User Experience Research Portfolio

Maria Perez

About Myself

Longitudinal analysis & Psychometrics
to measure intangible assets

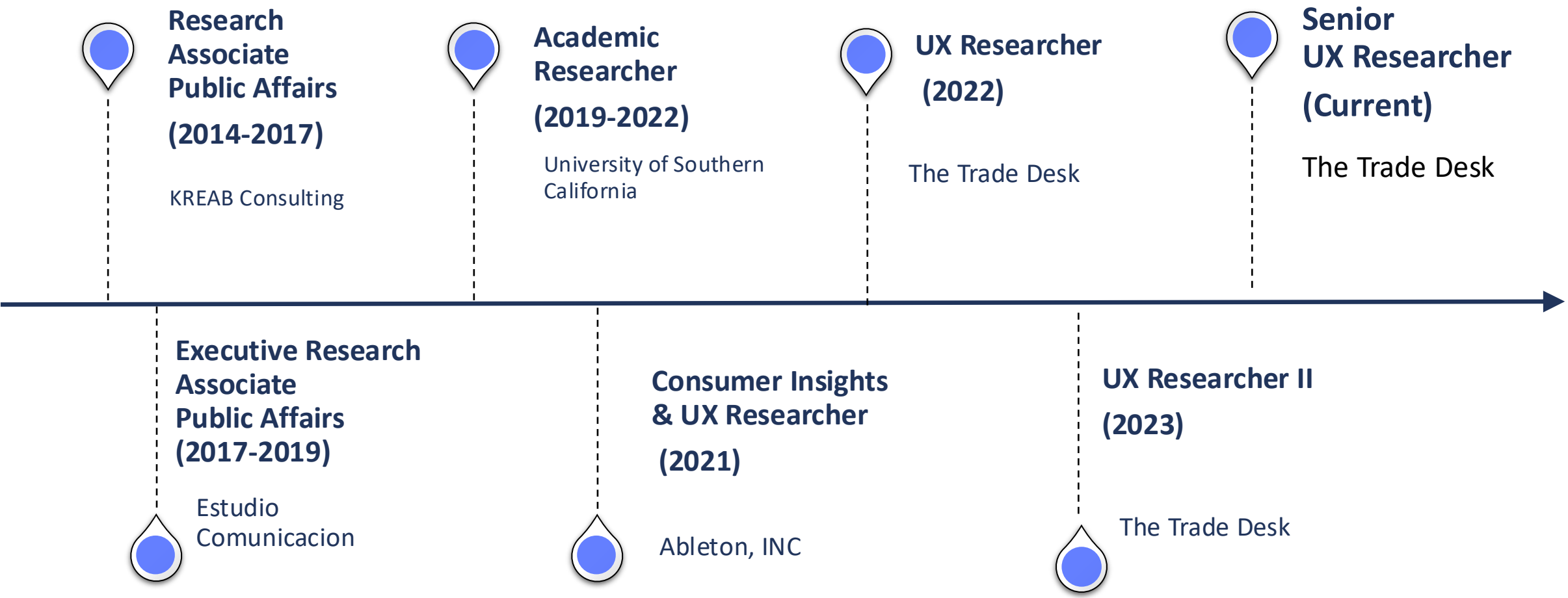
Ph.D. Political Science (USC)

M.S. Applied Econometrics (USC)

Ableton

The Trade Desk

Research Experience



Research Skills

Surveys & Measurement

- Survey design, analysis, and visualization
- Conjoint analysis (choice-based, adaptive)
- Questionnaire design & scale development
- Sampling strategies & weighting
- Response bias detection and mitigation

Experimental & Behavioral Data

•A/B testing and experimentation frameworks
•Hypothesis testing and statistical inference
•Funnel, cohort, and retention analysis
•Behavioral analytics interpretation (usage data)
•KPI and metric definition (e.g., NPS, CSAT, SUS)

Statistical Methods

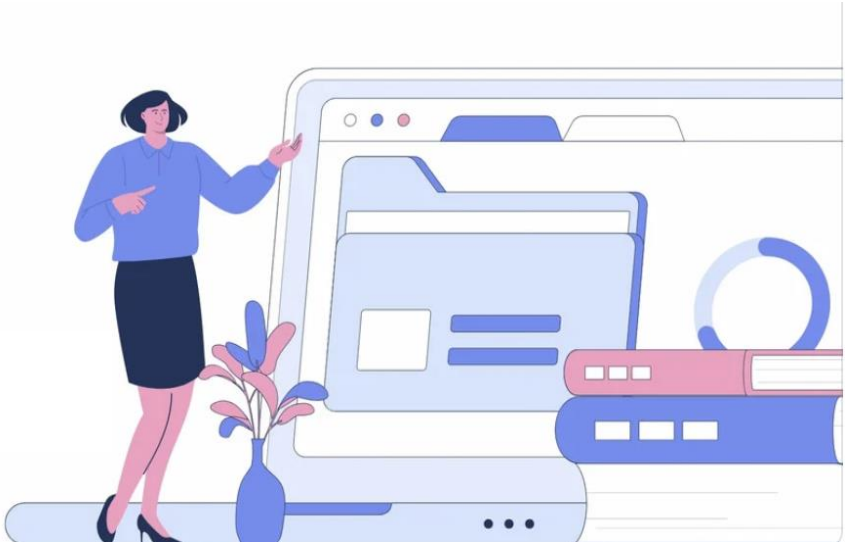
•Descriptive and inferential statistics
•Regression analysis (linear, logistic)
•Factor analysis & dimensionality reduction
•Segmentation & clustering
•Power analysis and sample size estimation

Interviews

•Structured and semi-structured interviews
•In-depth user interviews (remote & in-person)
•Stakeholder interviews (internal & external)
•Contextual inquiry

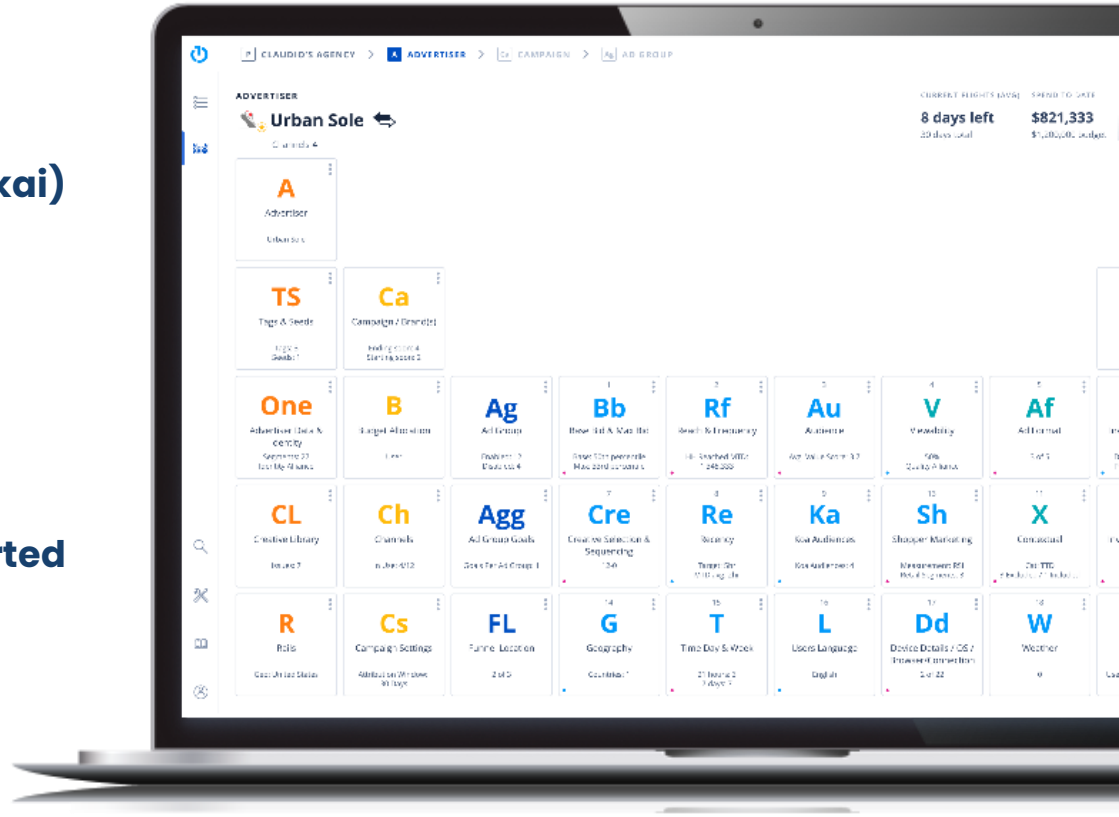
Observational & Generative Methods

•Usability testing (moderated & unmoderated)
•Diary studies
•Field studies & ethnographic research
•Task analysis and workflow mapping
•Journey mapping & service blueprints
•Persona and archetype development
•Jobs-to-Be-Done (JTBD) analysis



theTradeDesk®

1. **Demand Side Platform**
2. **Manage digital advertising campaigns across Connected TV (CTV), video, display, and audio**
3. **Partnered with Netflix and other major publishers**
4. **Moving Away from Legacy Platform (Solimar) to new (Kokai)**
5. **Consumer Products: SSO OpenPass**
6. **Agentic AI for Supported Workflows**



Audience Targeting is Not a Data Problem:

It's a Workflow Problem

The Problem

Advertisers today face:

- Thousands of available audience segments
- Limited visibility into quality or freshness
- No clear link between audience choices and performance

Limited insights into how and why are segments selected to create audiences,

The Goal

We set out to understand:

- How advertisers build audiences
- What drives segment selection
- How often audiences are refreshed
- Why automation adoption remains low
- How audience quality impacts performance

What prevents advertisers from using better data?

Workflow Behavior Insights

Backend event-level log data

- Audience Builder search logs
- Segment selection events
- Targeting data composition logs
- Audience update activity
- Adgroup-level performance reports

Behavioral Metrics Constructed

- Search Depth
- Selection Latency
- Segment Reuse Rate
- Audience Refresh Frequency
- **Data Enrichment**

Analytical Methos

- Timestamp-based clustering of incremental search queries
- Distributional analysis I.e. Segment diversity within audiences
- Regression analysis I.e. QRI (Audience Relevance) and CVR

82%
of audiences are
updated once or never

63%
of spend reuses 3P
segments

75%
of users ≤ 4 segment
searches before
selection

Behavioral data suggests advertisers may be:

- Using search as a navigation shortcut
- Selecting from the first few visible options
- Defaulting to familiarity over relevance

What about performance?

After controlling for segment diversity (Data Elements Per Impression) and data source (1PD vs 3PD), audience relevance is a strong predictor of conversion rate.

A 20-point increase in audience relevance is associated with an ~86% lift in CVR on average ($p < 0.01$), suggesting targeting quality drives performance more reliably than audience complexity or spend.

If relevance drives performance, why aren’t advertisers selecting the most relevant segments?

Hypothesis: Position Bias

When presented with ranked segment lists, advertisers may disproportionately select:

- Segments appearing earlier in the list
- Segments that feel familiar

If true, this implies:

Audience outcomes are influenced not only by data quality but also by interface ordering

Research Approach

Survey

To understand advertiser workflows at scale around audience creation

Stated Preferences

vs

A/B Testing

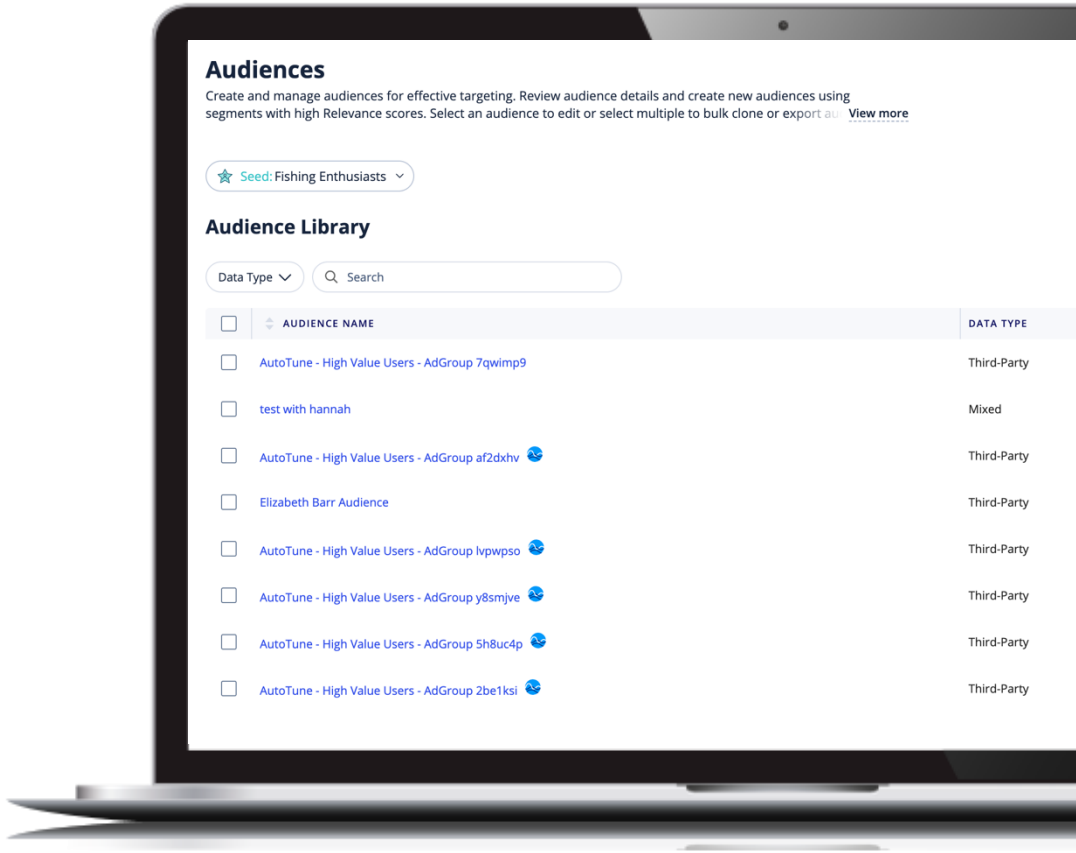
Run a randomized A/B experiment in Audience Builder

Revealed Preferences

Survey

To understand advertiser workflows at scale, we propose launching a survey measuring:

- Who defines audiences
- How audiences are constructed
- How segments are judged
- How hard it is to find segments
- Confidence in evaluating segment quality
- Importance of:
  - Price
  - Scale
  - Freshness



