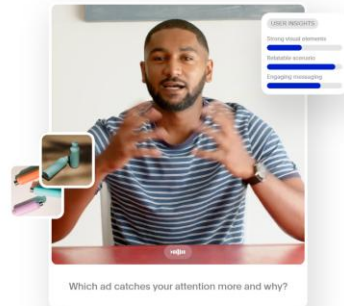
Sign in [Demo](#)

Announcing our Series B with \$100M raised to date • [Read more](#)

## Understand what your customers want, and why. Fast.

Listen's AI researcher finds your participants, conducts in-depth interviews, and delivers actionable insights in hours, not weeks.

[Book a Demo](#) [Try for Free](#)



*Try it out!*



**Personality Test**

Try the Listen AI Interviewer and see how well it understands you.

[Try our AI Personality Test](#)

# *Synthetic users:* When AI becomes the user

The screenshot shows the SyntheticUsers website. The header includes the logo 'SyntheticUsers', navigation links for 'Science', 'Tutorials', 'Pricing', and 'Investors', and buttons for 'Sign in' and 'Book demo'. A dark blue banner below the header says 'Download our latest Guide'. The main content area features a headline 'User research, without headaches.' and a sub-headline 'Synthetic Users lets you predict human behavior before the market does. Think of us like a recruitment agency for research participants, only faster and far more insightful.' There are two 'Book demo' buttons and a 'The science' button. A testimonial section on the right contains two quotes: one from Adam King, a Behavioural Scientist, and another from Johan Van Langendonck, Director of Strategy at Bridgestone Mobility Solutions.

