
# 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 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

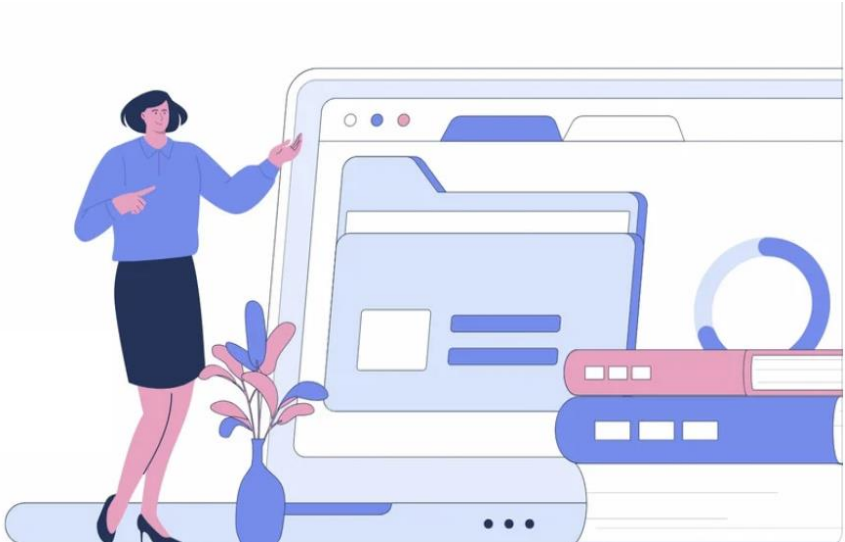
•Descriptive and inferential statistics
•Regression analysis (linear, logistic)
•Factor analysis & dimensionality reduction
•Segmentation & clustering
•Power analysis and sample size estimation