A/B Test

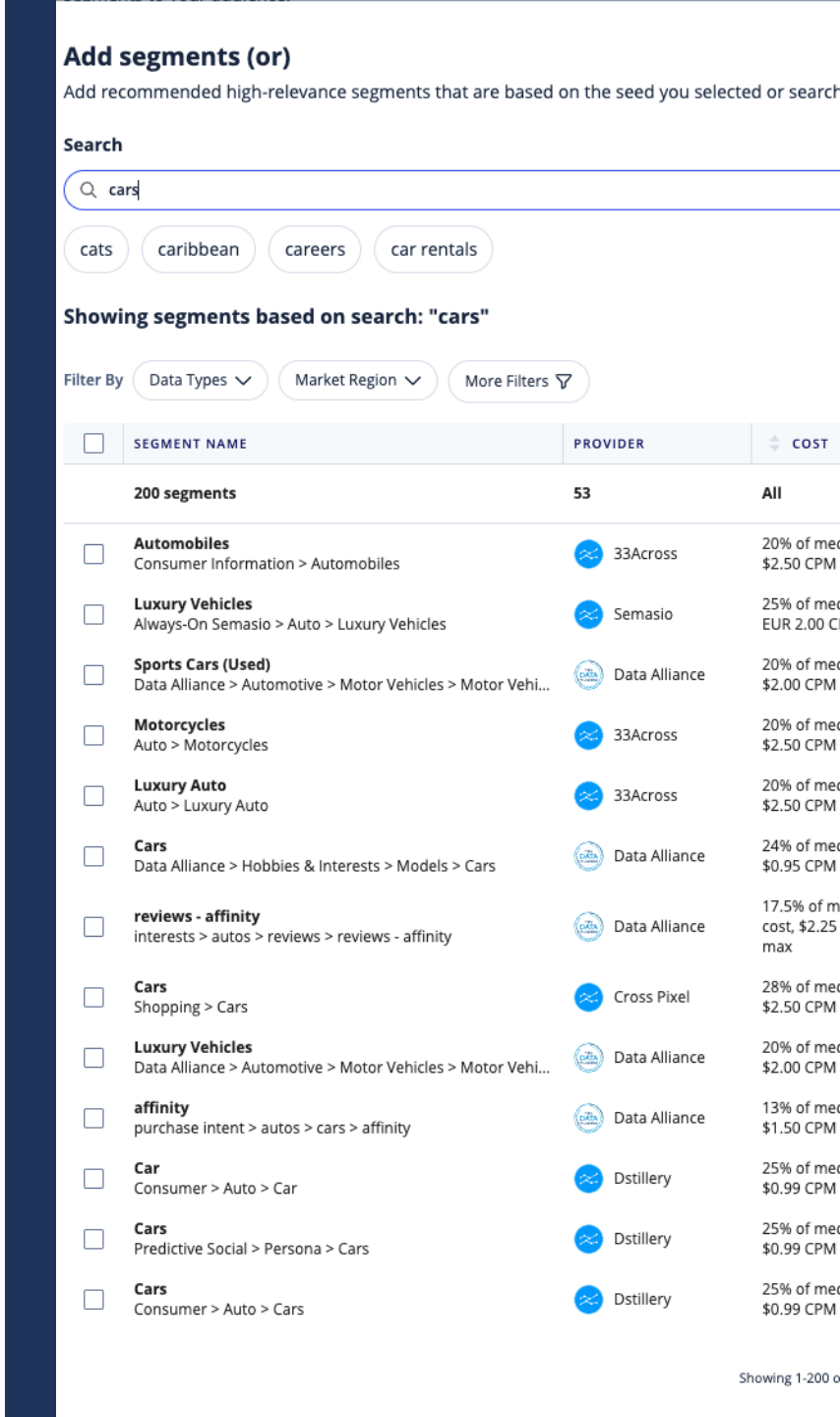
To To behaviorally test position bias we ran a randomized A/B experiment in Audience Builder:

Group	Segment Ordering
Control	Ranked by search relevance
Treatment	Reverse-ordered

No changes to:

- Segment content
- Relevance
- Metadata

Only list position differs.



A/B Test

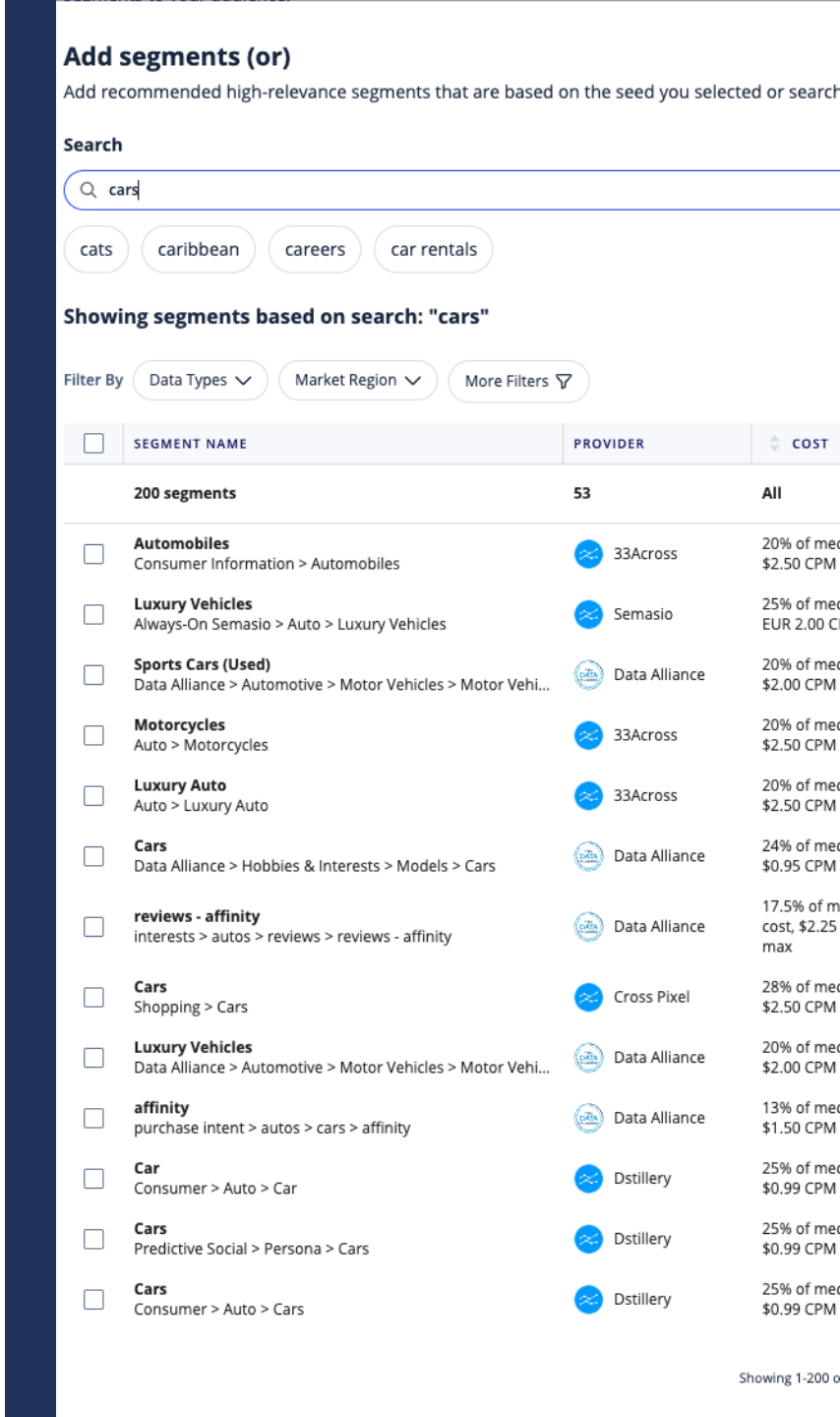
- **Randomization**: Search Query Event Level
- **Ranked by Relevance Score** : We don't change the segment pool, only the order they are displayed.

Field	Example
Search Term	cars
Assigned Group	treatment
Original Rank	8
Displayed Rank	1
Selected?	1

Segment Selected = 1 if selected , 0 otherwise

Independent variables:

- Displayed Position (1–10)
- Original Relevance Rank
- Segment Price



Findings

Survey

31% indicate audience strategy is defined externally but

68% rely primarily on search to explore segments

For 49% automation is viewed with caution — advertisers want transparency and control

19% rate relevance score as important

A/B Testing

Probability of selection decreases as position drops even when relevance is held constant

Top-3 Displayed Segment: 4.2× more likely to be selected

Segments originally ranked 7–10 have a selection rate of 41% when ranked Rank 1–3

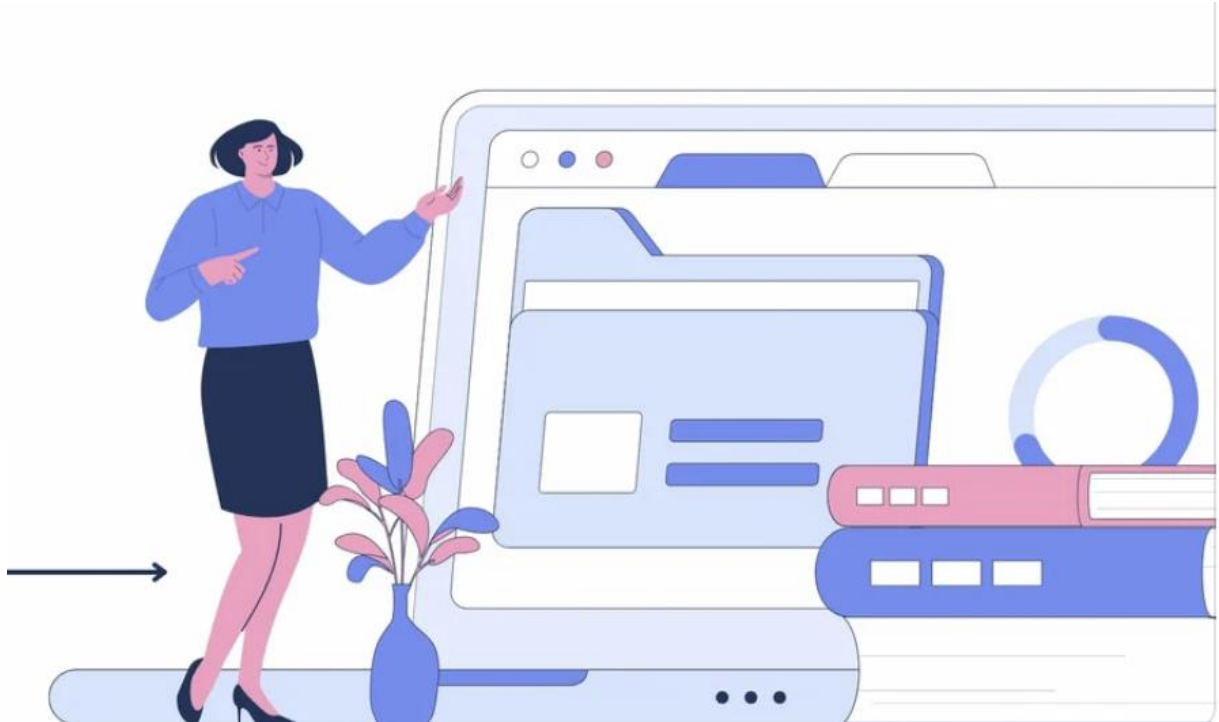
Position bias hypothesis proven

Product Recommendation

1. We need to close the loop between reporting (where users see impact of the audience) with audience selection
2. Promote high-performing searches: Improve back-end relevance scale to display not only semantic relevance
3. Communicate better in the UI how the audiences are ranked
3. UI reminders to update audiences and recommendations on good fits

“Make it easier for clients to select the good segments”

Measuring Clients' Trust & Satisfaction While Improving Product

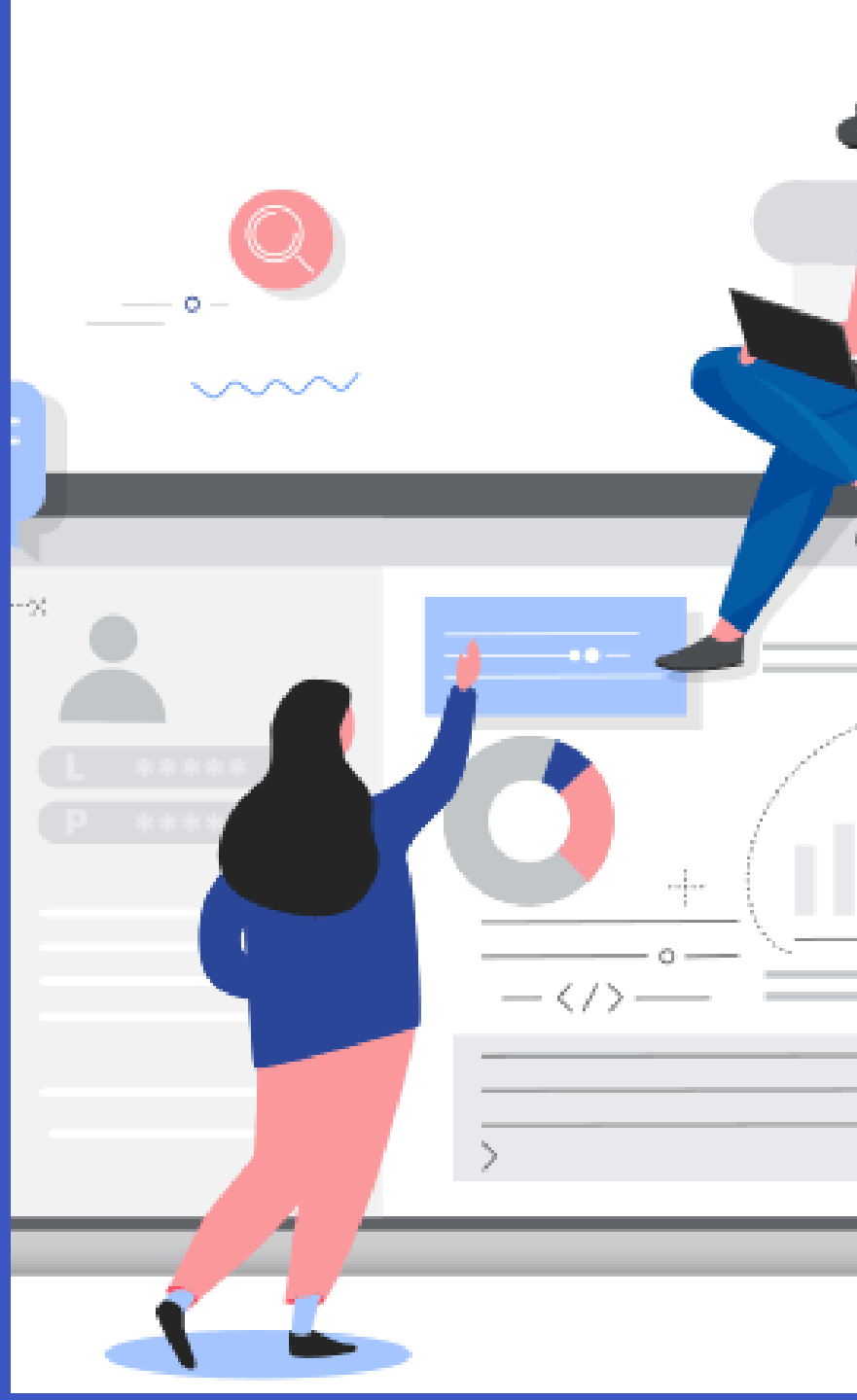


“If you can’t measure it, you can’t improve it.”

– Peter Drucker

The problem

- Lack of systematic way to measure **user satisfaction** with the different workflows in the platform.
- No insight into downstream effects: the company couldn't see how workflow friction cascaded into broader platform perceptions
- The company had no reliable way to detect where users were struggling—or why some workflows felt more frustrating or inefficient than others.
- Product development risk: decisions based on **anecdotes rather than empirical evidence**.