TRUSTED BY TEAMS AT



J.P.Morgan

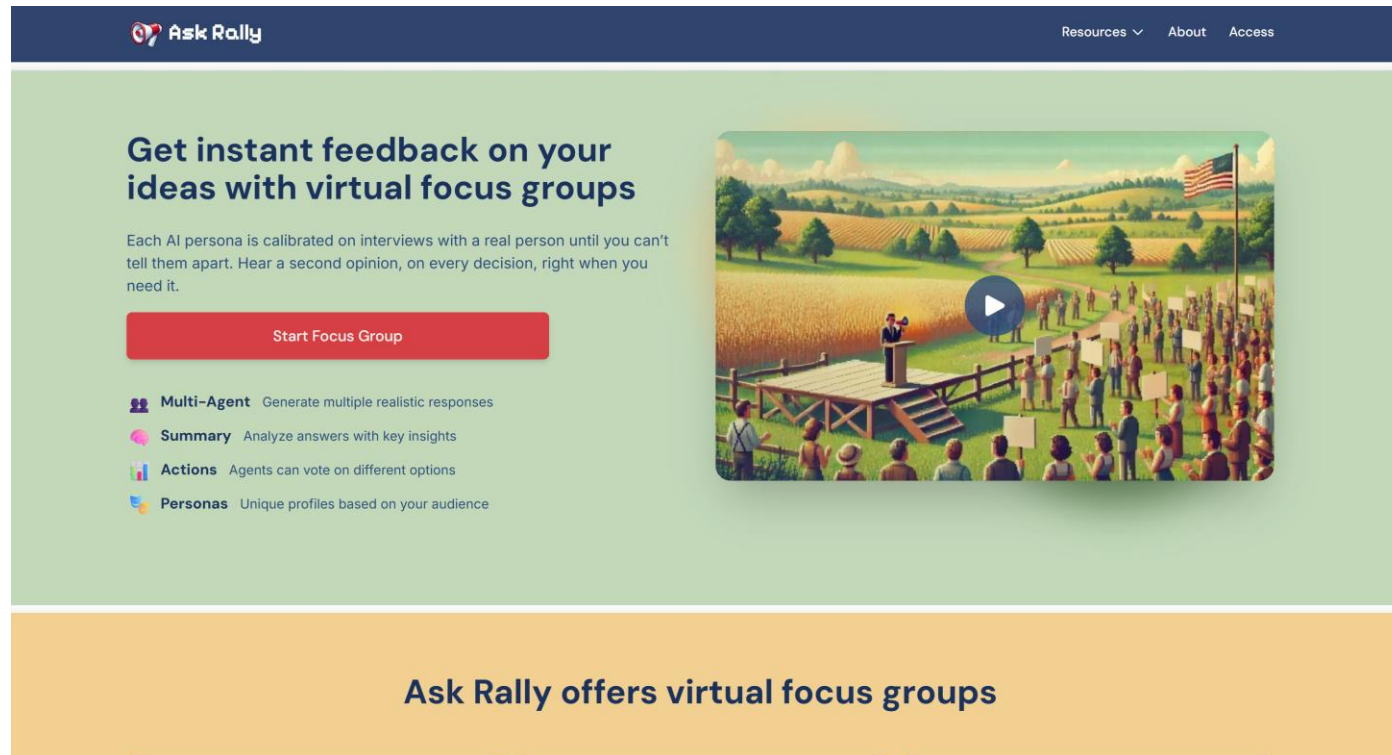
SAMSUNG



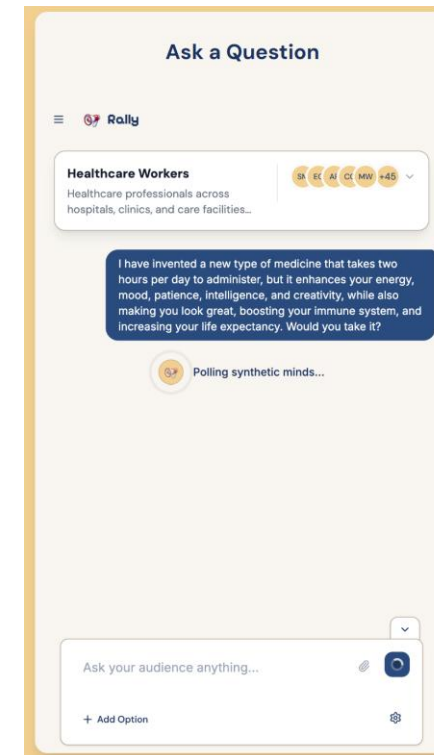
Capge

The screenshot shows the Vurvey Labs website. The header is yellow and contains a link to a TechCrunch article: 'Google names Vurvey Labs Top Start-up to Watch for 2026 (read on TechCrunch)'. Below the header is a navigation bar with 'Vurvey Labs' and a 'Menu' button. The main content area features a large heading 'Simulating the Future' and a sub-headline 'Vurvey Labs' advanced AI Models predict tomorrow's market by simulating human behaviors, not just language.' There are two buttons: 'Start Building' and 'Explore Populations'. At the bottom, there are two small images of people.

# *AI Personas:* When AI becomes the user



The screenshot shows the Ask Rally website landing page. At the top is a dark blue navigation bar with the Ask Rally logo on the left and 'Resources', 'About', and 'Access' on the right. The main content area has a light green background. On the left, there is a heading 'Get instant feedback on your ideas with virtual focus groups' followed by a paragraph: 'Each AI persona is calibrated on interviews with a real person until you can't tell them apart. Hear a second opinion, on every decision, right when you need it.' Below this is a red 'Start Focus Group' button. To the right of the text is a video player showing a man at a podium addressing a crowd in a field. Below the video player are four feature cards: 'Multi-Agent' (Generate multiple realistic responses), 'Summary' (Analyze answers with key insights), 'Actions' (Agents can vote on different options), and 'Personas' (Unique profiles based on your audience). At the bottom of the page is a dark orange banner with the text 'Ask Rally offers virtual focus groups'.



The screenshot shows the 'Ask a Question' interface in the Ask Rally application. At the top, it says 'Ask a Question'. Below that is a navigation bar with the Ask Rally logo. The main content area is white. It features a section for 'Healthcare Workers' with a dropdown menu showing 'SA', 'EC', 'AI', 'CC', 'MW', and '+45'. Below this is a text box containing a question: 'I have invented a new type of medicine that takes two hours per day to administer, but it enhances your energy, mood, patience, intelligence, and creativity, while also making you look great, boosting your immune system, and increasing your life expectancy. Would you take it?'. Below the text box is a poll button labeled 'Polling synthetic minds...'. At the bottom, there is a text input field with the placeholder 'Ask your audience anything...' and a microphone icon. Below the input field is a '+ Add Option' button and a settings icon.