### Interviews

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

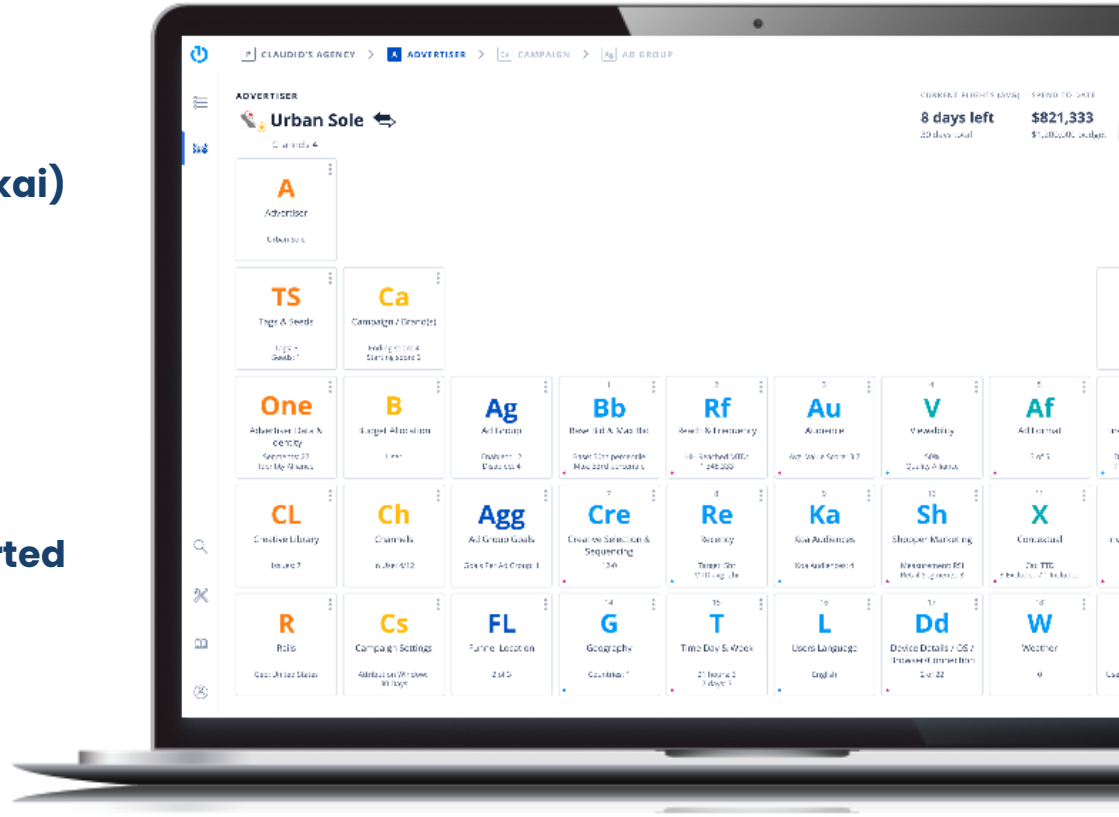
### Observational & Generative Methods

•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



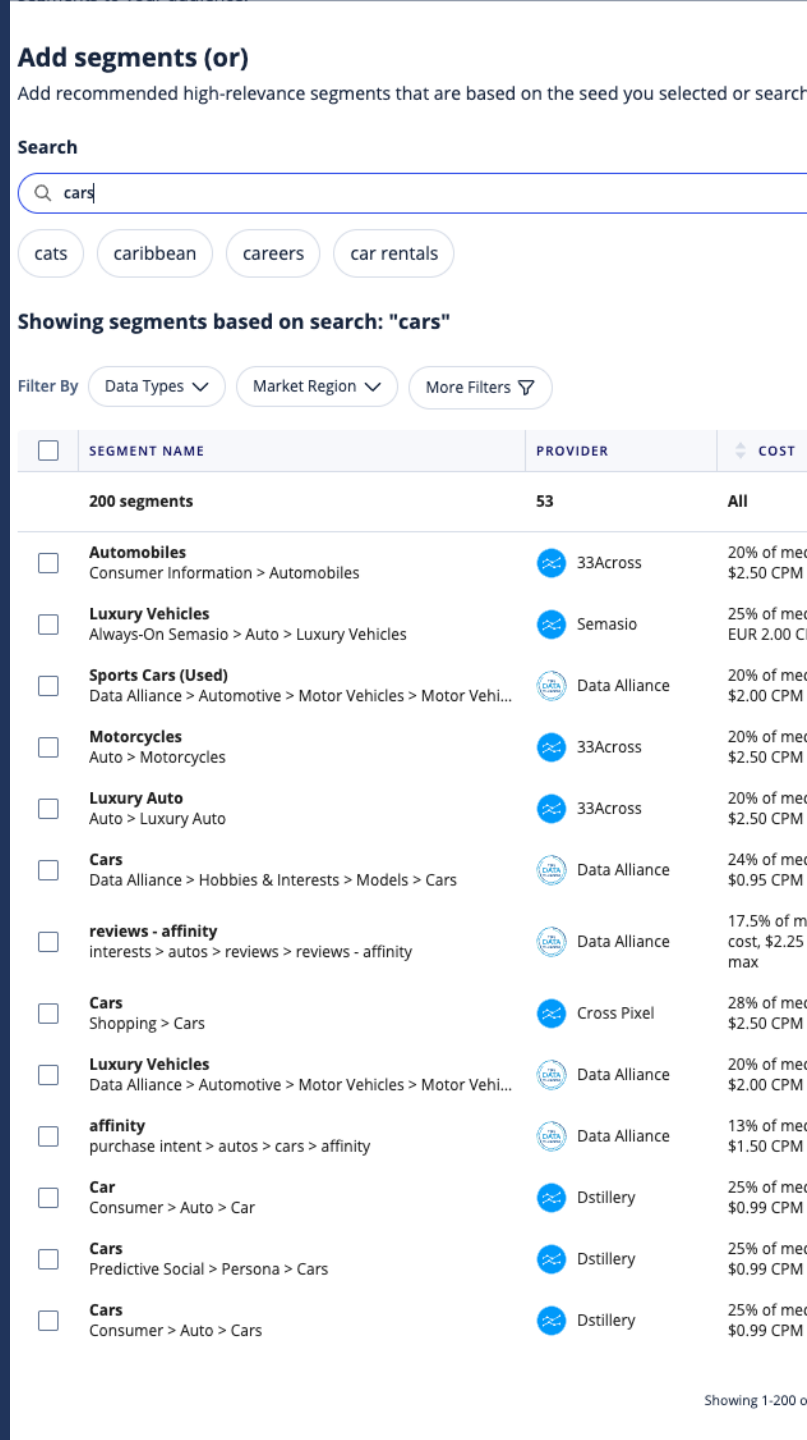
# theTradeDesk®

1. **Demand Side Platform (S&P 500)**
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:** OpenPass
6. **Agentic AI for Supported Workflows**