The problem

Leadership & Client Services care about customer trust

Product wants concrete guidance on how to benchmark their areas

Approach

Survey Design to

quantify UX Quality, Trust, and **Satisfaction**

Goals

- **Create a single, trusted signal of customer experience** – Establish a consistent quarterly view of CSAT and UX quality across TTD so leadership has one shared source of truth for experience health.
- **Track progress and impact over time** – Measure trends quarter over quarter to understand whether investments are improving ease-of-use, effectiveness, and overall customer satisfaction.
- **Focus and prioritize what matters most** – Identify and prioritize the most severe UX issues within key product areas and workflows so teams address the highest-impact problems first.

UX Quality

We measured two outcomes per 10 workflows:

Ease of Use

Ease of Use refers to *how simple, intuitive, and low-effort a workflow feels for the user*.

Effectiveness

Effectiveness refers to the user's perception of whether the workflow *successfully accomplishes what they need it to do*.

CSAT

- Customer Satisfaction Score (CSAT) measures customer satisfaction with a business.
- It’s one of the most straightforward ways to measure customer satisfaction, and it's obtained by asking a simple question, such as 'How satisfied were you with your experience?’
- To answer, there's a corresponding 1-5 survey scale

What is Trust?

What is Trust?

Reliable

The product is dependable, consistent, and performs as expected every time.

Motives

The underlying intentions and motivations and alignment with the interests of customers and stakeholders.

Integrity

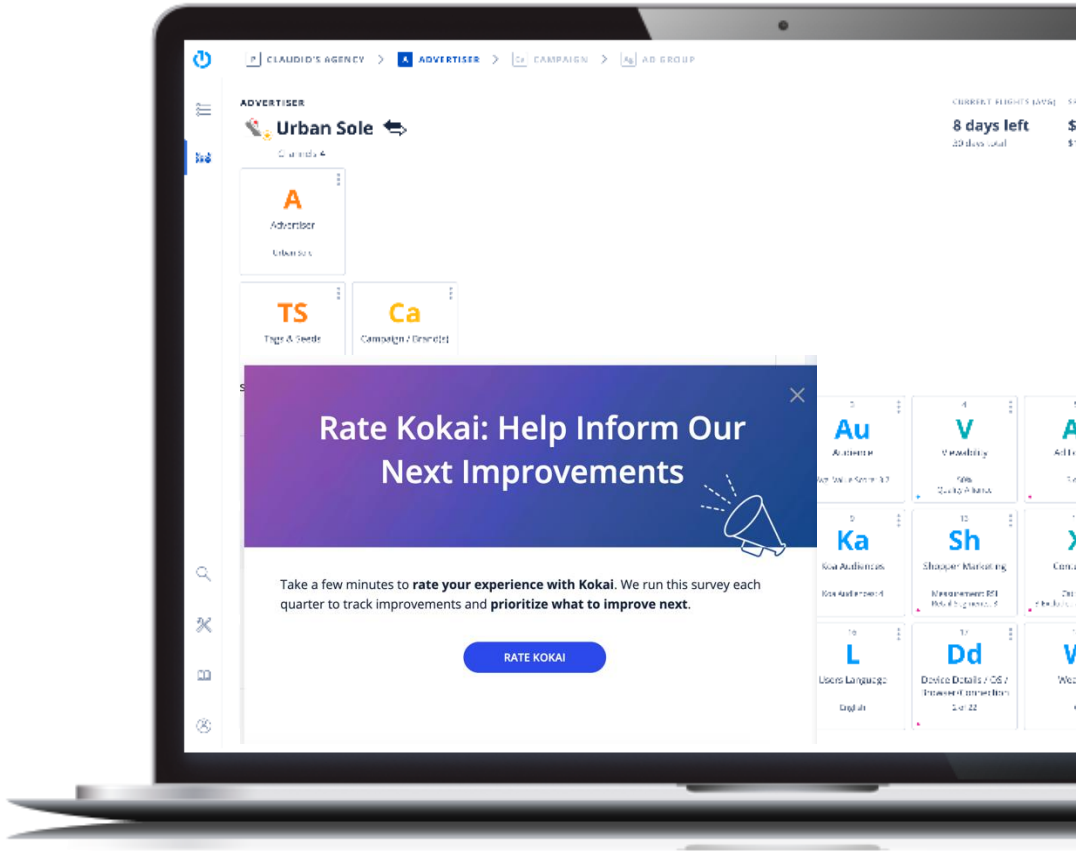
The ethical and responsible practices of a company such as transparency, integrity, fairness, and adherence to ethical standards.

Delivers Results

The perceived ability of a company to deliver quality product and achieve desired outcomes.

Sampling & Fielding

- **Frame:** Active external users in last 30 days with ≥3 sessions
- **Method:** In-product invite; stratified by role, spend tier, and region
- **N (3000):** 1000 completes
- **Scale:** 5-point Likert; “approval required” reverse-coded
- **QC:** Speeders removed, attention check, partials <80% dropped; minimal imputation



Analysis Plan

Descriptives

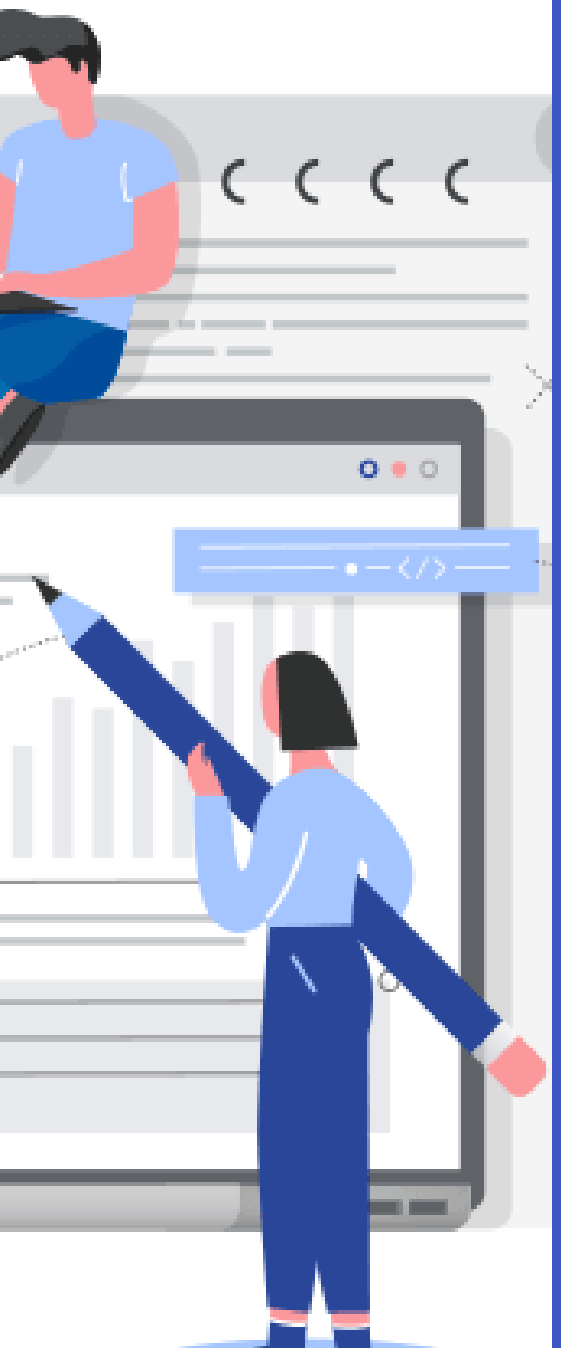
- Data cleaning
- Distributions
- Poststratification weights

Measurement validation (CFA)

- Loadings
- Reliability (α/CR, AVE)
- Discriminant validity (HTMT)

Structural Modeling

- MIMIC (Multiple Indicators, Multiple Causes) SEM: estimate direct and indirect paths
- Estimator**:** WLSMV (ordinal Likert) with bootstrapped CIs; controls
- Robustness: Model fit (CFI/TLI/RMSEA/SRMR)



What we learned

We evaluated a clear, theory-driven chain:

Workflow Time → Workflow Ease → Platform Ease → Platform Effectiveness → CSAT → Trust

Workflow ease flows upward to Platform Ease

Platform Ease is a latent construct built from users' ease-of-use experience across workflows. It's the shared signal across all workflows.

Not all workflows contribute to ease of use equally. The biggest drivers of how "easy" the platform feels overall are:

- **Editing campaigns** (β = .39 Importance ~36%)
- **Creating campaigns** (β = .33 Importance ~26%)

Users judge the platform's ease of use *primarily* based on these core workflows. Together, these account for ~62% of the shared ease-of-use signal.

Platform Ease Predicts Platform Effectiveness

Platform Effectiveness is another latent construct that reflects how well the platform helps users meet their needs across workflows.

Platform Ease is the dominant driver of Platform Effectiveness, accounting for ~64% of how users judge whether the platform works for them.

When users feel the platform is easy to use, they are much more likely to feel the platform helps them accomplish their goals.

Effectiveness is the Key Driver of CSAT

Platform effectiveness explains ~67% of the user satisfaction (CSAT)

The *majority* of CSAT is explained by whether the platform helps people accomplish their goals.

Ease → enables productivity
Productivity → drives satisfaction

Trust is a real, stable construct

The four trust items (Reliable, Motives, Integrity, Delivers Results) all load strongly onto a single underlying “Trust” dimension.

This means:

- Users clearly have a **coherent mental model of trust**
- We can measure and influence it reliably
- Improving upstream experience (ease → satisfaction) can shift trust meaningfully

The experience chain for users

We see a strong, intuitive progression in how people form opinions about the platform:

- **If workflows feel easier →**
- **The platform feels more effective →**
- **Users feel more satisfied →**
- **and ultimately trust increases.**

This chain holds tightly across all respondents. It shows that *improving the experience isn't cosmetic—it's foundational to satisfaction and trust.*

(Chain: (β = .65 → .73 → .46; all p < .001)

Time ≠ Engagement

Users feel friction—and it's measurable. When people spend more time in a workflow, they consistently report it feeling harder to use.

This friction shows up most clearly in:

- **Editing** (strongest pain)
- **Creating**

This tells us: **time is a real signal of user struggle**, not just an operational metric.

Spending more time in a workflow consistently predicts *lower perceived ease*

Model showed that more time in a workflow is a signal of struggle—not engagement—unlocking a new, actionable way to interpret user behavior.

More Time ≠ More Engagement

40%
of sessions that contain Inventory Controls also contain Inventory Selection

18%
of sessions that contain Inventory Selection also contain Inventory Controls

28%
of sessions that contain Creative Library also contain Creative Selection

10%
of sessions that contain Creative Selection also contain Creative Library

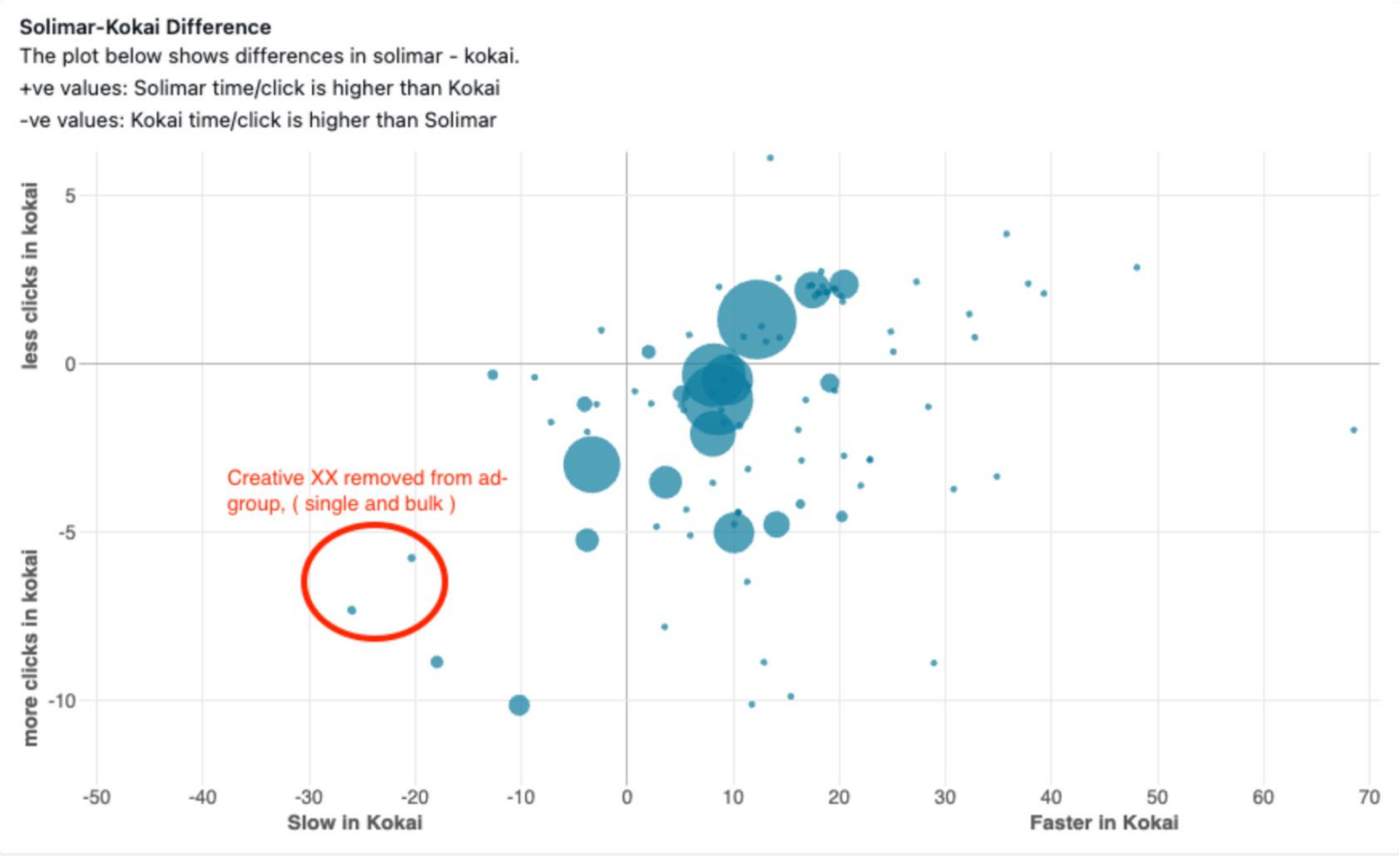
22%
of sessions that contain Campaign Overview also contain Campaign Settings

14%
of sessions that contain Campaign Settings also contain Campaign Overview

17%
of sessions that contain Ad Group also contain Budget

8%
of sessions that contain Budget also contain Ad Group

More Time ≠ More Engagement



Tailored Deliverables

Trust Index

Trust is calculated as the average across the 4 factors resulting in a score between 0 to 100.

Factor Score Rating of the individual factors comprising Trust.		How to Improve (Generalized)
77	**Reliability**	Improve communication channels with clients, improve transparency of process
74	**Competence**	Assure clients that TTD can meet their advertising goals and deliver on our promises.
73	**Motives**	Demonstrate TTD decisions and actions are guided by a genuine commitment to the best interests of our clients.
72	**Integrity**	Convey necessary information for clients to understand TTDs actions are just and honors commitments to partners and the Open Internet.