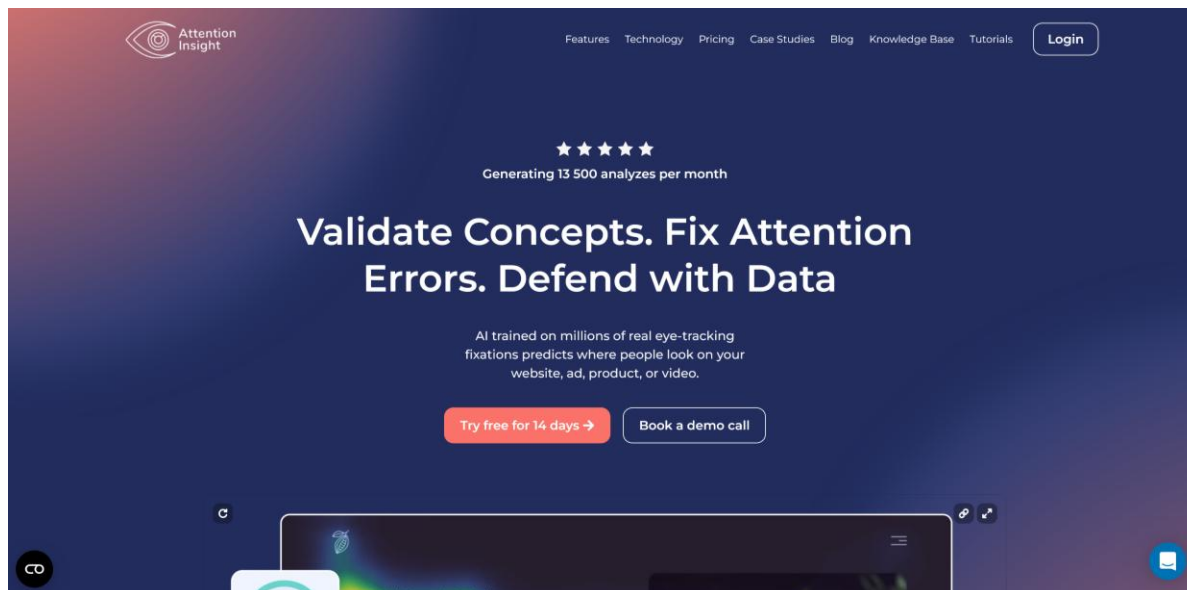
# *Synthetic users:* Simulation is not real evidence



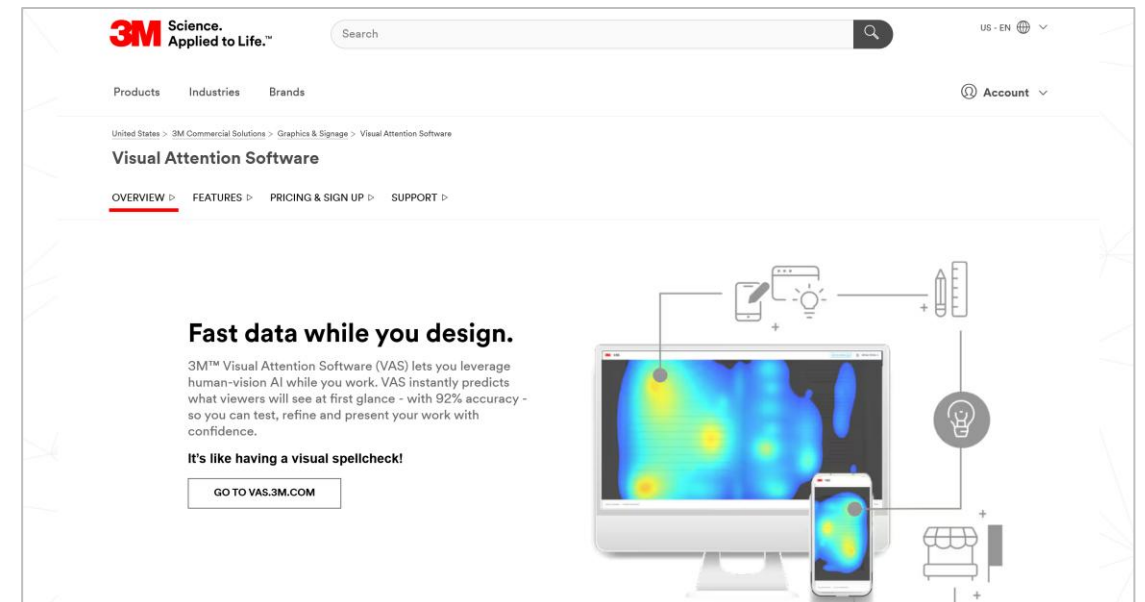
The screenshot shows the top of a Research Square preprint page. At the top left is the Research Square logo and a search bar. Below the logo is the text 'Systematic Review'. The main title is 'Synthetic Participants Generated by Large Language Models: A Systematic Literature Review'. Below the title are the authors 'Eduard Kuric, Peter Demcak, Matus Krajcovic'. A red warning box states 'This is a preprint; it has not been peer reviewed by a journal.' Below this is the DOI link 'https://doi.org/10.21203/rs.3.rs-9057643/v1' and the license 'This work is licensed under a CC BY 4.0 License'. The 'Abstract' section is visible, starting with 'In recent years, the prospect of Large Language Models (LLMs) for simulating participants within various research and data collection methods has been interrogated extensively. Its proponents cite aspirational promises, including high flexibility, adaptability, better representation and reduced research costs, all by leveraging the encoded wisdom of the internet crowd. Empirical studies paint a more nuanced but fragmented picture, with mixed results, heterogeneous methods and a saturation of different perspectives. In this systematic literature review, we delineate a clear and comprehensive conceptual understanding of LLM-generated participants and their comparative relationship to human samples. We synthesize the findings from 182 studies, obtained through a hybrid database and reference search, followed by a rigorous quality curation. Grounded in generalizable indicators, we present a standardized categorization of four fundamental issues that impact synthetic participants across diverse types of simulations – cognitive misalignments, distortions, misleading believability, and overfitting/contamination. Despite the survey revealing integrations of different LLMs, prompt engineering techniques, and participant or environment modeling methods, the fidelity improvements they demonstrated remain modest. At their most representative, LLMs may stochastically parrot data they were pre-trained on or fine-tuned with. To set appropriate expectations, explain their