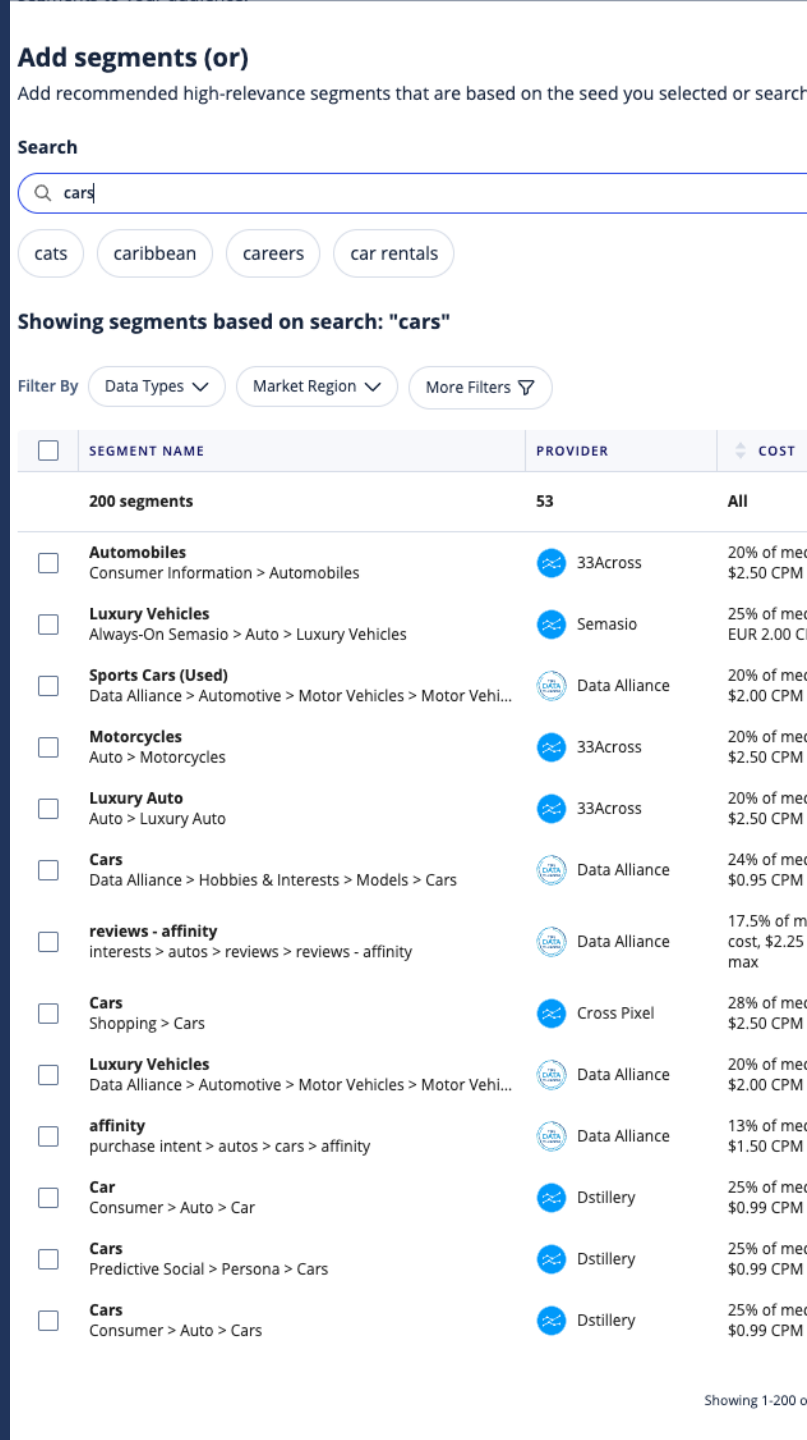
BUILDING THE TTD AI ASSISTANT

# How we solved workflow pain through quality AI



Phase 1

# We need AI... For what?



HOW IT STARTED

# Leadership said we needed AI.

## *No one had said AI for what.*
## *No one knew where to start.*

I proposed approaching AI as any other product feature by figuring out “what problem is AI trying to solve?”.

Phase 01

# Research Goals

## 1 Map where traders struggle

Identify which tasks are the most time-consuming, and which areas clients reach out the most about.

## 2 Identify critical AI support opportunities

Pinpoint the workflows where an AI assistant would save the most time and reduce the most friction.

## 3 Quantify the opportunity

Produce a dataset tying question frequency, effort, and perceived opportunity to concrete numbers for leadership and product prioritization.

## 4 Validate at scale

Have confidence in the insights are built on data, not just based on a handful of conversations.

Phase 01

# Survey at Scale

## Who

450 Active traders across segments — the people who live in the platform every day and feel the friction firsthand

## What

Open ended and Structured Likert-scale instruments measuring workflow easiness, efficiency, client question frequency, effort, and opportunities for AI.

## Output

Ease of use and effectiveness score per workflow, ranked critical workflows and client question frequency, quantified effort scores, and a mapped AI opportunity space.

Phase 01

# 360° Triangulation

Focus Group | Survey Design | Survey at Scale | Platform Analysis | AI Opportunity Map

## Focus Group

*Where it starts: focus group informs design before scale*

- **Discover unknowns**
  Surface pain points you didn't know to ask about
- **Calibrate language**
  Ensure traders interpret questions the same way
- **Test comprehension**
  Catch ambiguous or confusing questions before scaling
- **Shape dimensions**
  Define the right constructs: frequency, effort, AI readiness

## Platform Analysis

*Parallel track: platform data confirms what traders reported*

- **Clickstream data**
  Revealed navigation patterns and high-frequency workflow areas
- **Session duration**
  Showed which tasks take disproportionately long to complete
- **Re-entry patterns**
  Flagged tasks traders had to return to repeatedly — a friction signal
- **Support escalations**
  Identified highest-friction workflows where traders sought external help

THE APPROACH

# Closing the Gap

STATED

## What users say

Survey & focus group. Traders told us which workflows are the hardest and least efficient along with which clients' questions were hardest and most repetitive.

**Strength: intent & context**

**Blind spot: people misremember**

REVEALED

## What users do

Platform clickstream, session duration, re-entry patterns, support escalations. The backend showed where traders actually spent time and got stuck.

**Strength: actual behavior**

**Blind spot: no "why"**

THE APPROACH

# Key Findings & AI Opportunity Map

## Data Triangulation

### Highest session time: troubleshooting

Survey: #1 hardest question
Behavioral: longest platform sessions, most support escalations