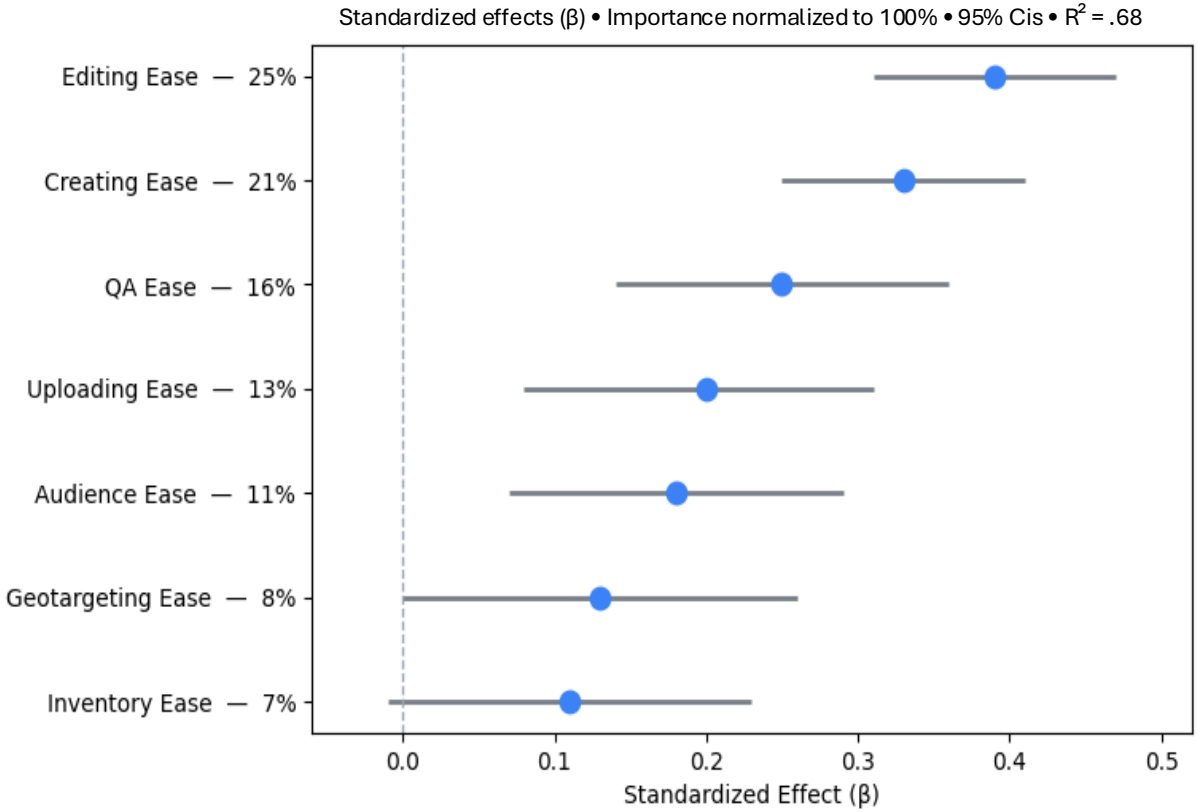
Trust Index

74.8

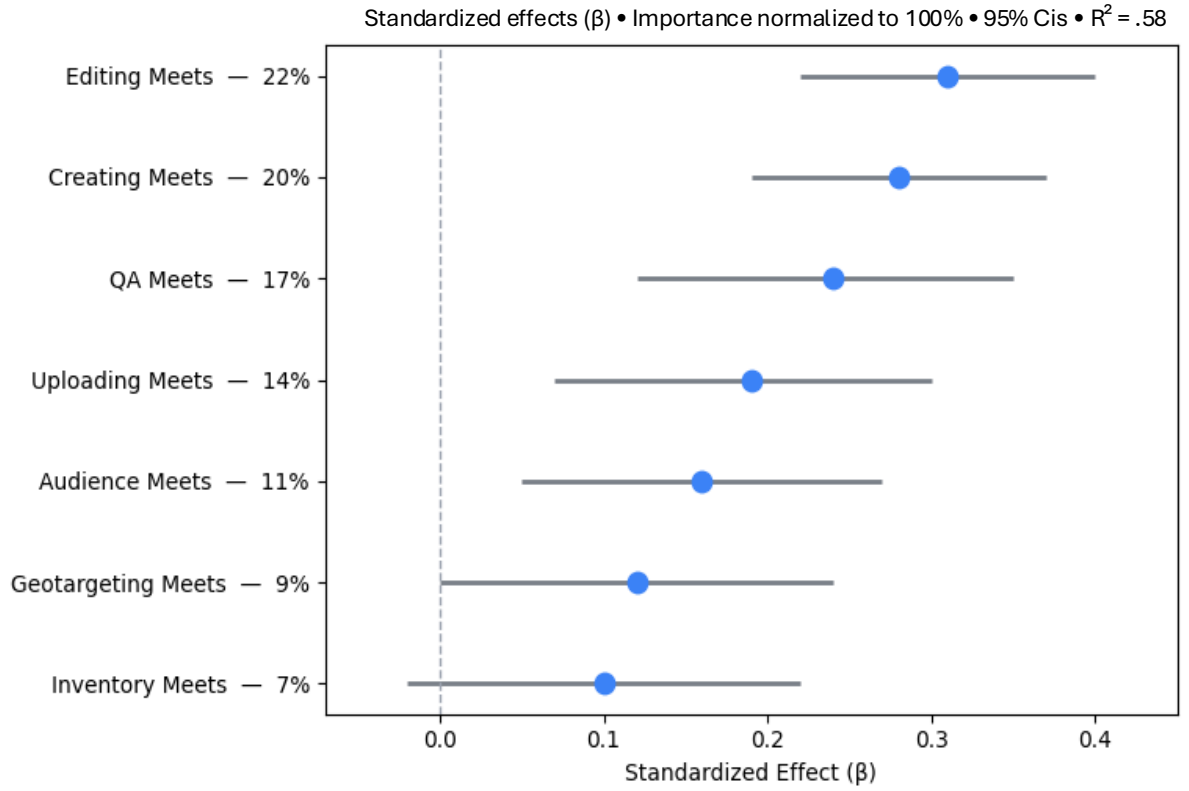
Out of 100

Driver Maps

What Drives Platform Ease?

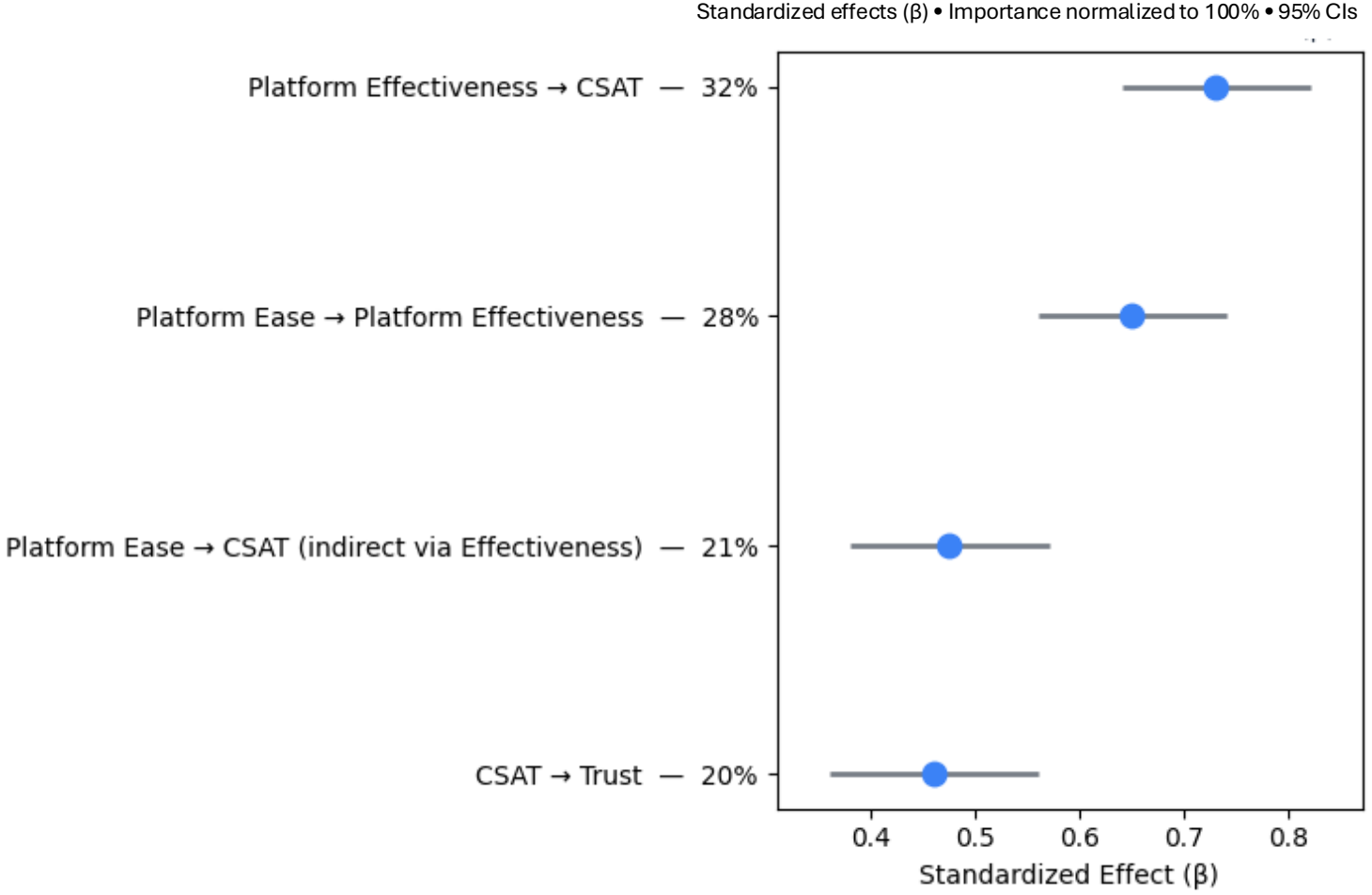


What Drives Platform Effectiveness?

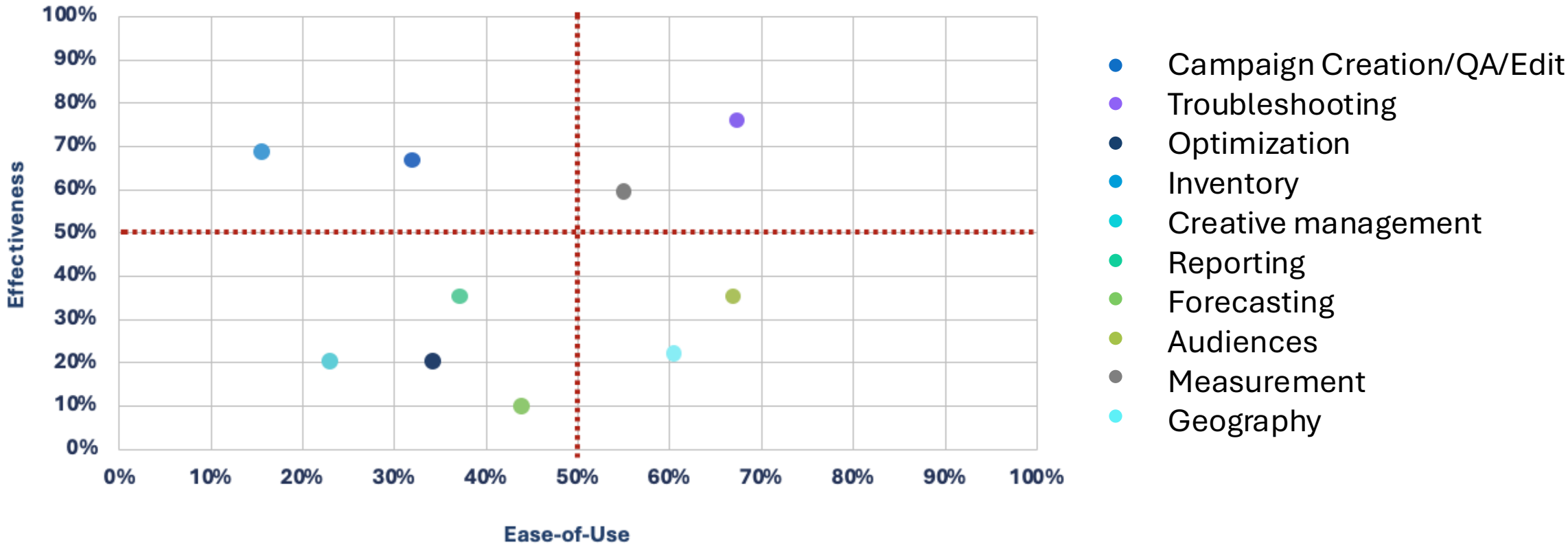


Driver Maps

Platform Chain:
Ease – Effectiveness- CSAT- Trust



Product Area Comparisons



Hold Co vs Indie

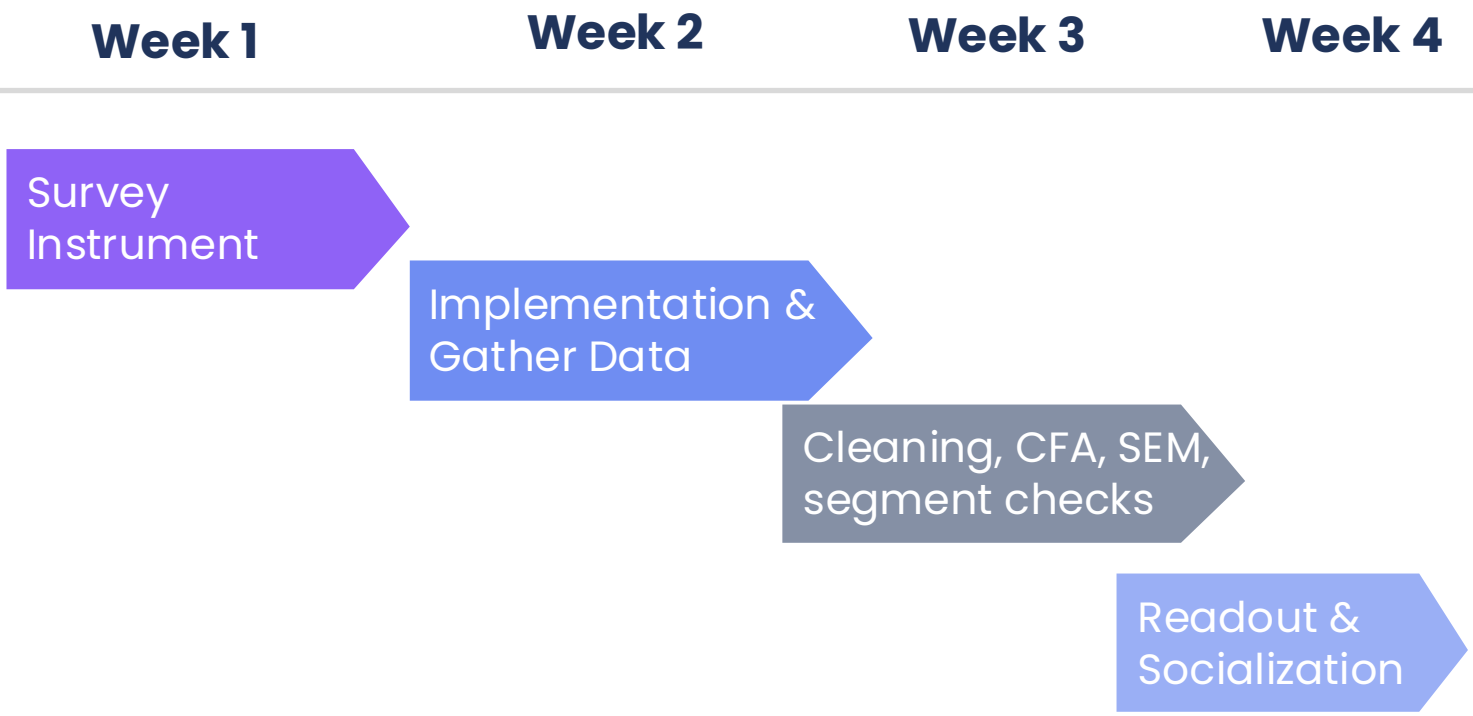
	Hold Co	Indie Agencies
Top CUJ	Monitor, Report, Troubleshoot, Create Campaigns	Monitor, Report, Troubleshoot, Create Campaigns. Manage inventory more than Hold Co*.
CSAT	Slightly higher than overall CSAT	Slightly lower than overall CSAT
Trust in TTD	Trusts TTD (85/100)	Trusts TTD (84/100) but trust KOA less than Hold Co*
% Time Spent in Kokai	79%	62%*