- Synthetic participants are **not equivalent to human** participants.
- Their believability can be **misleading**.
- They may **echo** patterns from training data rather than reveal real user insight.
- Best framed as **heuristic tools**, not research evidence.
- Useful for **hypotheses and preparation**; risky for validation or decision-making.

# *Predictive research:* Quick input, not validation



The screenshot shows the landing page for Attention Insight. At the top left is the logo, and to the right are navigation links: Features, Technology, Pricing, Case Studies, Blog, Knowledge Base, Tutorials, and a Login button. The main content area has a dark blue background with white text. It features a five-star rating and the text "Generating 13 500 analyzes per month". The primary headline reads "Validate Concepts. Fix Attention Errors. Defend with Data". Below this, it states "AI trained on millions of real eye-tracking fixations predicts where people look on your website, ad, product, or video." At the bottom of the main section are two buttons: "Try free for 14 days" and "Book a demo call". A partial view of a heatmap visualization is visible at the bottom of the page.



The screenshot shows the product page for 3M Visual Attention Software. The top navigation includes the 3M logo, the tagline "Science. Applied to Life.", a search bar, and a language selector for "US - EN". Below the navigation are links for "Products", "Industries", and "Brands", along with an "Account" link. The breadcrumb trail reads "United States > 3M Commercial Solutions > Graphics & Signage > Visual Attention Software". The product title "Visual Attention Software" is followed by a menu with "OVERVIEW", "FEATURES", "PRICING & SIGN UP", and "SUPPORT". The main content area features the headline "Fast data while you design." and a paragraph explaining that 3M Visual Attention Software (VAS) uses human-vision AI to predict viewer behavior with 92% accuracy. A call to action button says "GO TO VAS.3M.COM". To the right is a diagram showing a computer monitor displaying a heatmap, a smartphone with a heatmap, and various icons representing design and analysis tools.

# Agentic users: Still emerging



Computer Science > Computation and Language

[Submitted on 13 Apr 2025 (v1), last revised 19 Sep 2025 (this version, v3)]

## UXAgent: A System for Simulating Usability Testing of Web Design with LLM Agents

Yuxuan Lu, Bingsheng Yao, Hansu Gu, Jing Huang, Jessie Wang, Yang Li, Jiri Gesi, Qi He, Toby Jia-Jun Li, Dakuo Wang

Usability testing is a fundamental research method that user experience (UX) researchers use to evaluate and iterate their new designs. But what about evaluating study design itself? Recent advances in Large Language Model-simulated Agent (LLM Agent) research inspired us to design UXAgent to support UX researchers design before they conduct the real human-subject study. Our system features a Persona Generator module, an LLM Agent module, and a Universal Browser Co generate thousands of simulated users and to interactively test the target website. The system also provides a Result Viewer Interface so that the UX researchers generated qualitative (e.g., agents' post-study surveys) and quantitative data (e.g., agents' interaction logs), or even interview agents directly. Through a heuristic participants praised the innovation of our system but also expressed concerns about the future of LLM Agent usage in UX studies.