### Most re-entries: troubleshooting tasks

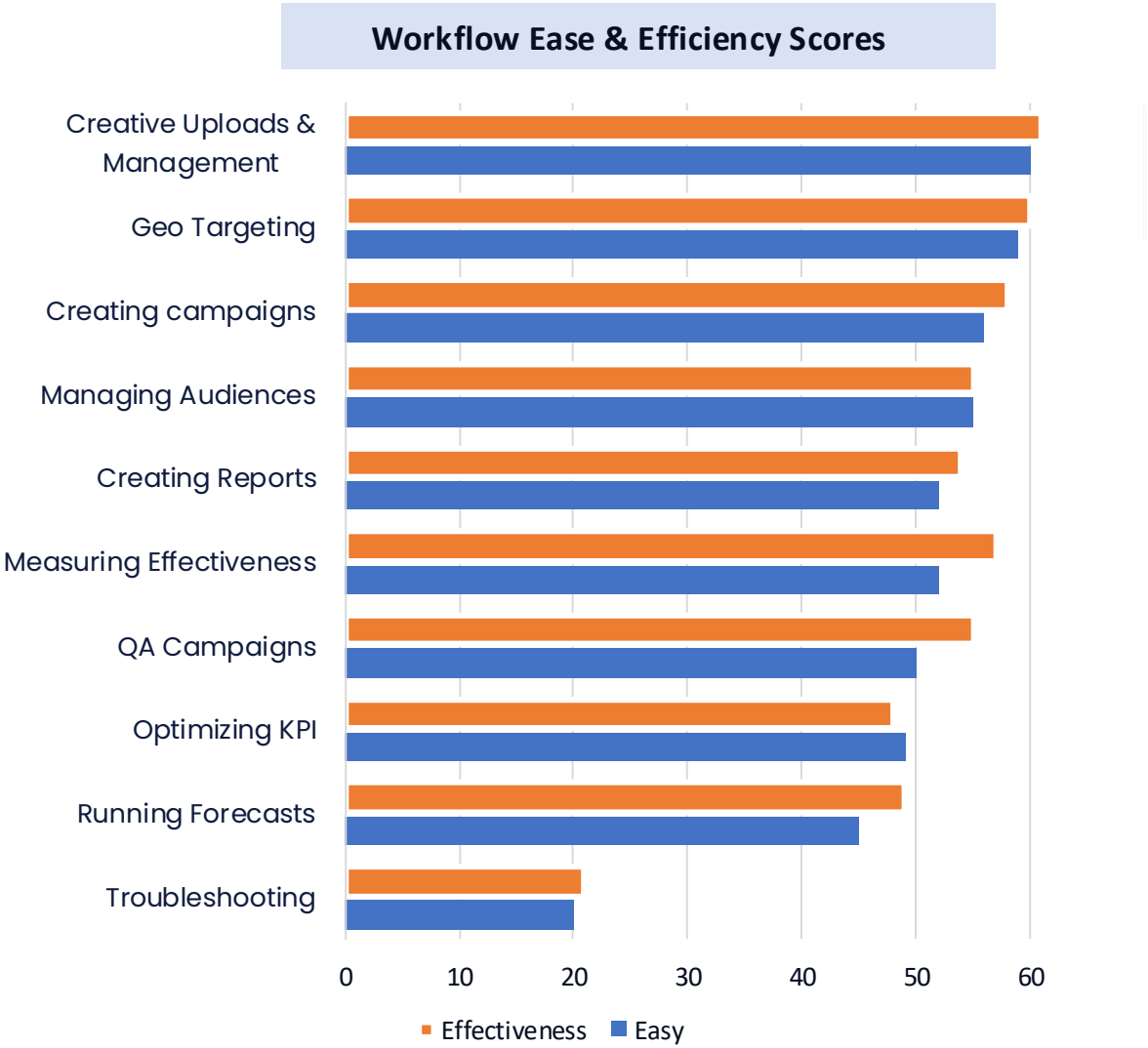
Survey: high frequency, 30+ min average
Behavioral: traders repeatedly visit report tiles before finding answer

### Noisiest clickstream: Kokai navigation

Survey: newer traders struggle most
Behavioral: random navigation paths, trial-and-error in UI

### Escalation trigger: integrations & costs

Survey: repetitive question from clients
Behavioral: issues that end ends in SME escalation or Jira ticket



# Traders were drowning in repetitive tasks and complex workflows

**80%** of traders spend 50+% of their time troubleshooting client's campaigns

**75%** of traders spend 2+ hours per week answering repetitive client questions.

*This is costing the company millions when their time could be optimized for more complex, added value tasks*

THE BET

# Three agents backed by in findings

*Both signals converged on three jobs-to-be-done — that convergence became the product scope.*

## Navigation

STATED

Platform navigation and feature questions ranked 2nd in client inbound.

REVEALED

Heavy help-doc search sessions; repeated lookups for the same concepts.

## Knowledge

STATED

Knowledge fragmented across Slack, Confluence, Jira, peer DMs.

REVEALED

Traders leave the platform to find answers. No single source works.