* Statistically significant differences

Collaboration & Feedback

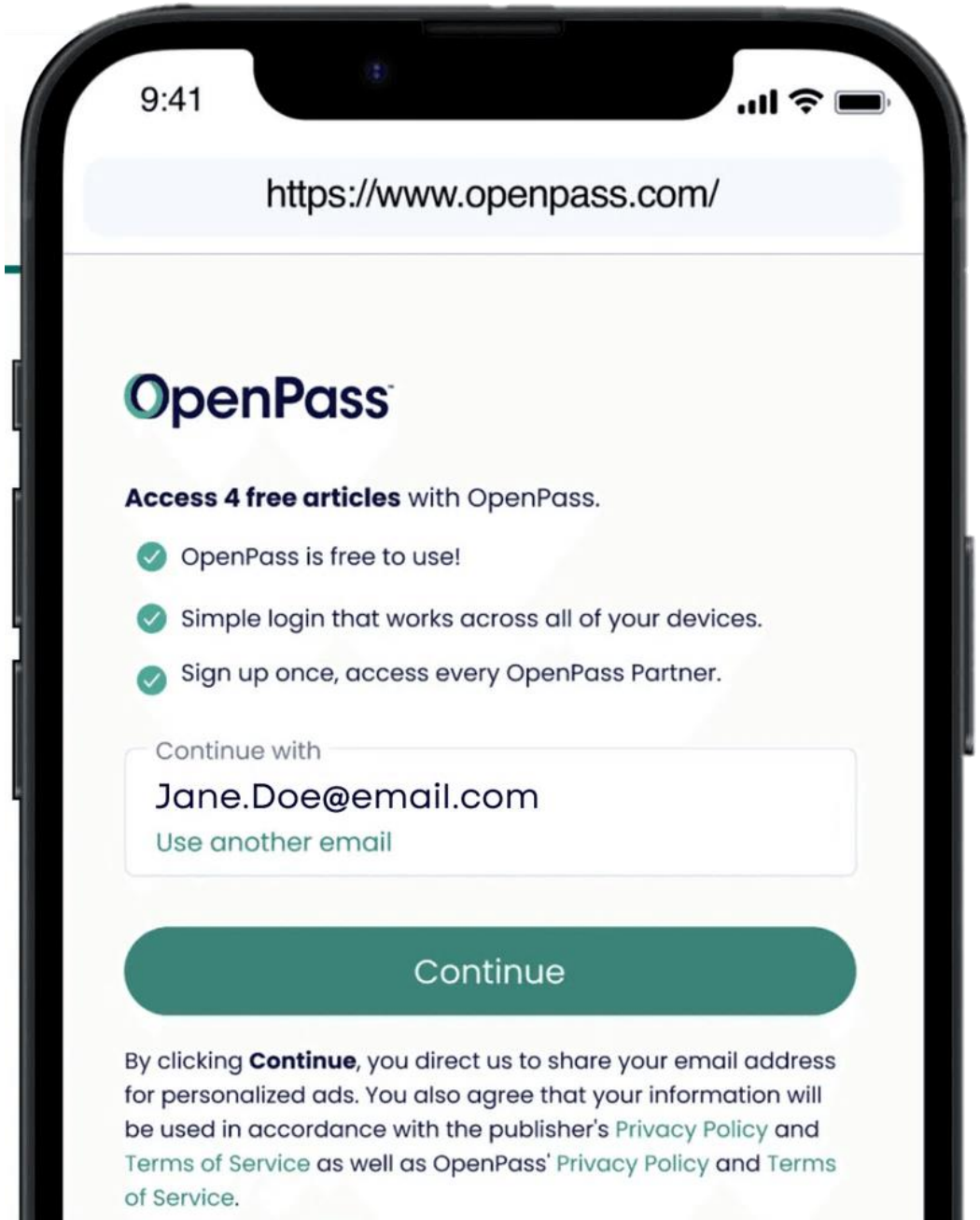
- Product Area PMs
- Data Science
- Qual UXR
- Content team
- Client Services
- Senior Leadership

Timeline


Week 1
Week 2
Week 3
Week 4
Survey Instrument
Implementation & Gather Data
Cleaning, CFA, SEM, segment checks
Readout & Socialization

OpenPass

A/B Survey > A/B Testing



OpenPass is The Trade Desk sign-on (SSO) solution that allows organizations to increase authenticated users on their websites and apps.

Consumers can access premium content across the open internet using just their first name, last name, and email address (using their phone number is optional).

Objective

Identify what type of sign-in text format and content is preferred by users to sign-in into OpenPass

Problem

Many types of copy, design, not enough sample size,

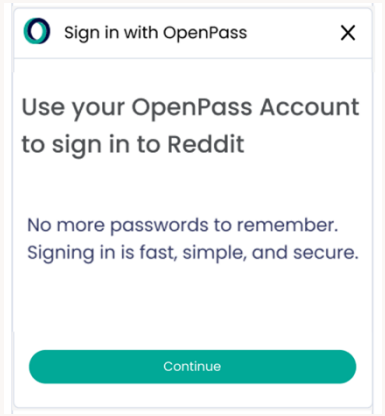
NOT ENOUGH TIME!