Subjects: **Computation and Language (cs.CL)**; Human-Computer Interaction (cs.HC)

Cite as: [arXiv:2504.09407](https://arxiv.org/abs/2504.09407) [cs.CL]  
(or [arXiv:2504.09407v3](https://arxiv.org/abs/2504.09407v3) [cs.CL] for this version)  
<https://doi.org/10.48550/arXiv.2504.09407>

### Submission history

From: Yuxuan Lu [[view email](#)]  
[v1] Sun, 13 Apr 2025 02:34:22 UTC (5,562 KB)  
[v2] Mon, 21 Apr 2025 05:22:55 UTC (6,247 KB)  
[v3] Fri, 19 Sep 2025 17:52:50 UTC (6,004 KB)



Computer Science > Human-Computer Interaction

[Submitted on 13 Apr 2025 (v1), last revised 19 Sep 2025 (this version, v3)]

## AgentA/B: Automated and Scalable Web A/BTesting with Interactive LLM Agents

Dakuo Wang, Ting-Yao Hsu, Yuxuan Lu, Hansu Gu, Limeng Cui, Yaochen Xie, William Headean, Bingsheng Yao, Akash Veeragouni, Ji

A/B testing experiment is a widely adopted method for evaluating UI/UX design decisions in modern web applications. Yet, traditional A/B testing remains co scale and live traffic of human participants, and the long time of waiting for the testing result. Through formative interviews with six experienced industry prac in current A/B testing workflows. In response, we present AgentA/B, a novel system that leverages Large Language Model-based autonomous agents (LLM interaction behaviors with real webpages. AgentA/B enables scalable deployment of LLM agents with diverse personas, each capable of navigating the dyna multi-step interactions like search, clicking, filtering, and purchasing. In a demonstrative controlled experiment, we employ AgentA/B to simulate a between-s this [http URL](#), and compare agent behaviors with real human shopping behaviors at a scale. Our findings suggest AgentA/B can emulate human-like behavi

Subjects: **Human-Computer Interaction (cs.HC)**; Computation and Language (cs.CL)

Cite as: [arXiv:2504.09723](https://arxiv.org/abs/2504.09723) [cs.HC]  
(or [arXiv:2504.09723v3](https://arxiv.org/abs/2504.09723v3) [cs.HC] for this version)  
<https://doi.org/10.48550/arXiv.2504.09723>

### Submission history

From: Yuxuan Lu [[view email](#)]  
[v1] Sun, 13 Apr 2025 21:10:56 UTC (2,379 KB)  
[v2] Mon, 21 Apr 2025 23:57:49 UTC (2,379 KB)  
[v3] Fri, 19 Sep 2025 17:56:58 UTC (2,645 KB)

# The *six stages* of UXR

01



DISCOVER

02



PLAN

03



CAPTURE

04



ANALYSE

05



DECIDE

06



ORGANIZE



# *Analyze:* Turn raw evidence into meaning

## WHAT

Understand patterns, differences, and what they mean for the product or service



## HOW

Review transcripts, tag observations, cluster themes, compare participants, identify outliers, validate findings against evidence



## WHAT AI CAN DO FOR US

*Summarizes data, creates first-pass tags, clusters observations, surfaces recurring themes, and helps explore possible interpretations*

**AI helps find patterns faster, but it cannot decide what matters.**

# AI analysis: Finding patterns in research data

CoLoop

Customers Use Cases Blog Documentation

Log in

Book a demo

## Turn raw qualitative data into strategic insights faster

CoLoop is your AI analysis tool for qualitative research projects. We provide an end-to-end flow for recording, synthesising and analysing research material into high impact insights.

Book a demo

DoReveal

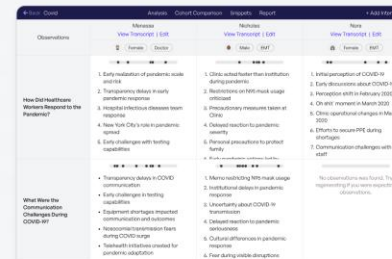
Home Data Privacy Pricing Prompt Library Register/Login

Book a Demo

## Three Ways DoReveal Enables Analysis

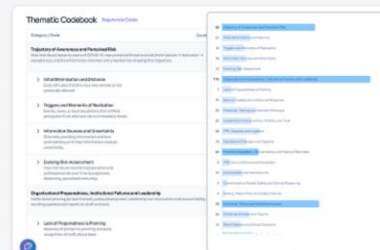
DoReveal is built by researchers, for researchers. We understand the researcher's mindset and workflows. Here are three ways we help you dive deep into your data.

Try it free →



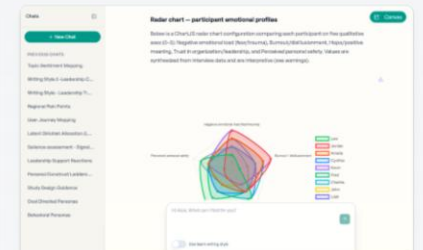
### Structured grids

Build analysis grids that keep evidence, themes, and decisions organized across interviews and cohorts.