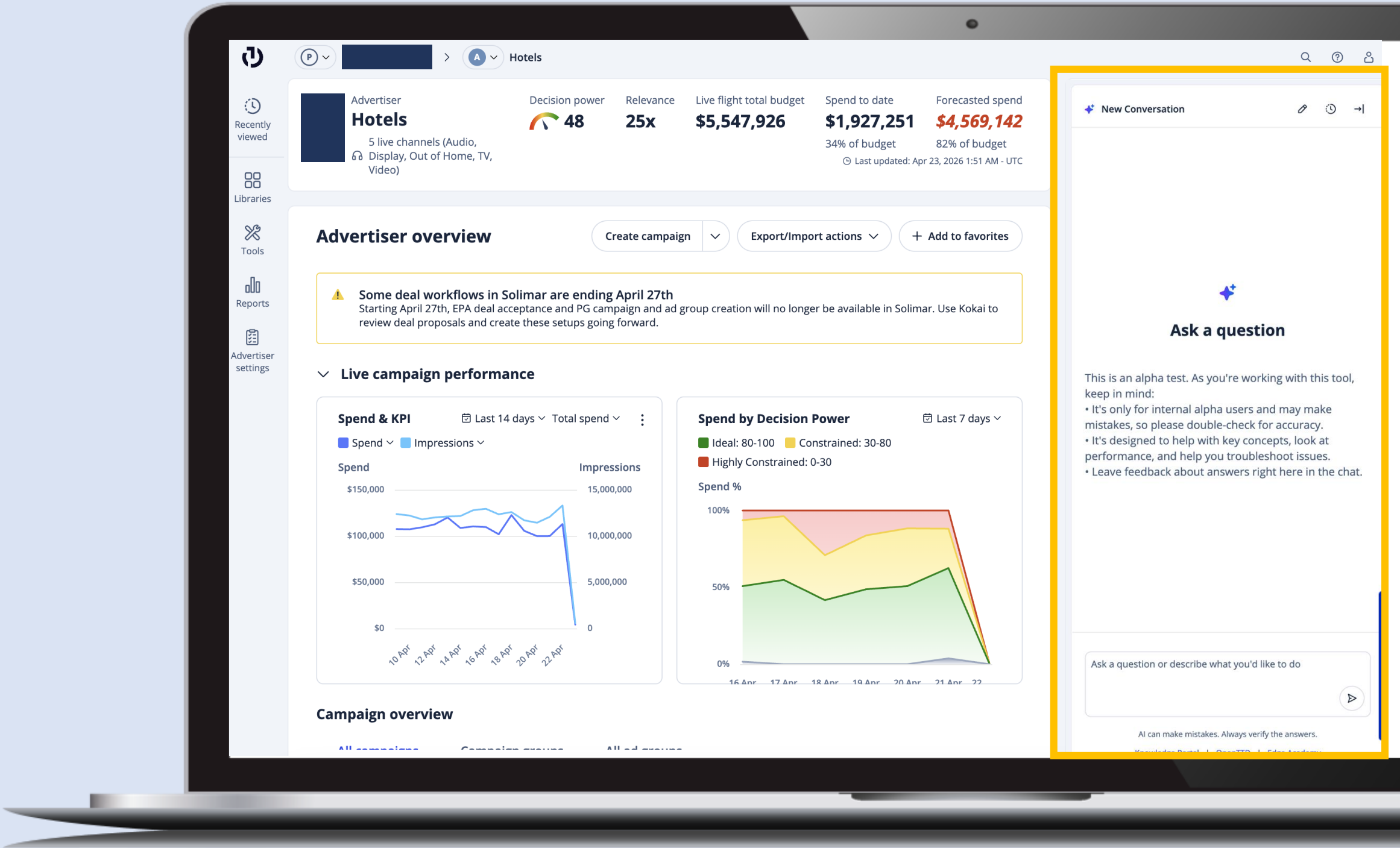
## Troubleshooting

STATED

"Why isn't my campaign spending?" — the #1 question clients ask traders.

REVEALED

Traders spent most time in campaign / ad-group diagnostic surfaces.


Hotels
Recently viewed
Libraries
Tools
Reports
Advertiser settings
Advertiser
Hotels
5 live channels (Audio, Display, Out of Home, TV, Video)
Decision power
48
Relevance
25x
Live flight total budget
$5,547,926
Spend to date
$1,927,251
34% of budget
Forecasted spend
$4,569,142
82% of budget
Last updated: Apr 23, 2026 1:51 AM - UTC
Advertiser overview
Create campaign
Export/Import actions
Add to favorites
Some deal workflows in Solimar are ending April 27th
Starting April 27th, EPA deal acceptance and PG campaign and ad group creation will no longer be available in Solimar. Use Kokai to review deal proposals and create these setups going forward.
Live campaign performance
Spend & KPI
Last 14 days
Total spend
Spend
Impressions
$150,000
$100,000
$50,000
$0
15,000,000
10,000,000
5,000,000
0
10 Apr
12 Apr
14 Apr
16 Apr
18 Apr
20 Apr
22 Apr
Spend by Decision Power
Last 7 days
Ideal: 80-100
Constrained: 30-80
Highly Constrained: 0-30
Spend %
100%
50%
0%
16 Apr
17 Apr
18 Apr
19 Apr
20 Apr
21 Apr
Campaign overview
New Conversation
Ask a question
This is an alpha test. As you're working with this tool, keep in mind:
• It's only for internal alpha users and may make mistakes, so please double-check for accuracy.
• It's designed to help with key concepts, look at performance, and help you troubleshoot issues.
• Leave feedback about answers right here in the chat.
Ask a question or describe what you'd like to do
AI can make mistakes. Always verify the answers.

SOLUTIONS

# Troubleshooting is incredibly complex

## Troubleshooting

### What is it

Troubleshooting is the process of diagnosing and resolving issues that prevent campaigns from spending, winning bids, or reaching their goals

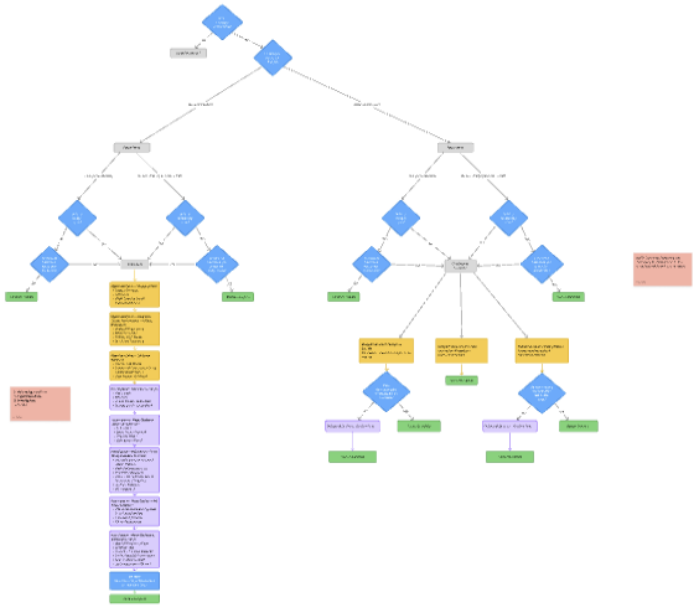
### WHAT IT ENTAILS

1. Traders follow a mental decision tree to find and isolate the root cause.
2. Begin by verifying campaign/ad group status, spend trends, and how budgets are set up.
3. Inspect ad group bids, targeting, creative approval/status, format compatibility with supply inventory, audience reach, and deal eligibility
4. Compare settings with inventory and campaign requirements to find mismatches or blockers.
5. Surface the most critical spend-blocking reasons, then take direct action—adjust targeting, budget caps, creative formats, or deal parameters as needed.