A/B Survey

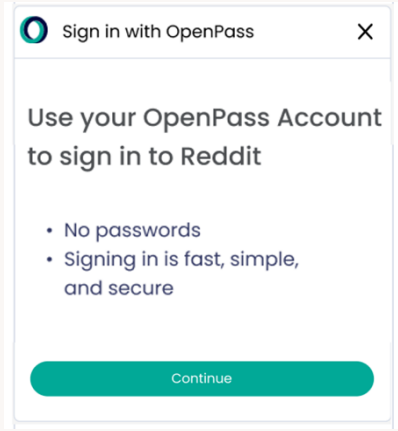
Conjoint Attributes and Levels

1. Text Format

Text

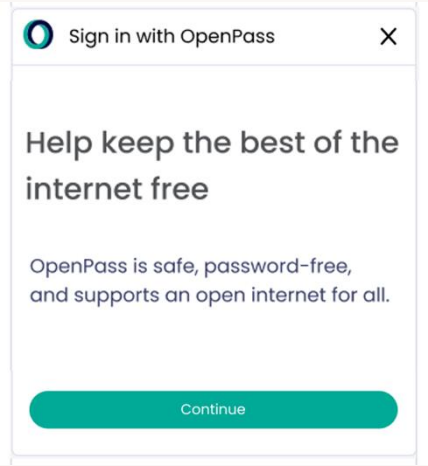


Bullet

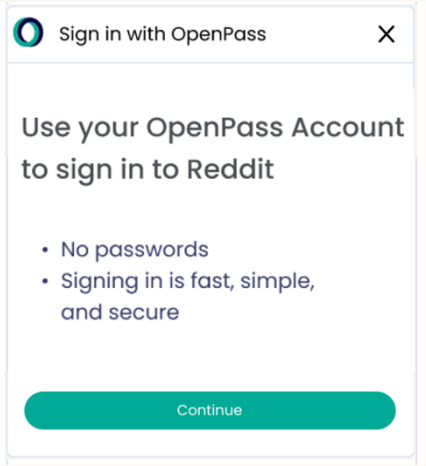


2. Content

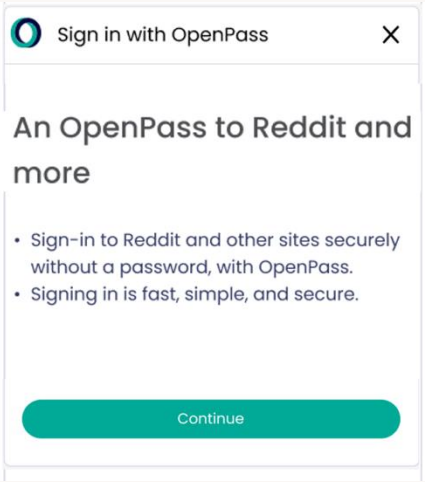
Free



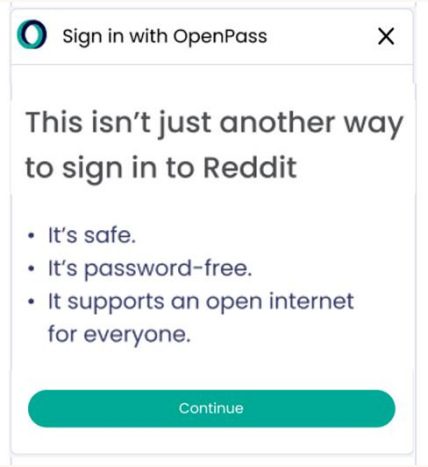
Neutral



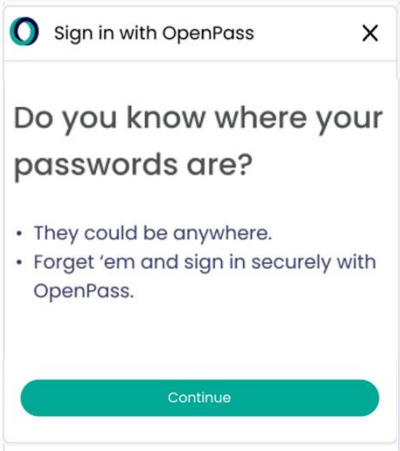
More



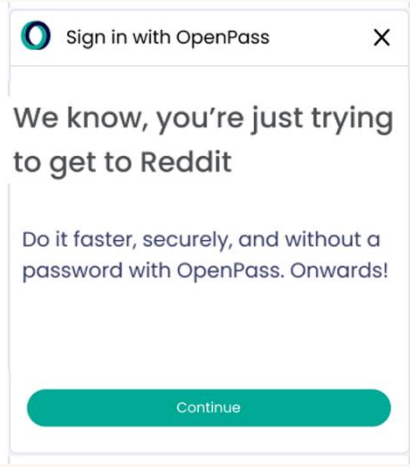
Intrigue



Fear



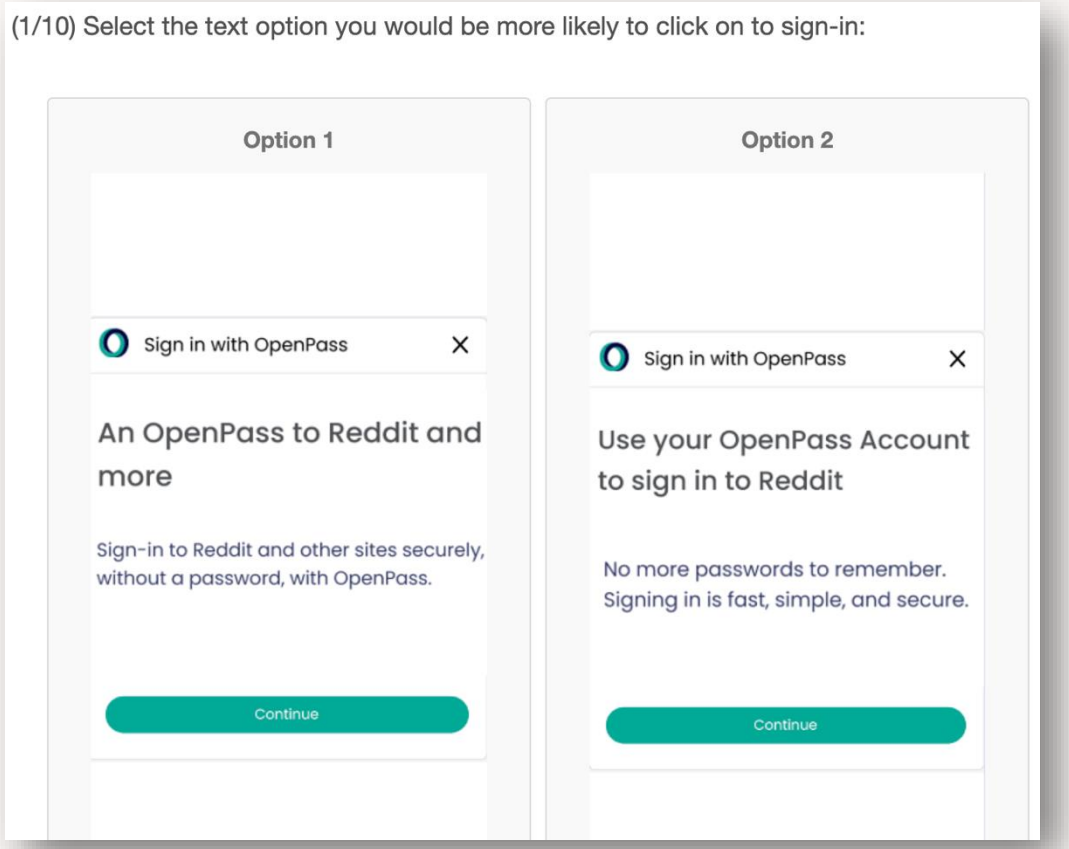
Humor



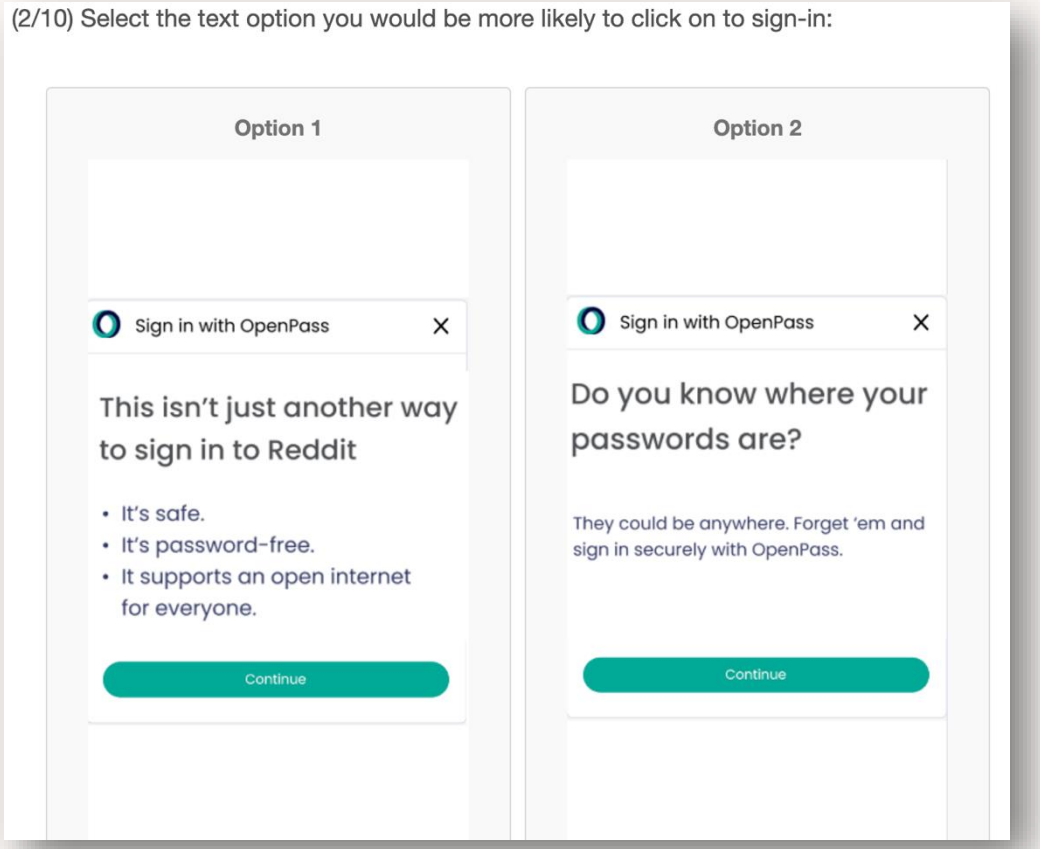
Conjoint Methodology

Each of the 200 participants complete 10 tasks: in each tasks they choose the option/scenario where they are more likely to complete the sign-in. For example:

Task 1:

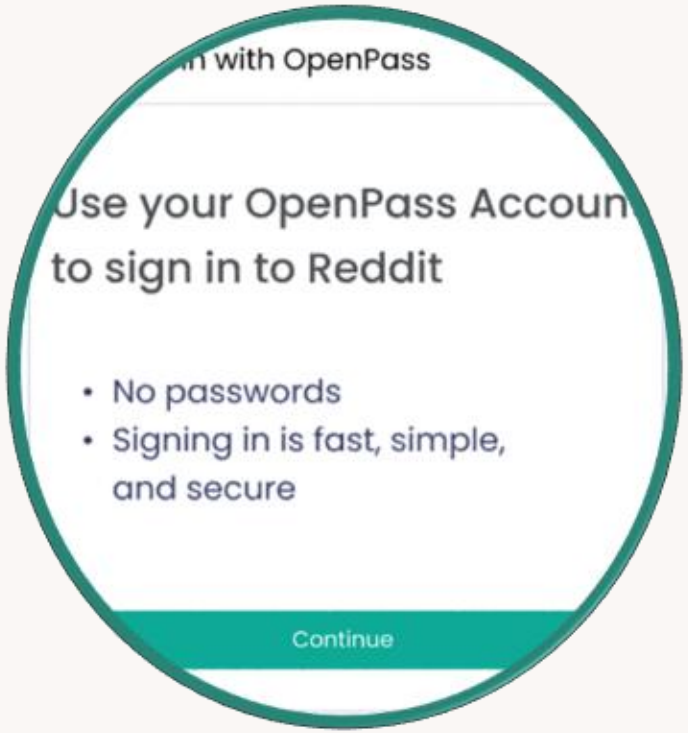


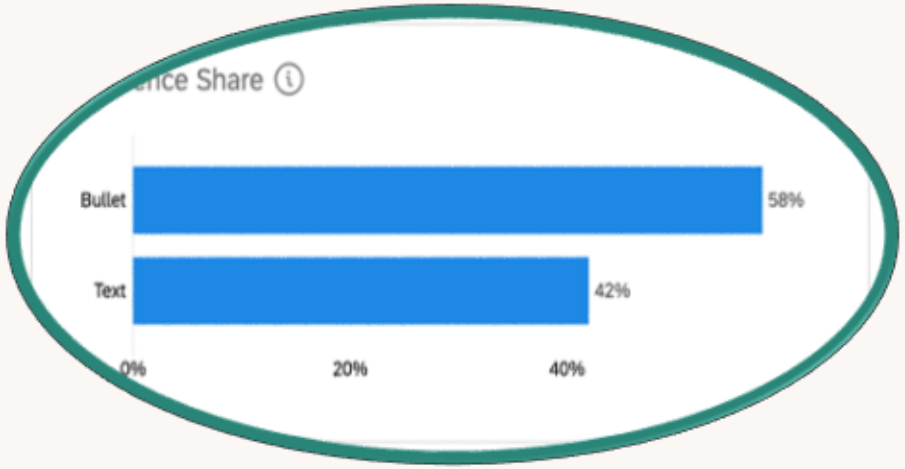
Task 2:

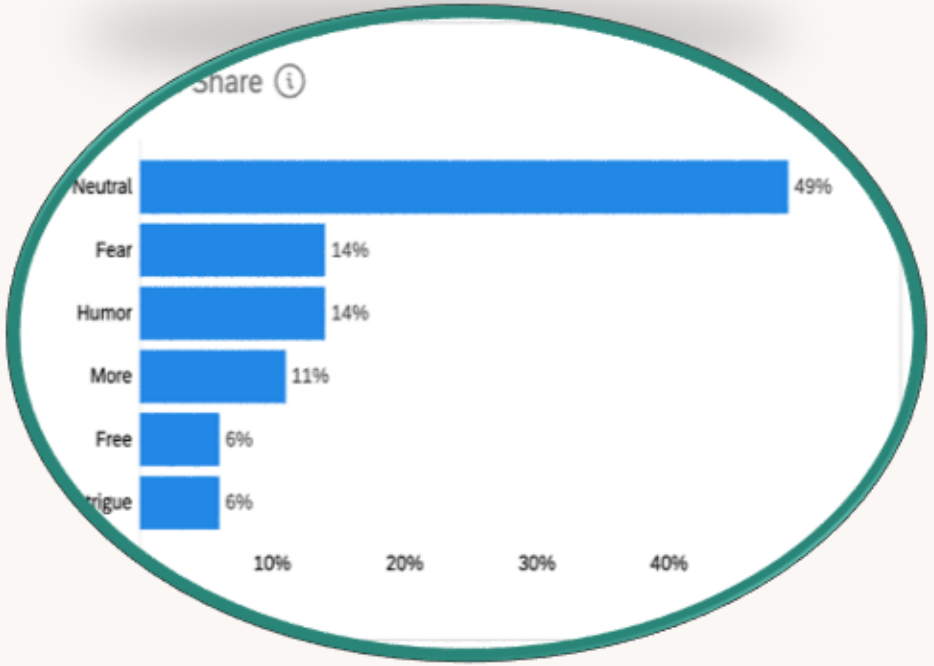


Conjoint Methodology

Based on the combination of text format and content, the preferred sign-in by users is the neutral sing-in, combined with bullet points.







Benchmarking OpenPass Sign-In Experience

The Problem

Publishers hesitate to adopt OpenPass without proof that its **sign-in experience matches or exceeds competitors**.

Existing perceptions of usability are influenced by **brand recognition**, not the actual product experience.

Lack of empirical evidence makes it harder to **differentiate OpenPass** and persuade clients.

Research Challenge

- **Our leading's competitor strong brand recognition** biases user evaluations of sign-in flows.
- Directly comparing OpenPass vs. Competitor risks **unfair judgments** not based on usability.
- Needed a **fair, brand-neutral benchmark** to validate OpenPass's sign-in experience.

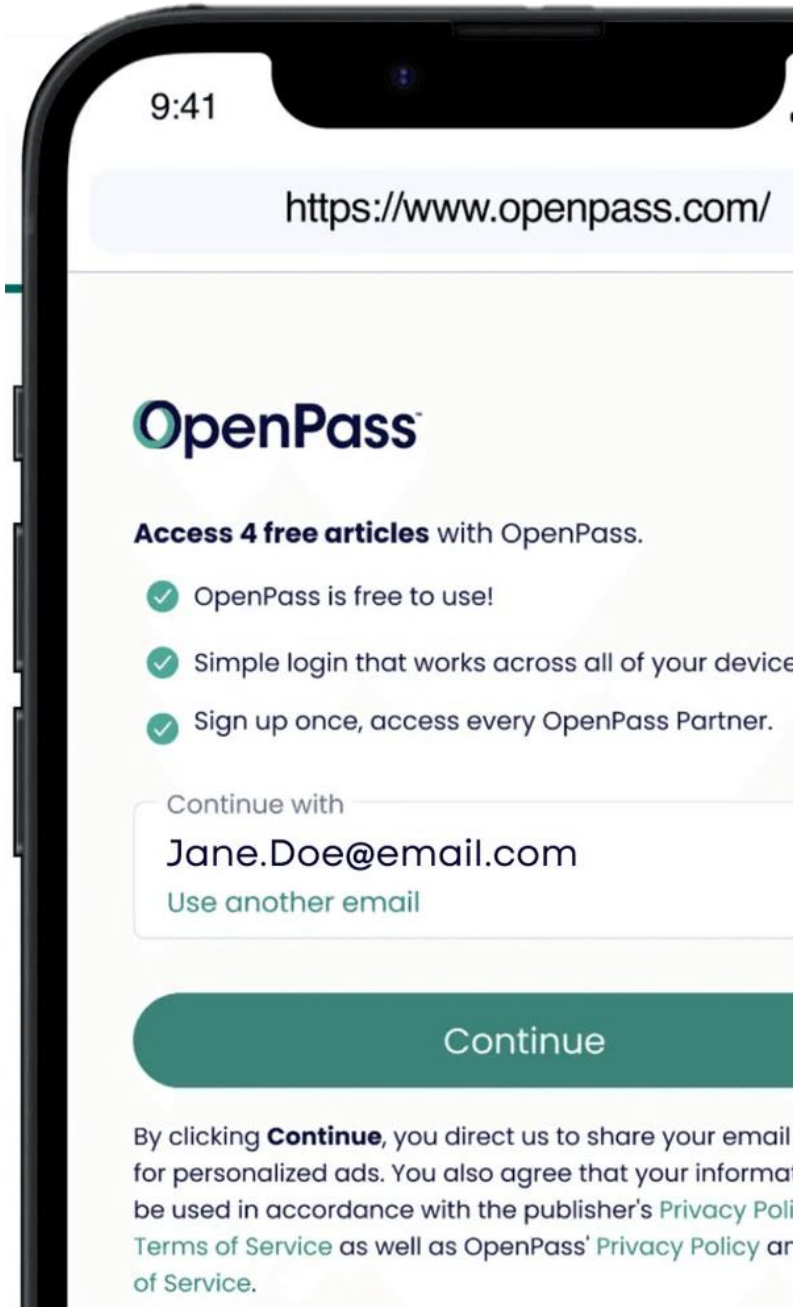
Research Approach

- **White-label competitor test:** Used Competitor's sign-in flow but replaced branding with OpenPass to remove brand bias.
- Conducted an **A/B test**:
  - Condition A → OpenPass UX
  - Condition B → White-labeled Competitor's UX
- Survey to measure and benchmark usability and satisfaction.
- In-depth interviews.
- Participants recruited via **dscout** (n ≈ 500).

Design

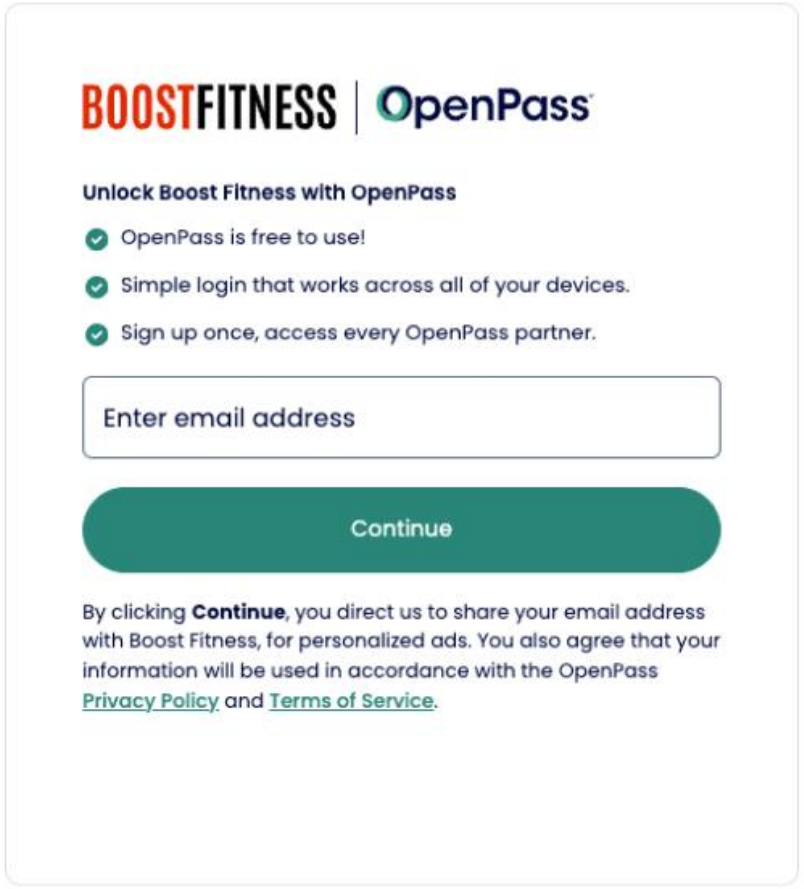
- Participants completed **two tasks**:
  - New user sign-in
  - Returning user sign-in

- After each task, participants rated:
  - **Overall experience** (1 = Very poor, 5 = Very good)
  - **Ease of use** (1 = Very difficult, 5 = Very easy)
  - **Ease of use vs. other providers** (Google, Facebook, Microsoft)
  - **Trust in provider's handling of private data**
  - Whether they **read Terms & Conditions**, and clarity if read