### Thematic analysis

Surface patterns, contradictions, and nuance with grounded thematic synthesis that stays close to participant language.



### Agentic AI chat

Interrogate your data with an AI partner that can compare cohorts, trace evidence, and draft insights on demand.

# AI analysis: Use responsibly

The screenshot shows a web application interface for analyzing user testing notes. The main document, titled "User Testing Notes.docx", contains the following text in Greek:

**LP**  
Τα διαβάζει όλα, αλλά δεν γίνεται ξεκάθαρο  
Ποιά η διαφορά από το δάνειο; Θα ήθελε διαφορές π.χ. σε ένα πίνακα (έτσι θα γινόταν ξεκάθαρο ότι είναι πιστωτικό όριο αλλά και ότι μπορείς να το πάρεις ανεξάρτητα από την αγορά του προϊόντος)  
Διαβάζει «το διαθέσιμο όριο ανανεώνεται» αλλά δεν το καταλαβαίνει  
«εύελικτες επιλογές δόσεων» νόμιζε ότι αφορά την ΙΔΙΑ αγορά  
3 boxes δόσεων – δεν καταλάβαινε τις διαφορές, πρέπει να φαίνεται ότι σχετίζεται με τον αριθμό δόσεων  
Calculator – την ενδιέφερε

**Checkout**  
Θα ήθελε να βλέπει και να επιλέγονται οι περισσότερες δόσεις που είναι άτοκες;  
Επιτόκιο – 0% το είδε, αλλά θα μπορούσε να είναι πιο έντονο  
Συνολική τιμή – δεν ήταν ξεκάθαρη  
Έξοδα φακέλου – την ενδιέφεραν αλλά δεν τα βρήκε

**Modal**  
Το περίμενε, δεν την ξένισε  
Δεν έδωσε σημασία στα βήματα

**Τράπεζα**  
Ποια δικαιολογητικά;; ήθελα να ξέρει νωρίτερα  
Είδε ότι μπορεί να πάρει μεγαλύτερο ποσό και της άρεσε, αλλά δεν κατάλαβε ότι μπορεί να πάρει μεγαλύτερο επί τόπου  
OK τα 2 δικαιολογητικά που ζητάει  
Ευαίσθητοι συμβαλλόμενοι – δεν καταλαβαίνει τι είναι  
Που είναι οι δόσεις;;; Που θα εμφανιστούν; Θα ήθελα να έχω ένα IBAN  
Λογαριασμός σύνδεσης – που εμφανίζεται;  
METABΑΣΗ ΣΕ ΑΓΟΡΑ νόμιζε ότι θα γυρίσει στο [redacted] θέλει αλλαγή λεκτικού

The right sidebar shows a list of notes with a "Select 20 notes" button at the top. The notes are categorized by "Content" (Landing Page) and "Usability" (Checkout). A modal window is open over the "Checkout" note, showing the text "Είδε το επιτόκιο - αλλά θα μπορούσε να φαίνεται περισσότερο" and buttons for "Save changes" and "Cancel".