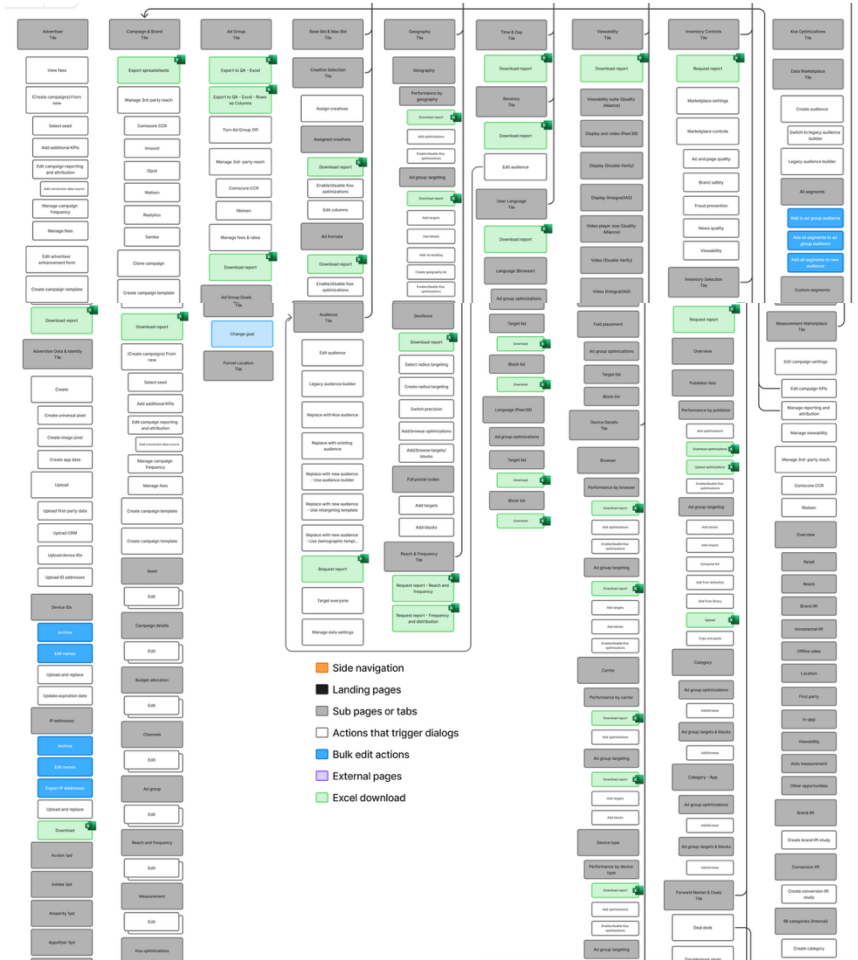
SOLUTIONS

# Troubleshooting is the most complex

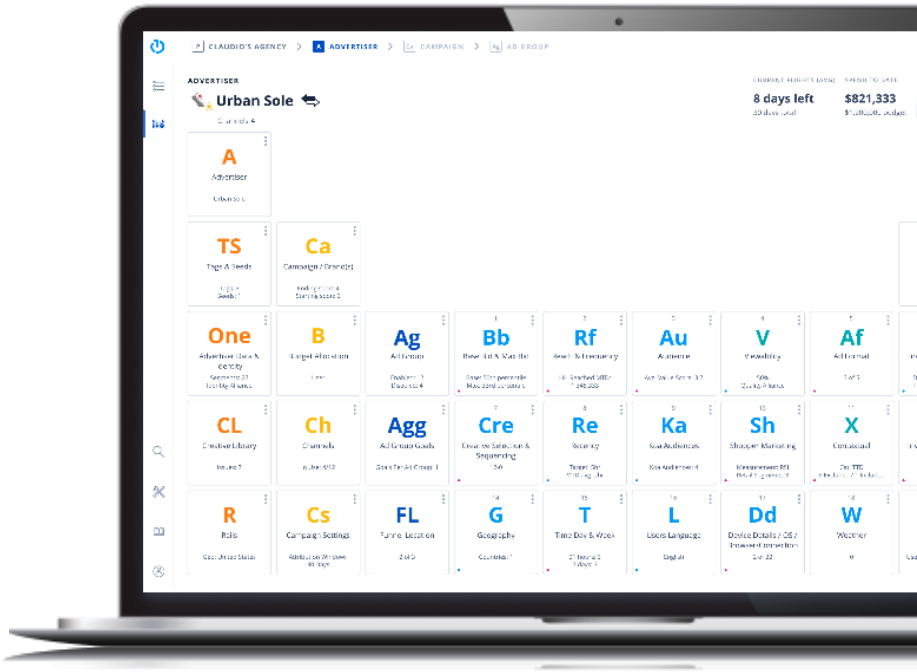
## Decisions



## Sitemap Navigation

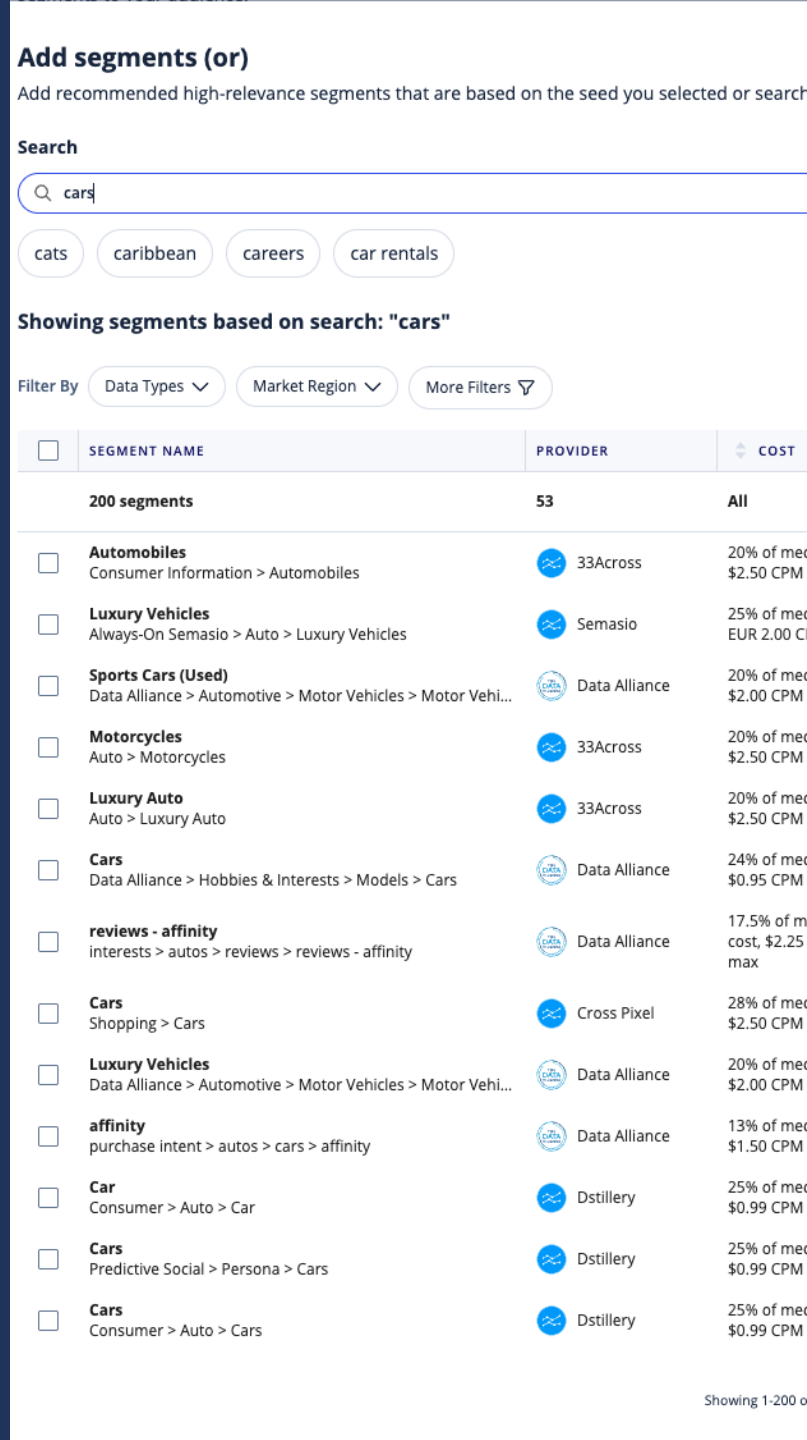



## Fragmented UI



PHASE 02

# What is a good answer ?



PHASE 02 · DEFINITION

# Before evaluating quality, we had to define it.

THE APPROACH

# Quantifying “goodness”

**Goal:** Quantify user preferences for AI chatbot response and define an “ideal answer profile.”

## Challenge

While “good responses” are often discussed qualitatively (e.g., helpful, concise, actionable) but:

- The agent needs parameters to model the output
- We have not measure of what dimensions contribute to a good answer
- How each dimension influences overall preference

## Solution

We used a choice-based conjoint survey:

- Captures real-world trade-offs (not isolated ratings)
- Estimates relative importance of attributes
- Enables decomposition of “ideal response” into measurable components
- Quicker and cheaper than A/B testing
- Directly translates into AI evaluation

PHASE 02 · DISCOVERY

# Conjoint Design: Attributes and Levels

500 participants evaluated 10 pairwise choice tasks. Each task showed two AI answers differing across 6 attributes with varying levels.