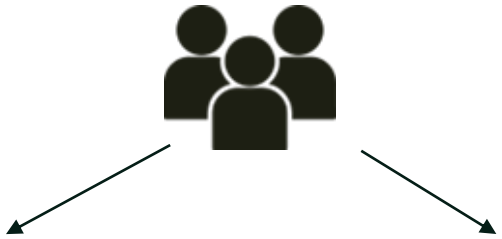


A/B Test

OpenPass Sign-In

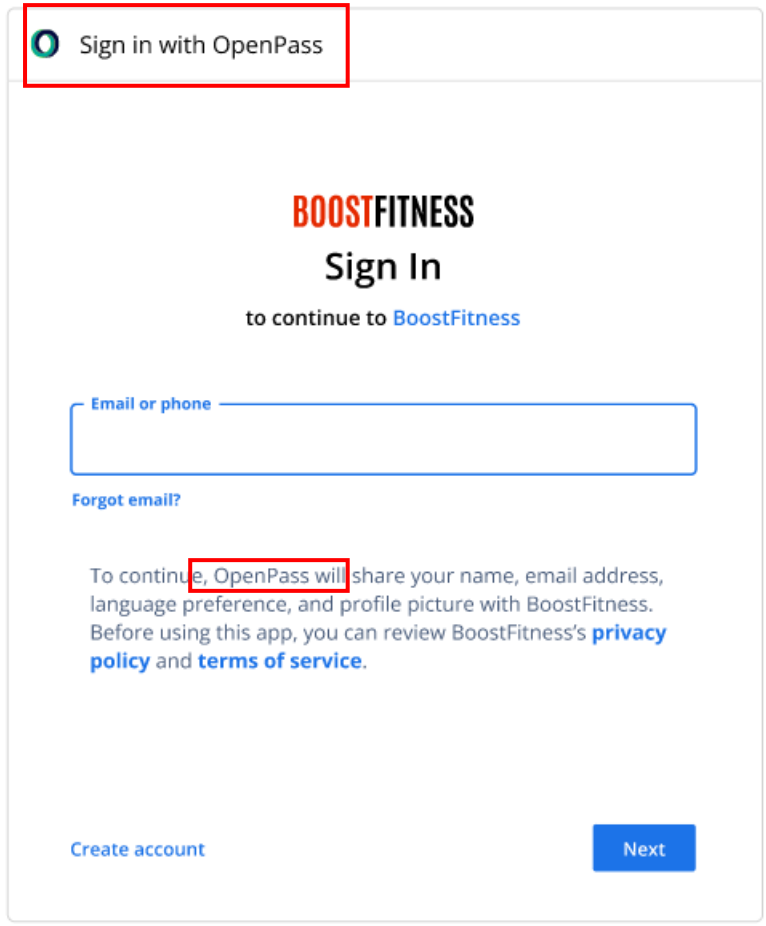


300 consumers

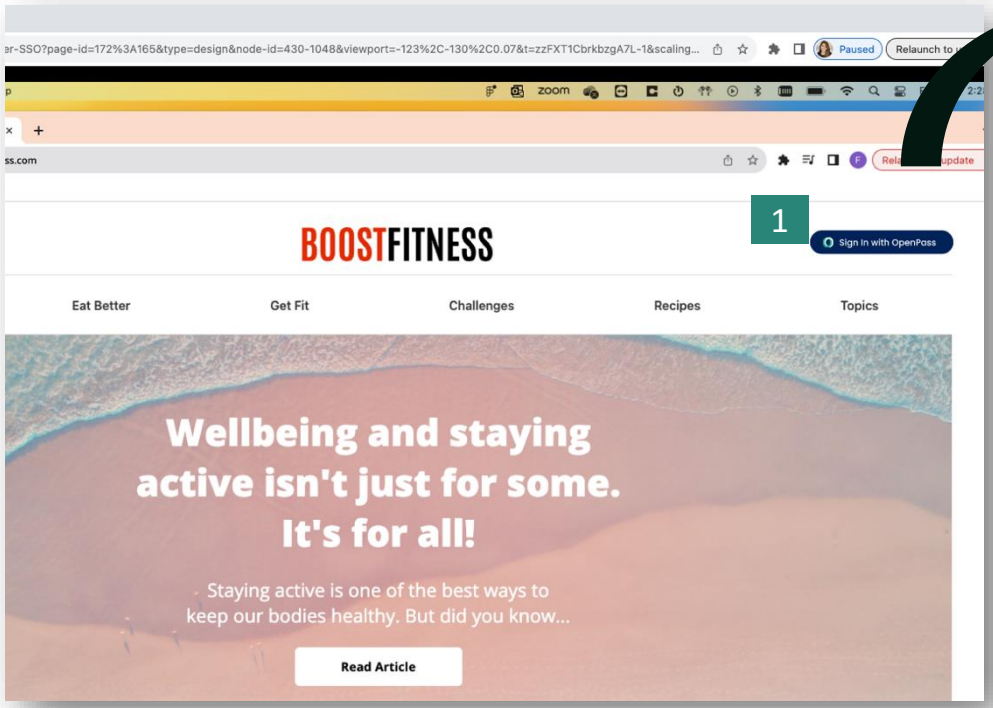


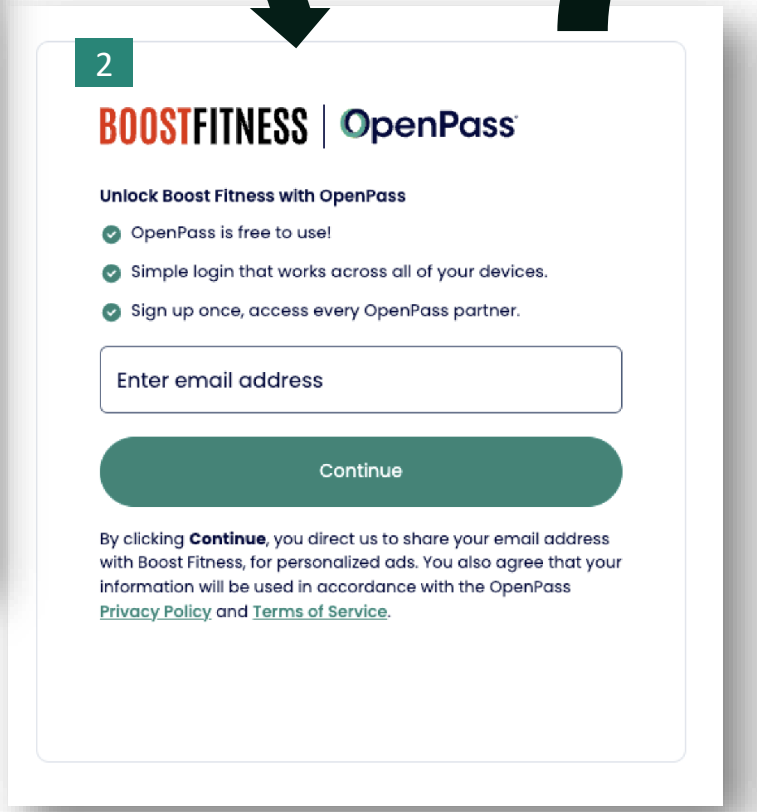
*Control for bias towards Google's brand recognition.
Fair test and comparison of the sign-in experiences.

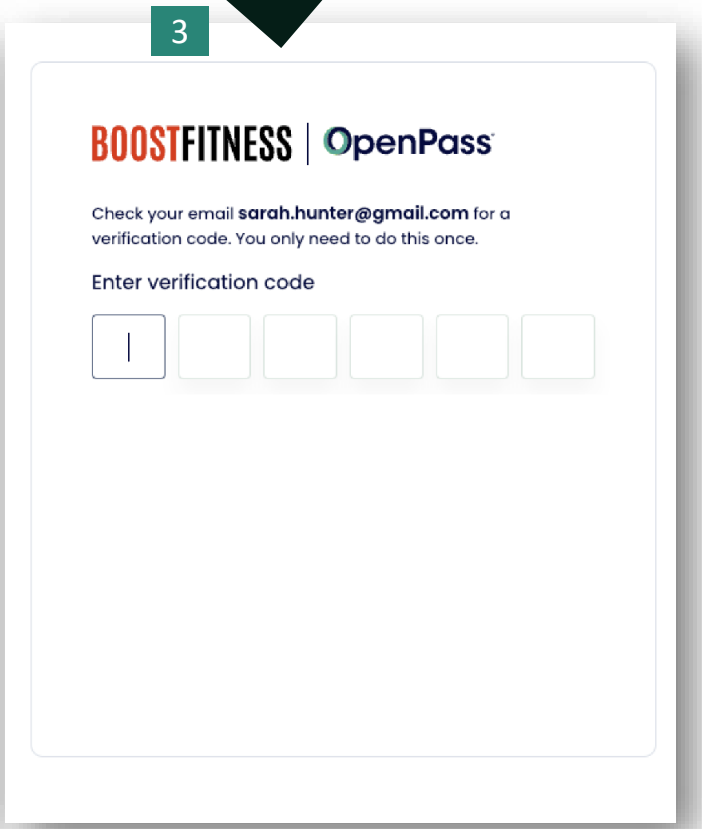
White Label Sign-In*



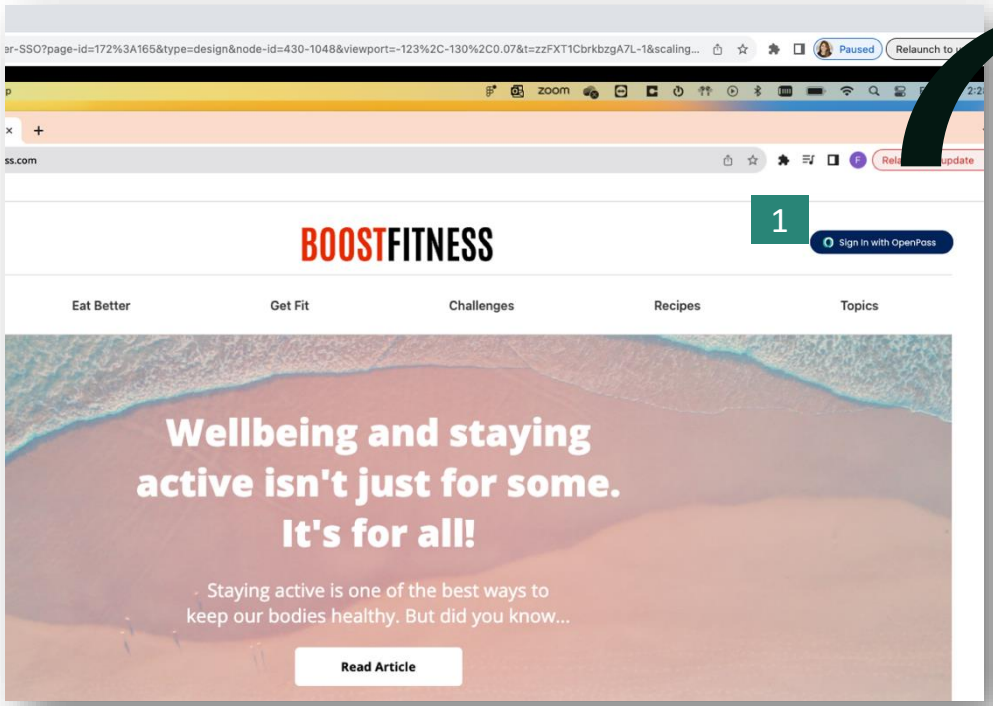
OpenPass Workflow

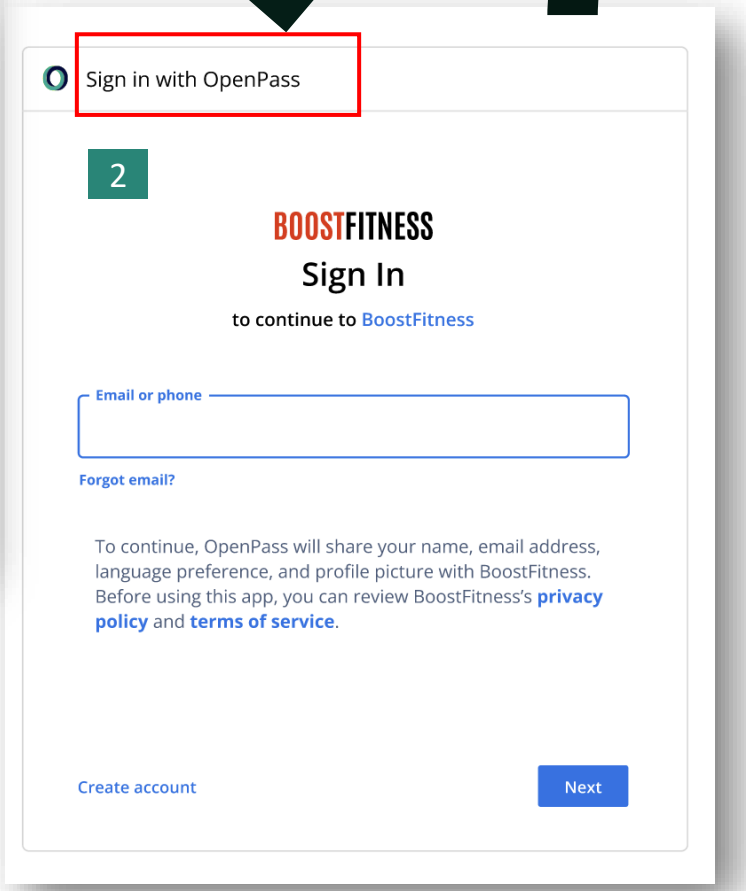


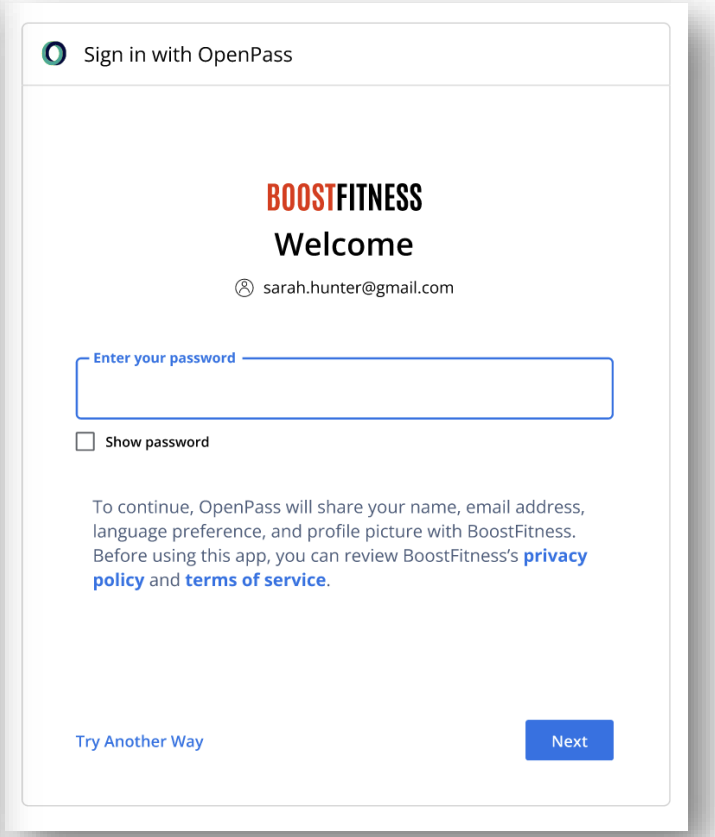




OpenPass Workflow







Survey Questions

Participants completed a new user and returning user sign-in in experience

For each of these experiences they were asked to rate:

1) Overall experience (1=Very poor, 5= Very good)
2) Ease of use (1=Very difficult, 5=Very easy)
3) Ease of use compared to other sign-in providers (1= Significantly difficult, 5= Significantly easier)
4) Rate trust of providers' handling private data (1=Very low trust, 5= Very high trust)
5) Did they read the terms and conditions (Yes, No)
    - If so, rate the clarity (1=Very unclear, 5=very clear)