THE AI-ENABLED WORKFLOW

# The *six stages* of UXR

01



DISCOVER

02



PLAN

03



CAPTURE

04



ANALYSE

05



DECIDE

06



ORGANIZE

# The *six stages* of UXR

01



**DISCOVER**

02



**PLAN**

03



**CAPTURE**

04



**ANALYSE**

05



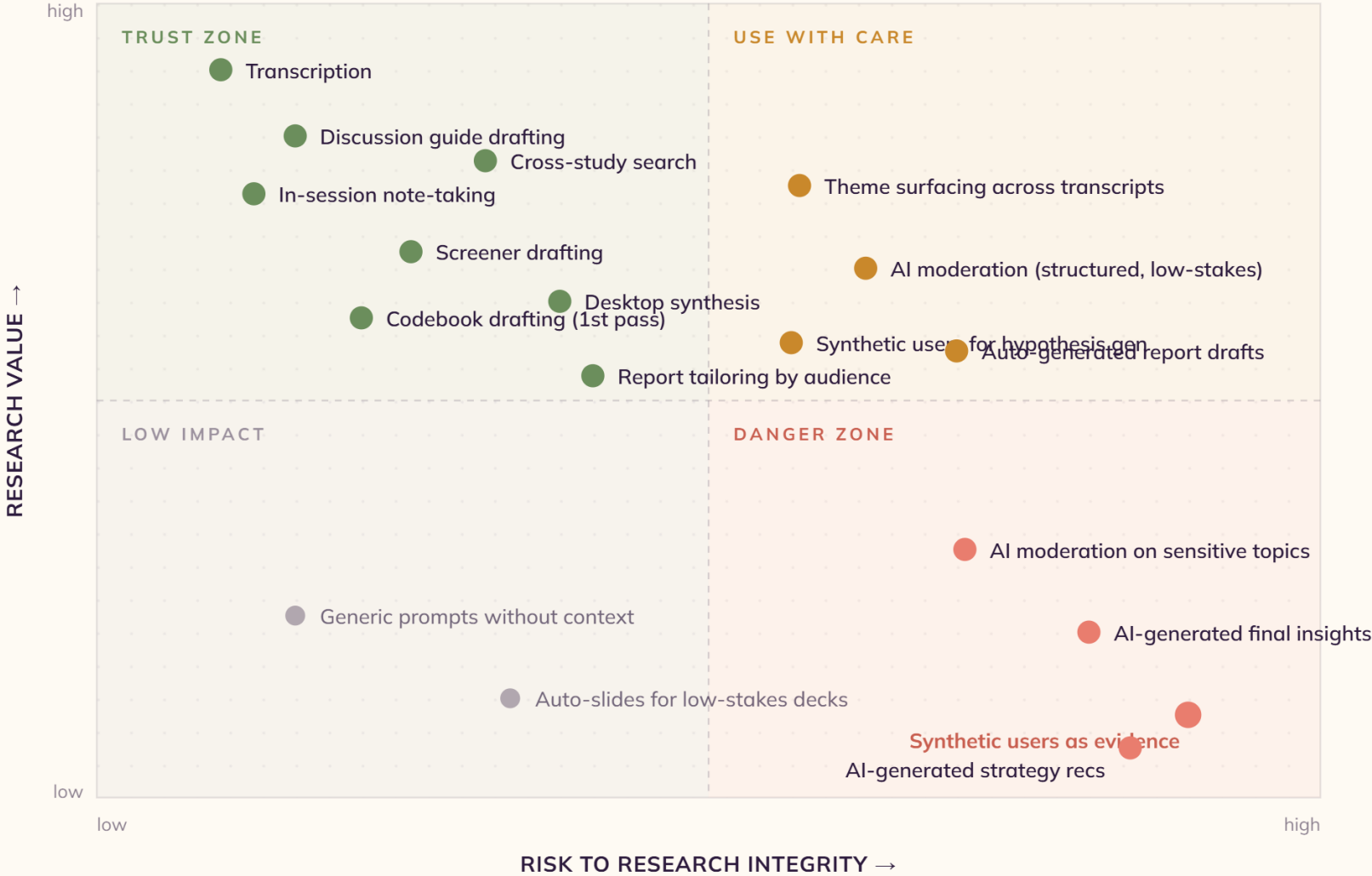
**DECIDE**

06



**ORGANIZE**

# Research value *vs* Risk to integrity



# If in doubt, stick to *safe defaults*

## 01 Discover

Use AI *to map the space*, then verify sources manually.

## 02 Plan

Use AI *to draft and critique*, not to choose the method.


## 03 Capture

Use AI *to record and structure sessions*, not to replace participants or moderators on sensitive work.

## 04 Analyze

Use AI *to find patterns*, then validate against raw data.

# AI x *UXR* Framework

 K Consulting · UX/CX Research REFERENCE FRAMEWORK · 2026

---

## AI x UX Research

*What helps, what hurts & what still needs you.*

**CORE IDEA**

Across the UX research lifecycle, AI is strongest when it helps you *process, structure, compare, retrieve, and draft*. It is weakest when it is asked to replace human judgment, participant reality, or strategic accountability.

<https://claude.ai/public/artifacts/335021c0-33a1-4e9f-a5aa-053e1ee8f4a3>



SECTION FOUR

# 04

---

**From notes to insights: using AI responsibly**

# When analyzing, *know the risks*

## 01 Hallucinated or misattributed evidence

Quotes, speakers, or observations may be invented, distorted, or assigned to the wrong participant.

## 02 Theme frequency over significance

AI surfaces what appears often, but rare moments can be strategically or experientially critical.

## 03 Flattening of contradiction

Differences between users are smoothed into one coherent-sounding theme.

## 04 Confirmation of the research framing

AI reflects the structure of your guide, hypotheses, or prompt back to you.

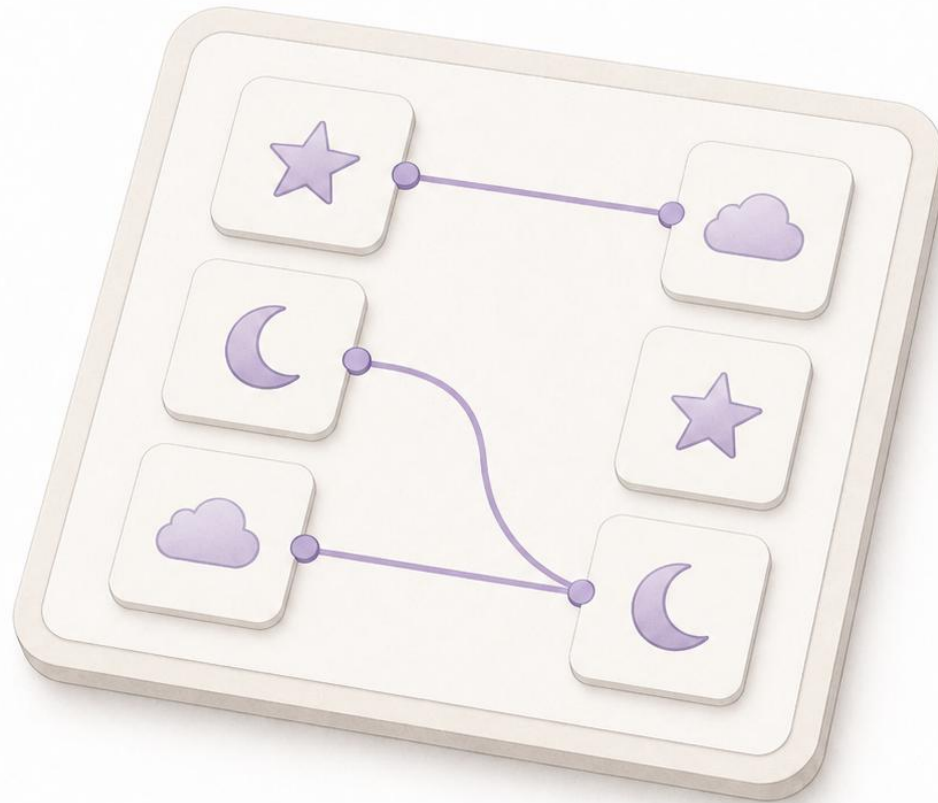
## 05 Context collapse across sessions

Participant context, session conditions, segment differences, or provenance get lost across transcripts.

## 06 Plausible structure masking shallow analysis

The output looks like a research readout, but may not contain real interpretation or evidence discipline.

# Let's match the risks to the *guardrails*



<https://claude.ai/public/artifacts/3a7e6f64-32bc-46e4-87c1-a57d73769ebf>



## RISKS

01

**Hallucinated or misattributed evidence**

02

**Theme frequency over significance**

03

**Flattening of contradiction**

04

**Confirmation of the research framing**

05

**Context collapse across sessions**

06

**Plausible structure vs shallow analysis**

## GUARDRAILS

A

**Structured extraction prompts**

Ask the AI to retrieve specific instances from the source material, including direct quote, speaker, timestamp, session, and context.

*Find every instance where a participant expressed frustration. Quote directly and include speaker, timestamp, and source session.*

B

**Adversarial prompting**

Ask the AI to actively search for evidence that contradicts, complicates, or weakens the synthesis it has produced.

*What evidence in these transcripts contradicts the themes you identified? What does not fit?*

C

**Separate retrieval from interpretation**

Use AI for lower-risk tasks such as locating, sorting, and organizing evidence, while keeping analytical judgment and insight generation human-led.

D

**Verification sampling**

Treat AI synthesis as a first-pass hypothesis and manually verify selected claims, quotes, and themes against the original transcript or recording before reporting.

E

**Session-level analysis before cross-session synthesis**

Analyze each participant or session separately first, preserving speaker, segment, task, context, and provenance before synthesizing across the dataset.

F

**Explicit uncertainty prompting**

Ask the AI to state where the evidence is strong, where it is thin, and what would make each finding less reliable.

*How confident are you in each finding, and what would make you less confident?*

SECTION FIVE

# 05

---

**Cognitive offloading & the future of expertise**

# Cognitive *offloading*



How do we keep **learning**, building **intuition**, tacit **knowledge** and **expertise**?

## AGENDA

# What we've *covered*.

- 01** Intro
- 02** What LLMs actually do
- 03** The AI-enabled UXR workflow
- 04** From notes to insights: Using AI responsibly
- 05** Cognitive offloading & the future of expertise



K CONSULTING

# Thank you.

*Questions, pushback, and stories welcome!*

---

**Katerina Karagianni**

UX & CX Consultant & Researcher



<https://www.linkedin.com/in/katerinakaragianni/>