What we found …

New User Sign-In Experience

1. Most participants (81%) reported a good or very good overall experience with the OpenPass sign-in, same as the white-labeled Google (81%). There is no statistically significant difference in overall experience between the two groups

2. Participants rated the ease of use for the OpenPass UX as 4.6 out of a 5-point scale, aligning closely with the white-labeled Google UX (4.7 points). There is no statistically significant difference in the ease of use between the two groups

3. Most reported the OpenPass UX was similar (64%) or easier (34%) compared to other providers (I.e., Google, Facebook or Microsoft). In white-labeled Google group, 46% reported it as easier compared to other providers (statistically significant difference)

OpenPass' UX ease of use is comparable to Google's

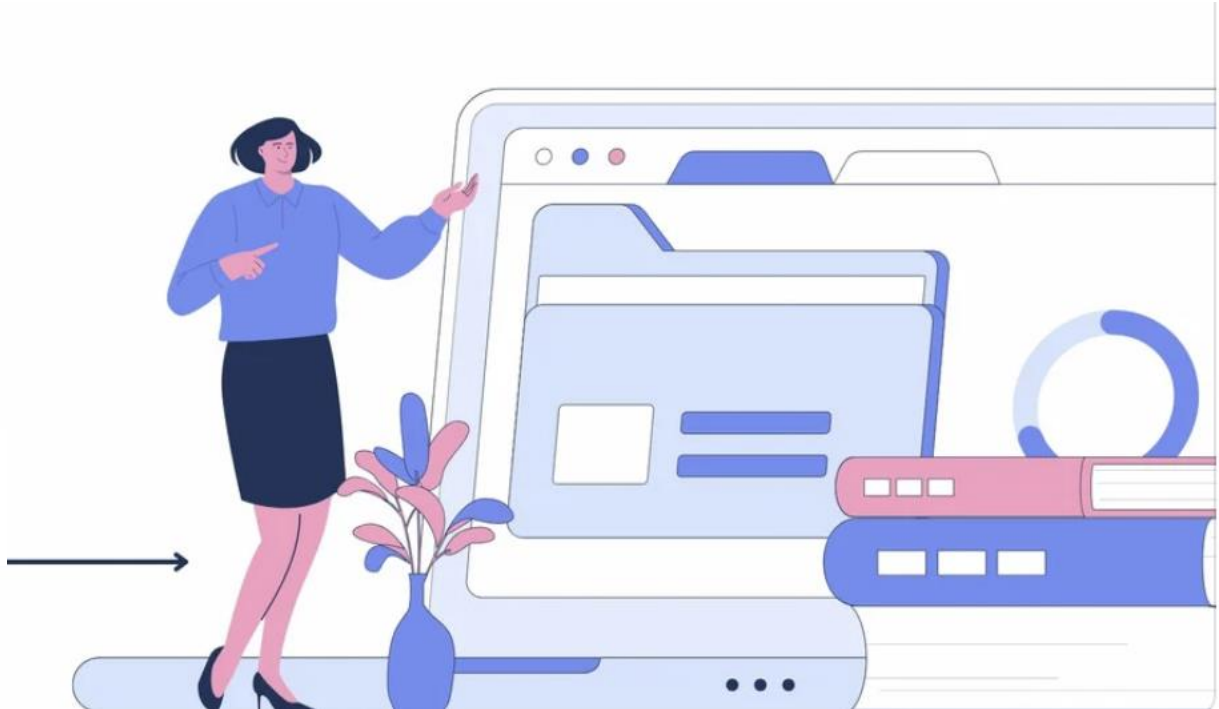
Research Impact

Participants completed a new user and returning user sign-in in experience

For each of these experiences they were asked to rate:

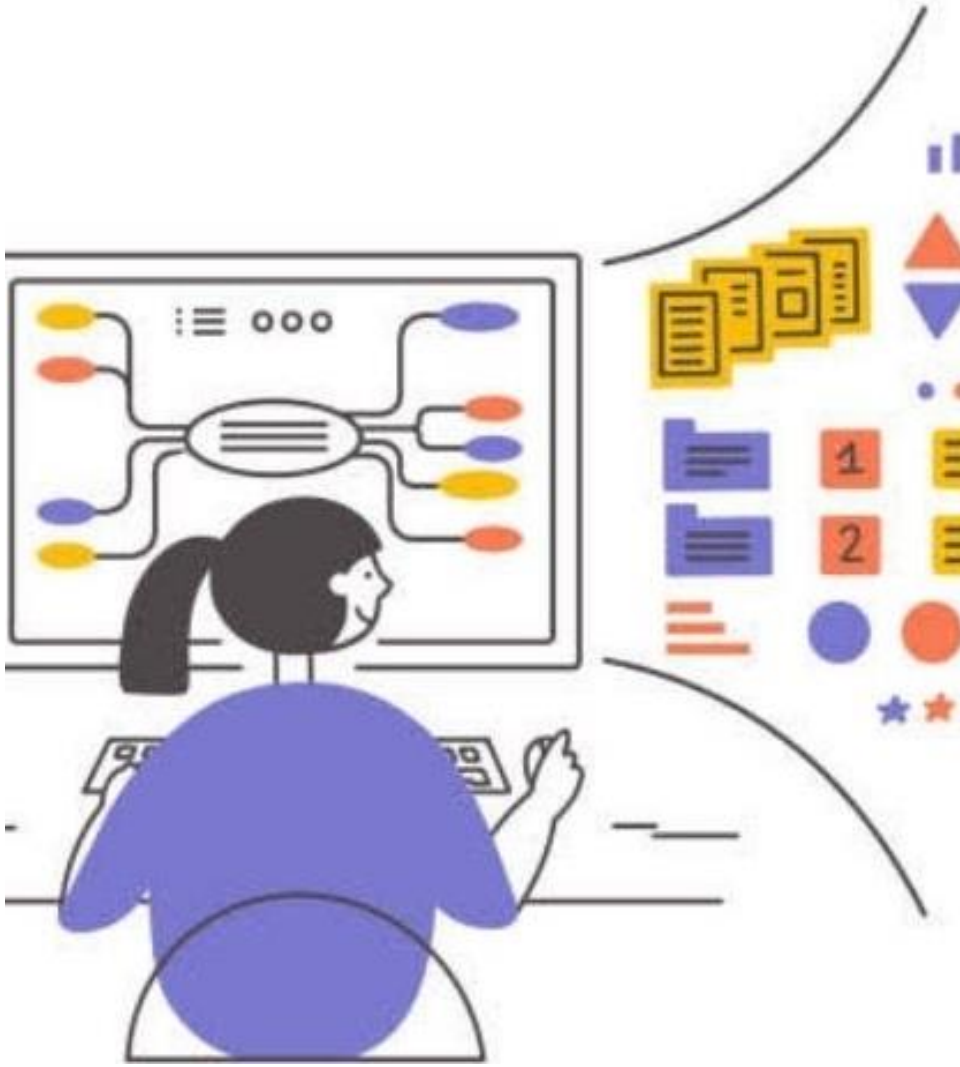
1) Overall experience (1=Very poor, 5= Very good)
2) Ease of use (1=Very difficult, 5=Very easy)
3) Ease of use compared to other sign-in providers (1= Significantly difficult, 5= Significantly easier)
4) Rate trust of providers' handling private data (1=Very low trust, 5= Very high trust)
5) Did they read the terms and conditions (Yes, No)
    - If so, rate the clarity (1=Very unclear, 5=very clear)

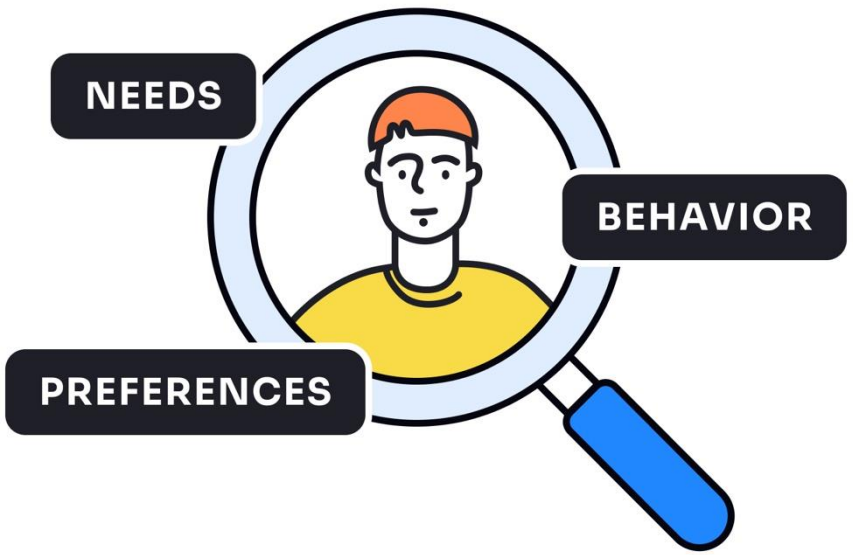
Bringing AI into User Workflows



The problem

- Internal users of the TTD platform face manage complexity across campaign setup, optimization, troubleshooting, and reporting workflows. Existing documentation and tools require context switching, deep platform knowledge and navigation.
- The goal was to develop an internal AI assistant to help improve the experience and efficiency of these tasks.
- Before the AI tool could be design there was a critical need to understand:
  - Where AI assistance would meaningfully reduce cognitive load rather than add complexity.
  - How accuracy, latency, and explainability impact trust in high-stakes advertising decisions.
  - How do users expect the model to answer their questions and support their workflow.





Why this matters

Without grounding in real user workflows, AI systems can increase cognitive load, surface incorrect guidance, or fail to earn user trust—especially when users rely on them for high-stakes decision-making.

The key is to shifts AI development from a purely technical exercise to a **user-centered system,** aligning model behavior, product design, and workflow integration before broader rollout. This ensures the AI release would be:

- **Useful**, not distracting, within real workflows
- **Trusted**, by reducing hallucinations and unclear reasoning
- **Actionable**, by supporting decisions at the moments users need help

Research Approach

Technical Evaluation & Model Learning

This track focused on improving model performance and operational readiness through real-world usage and feedback.

- Guided training sessions where users completed real tasks while interacting with the AI, generating high-quality prompts, corrections, and failure cases
- Continuous feedback collection through in-product signals and structured review sessions to identify hallucinations, unsupported queries, and data misinterpretations
- Analysis of response quality, latency, and error patterns to inform prompt optimization, guardrails, and iterative model improvements

User-Centered Experience & Workflow Integration

This track focused on ensuring the AI fit naturally into existing workflows and supported real user needs.

- Participatory design sessions with users to map workflows, identify high-friction moments, and co-design where AI assistance would be most valuable
- Usability testing and qualitative interviews to evaluate comprehension, trust, and decision-making when using AI-generated guidance
- Iterative validation of integration concepts to ensure the AI reduced cognitive load rather than introducing new steps or distractions

Design

Model Training & Guided Usage Sessions

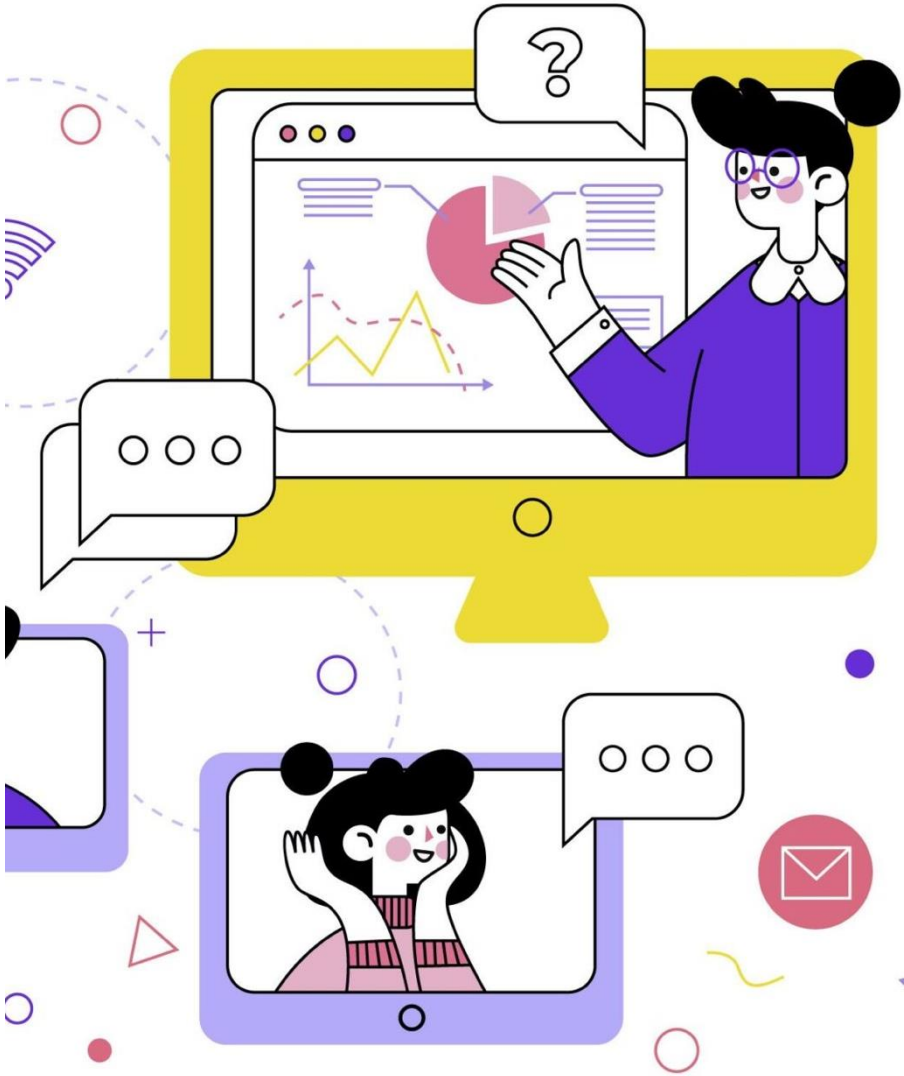
Structured sessions where participants use the AI tool with real tasks. These sessions generate labeled feedback, corrections, and failure cases to inform model tuning and prompt optimization.

Ongoing Feedback & Evaluation

Users provide in-product feedback and participate in follow-up sessions to assess response quality, trust, and workflow alignment. This captures both behavioral data and qualitative insight into where the AI succeeds or breaks down.

Design & Usability Testing Workshops

Collaborative sessions with users to map workflows, identify decision points, and co-design where AI assistance is most valuable (e.g., setup, troubleshooting, reporting). Outputs include opportunity areas and integration concepts grounded in real usage.



Research Impact

This research directly informed both model development and product design by:

- Prioritizing AI capabilities based on real workflow needs
- Identifying integration points that reduced cognitive load and context switching
- Producing labeled feedback and failure cases for ongoing model improvement
- Aligning stakeholders around a shared understanding of when AI adds value—and when it should stay out of the way

The outcomes enabled more confident iteration, reduced risk ahead of broader release, and ensured the AI evolved as a trusted workflow partner, not just an experimental feature.