## Actionable

**Level 1:** Generic guidance

**Level 2:** Specific next steps

**Level 3:** Redirecting

*Preferred →* ***Specific next steps***

## Personalized

**Level 1:** One-size-fits-all

**Level 2:** Tailored to advertiser

*Preferred →* ***Tailored to context***

## Concise

**Level 1:** Extensive (400+ words)

**Level 2:** Mid (150 to 400 words)

**Level 3:** Brief (≤150 words)

*Preferred →* ***Brief (≤150 words)***

## Sourced

**Level 1:** Expand citations

**Level 2:** Inline citations

**Level 3:** No citations

*Preferred →* ***Inline citations***

## Format

**Level 1:** Plain prose

**Level 2:** Headers

*Preferred →* ***Headers + bullets***

## Follow-up

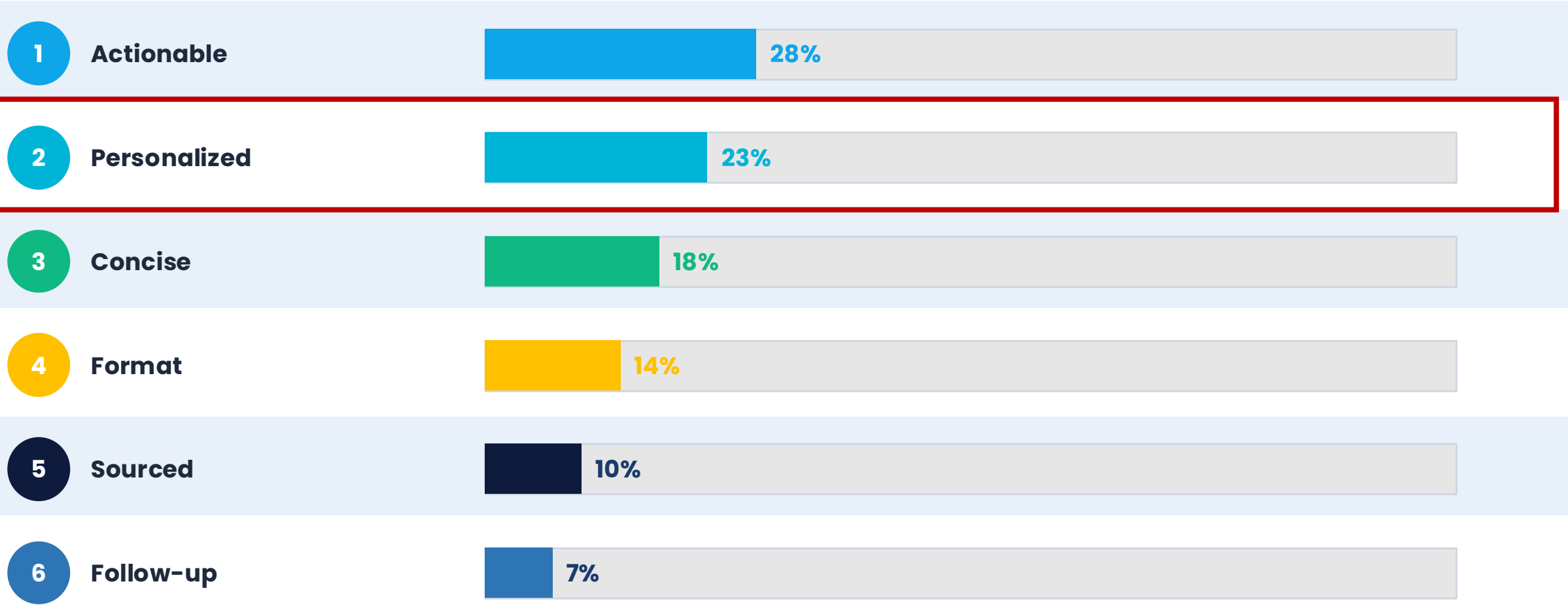
**Level 1:** No follow-up

**Level 2:** Follow-up suggestions

*Preferred →* ***No follow-up***

# Relative Attribute Importance

% of total utility explained by each attribute (N=500)

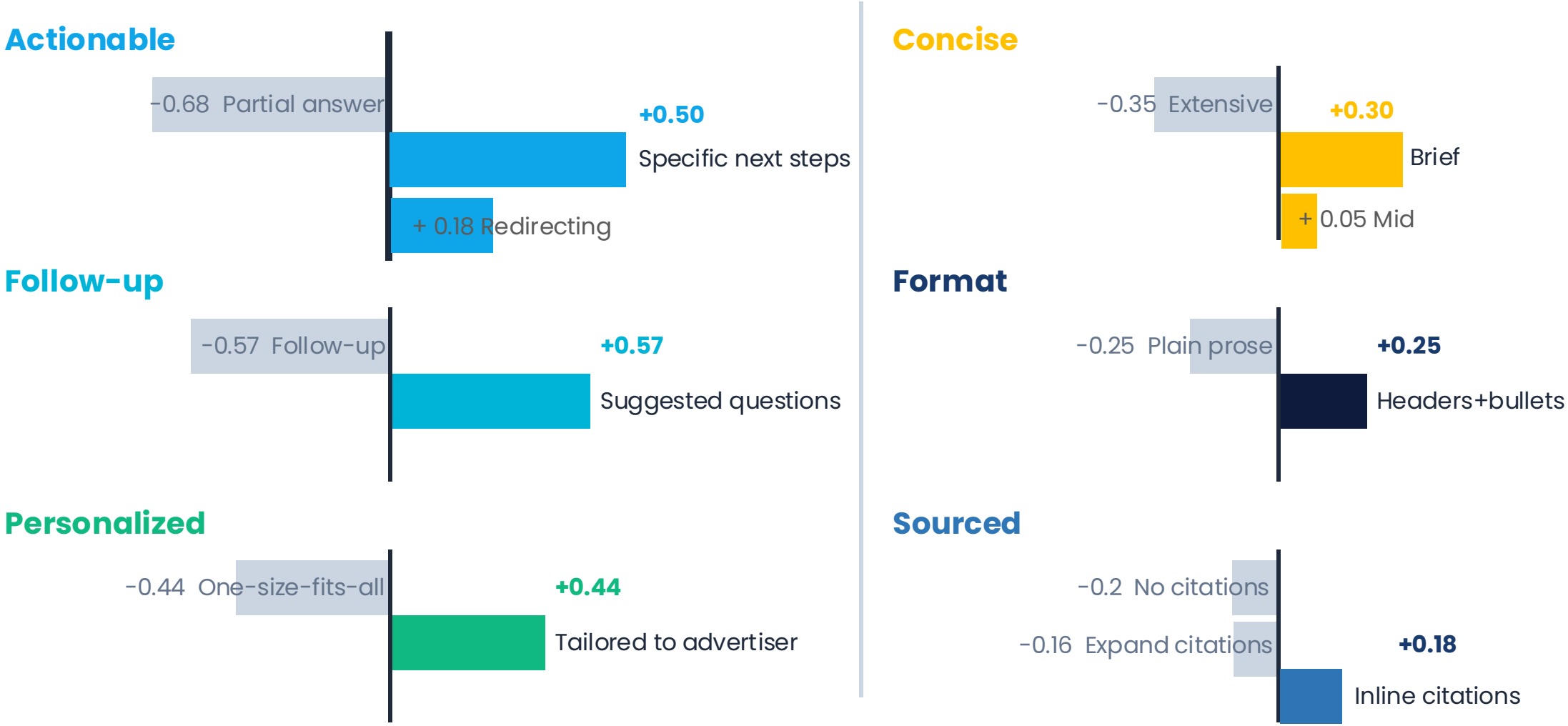



Key Insight: Personalization and Actionability together account for 51% of total preference

# Preferrable components

Part-Worth Utilities by Attribute Level

**Higher utility = stronger preference | Utilities centered at 0 within each attribute**



PHASE 02 · DISCOVERY

# Traits of a “good” answer

DOMINANT THEME

**Resolution & Guidance**

A good answer is not one that is “factually correct” but one that understand your needs and guides you to a solution.

## Actionable

Tells the user what to do next and where to find it in-platform.

## Format

Digestible and scannable information

## Concise

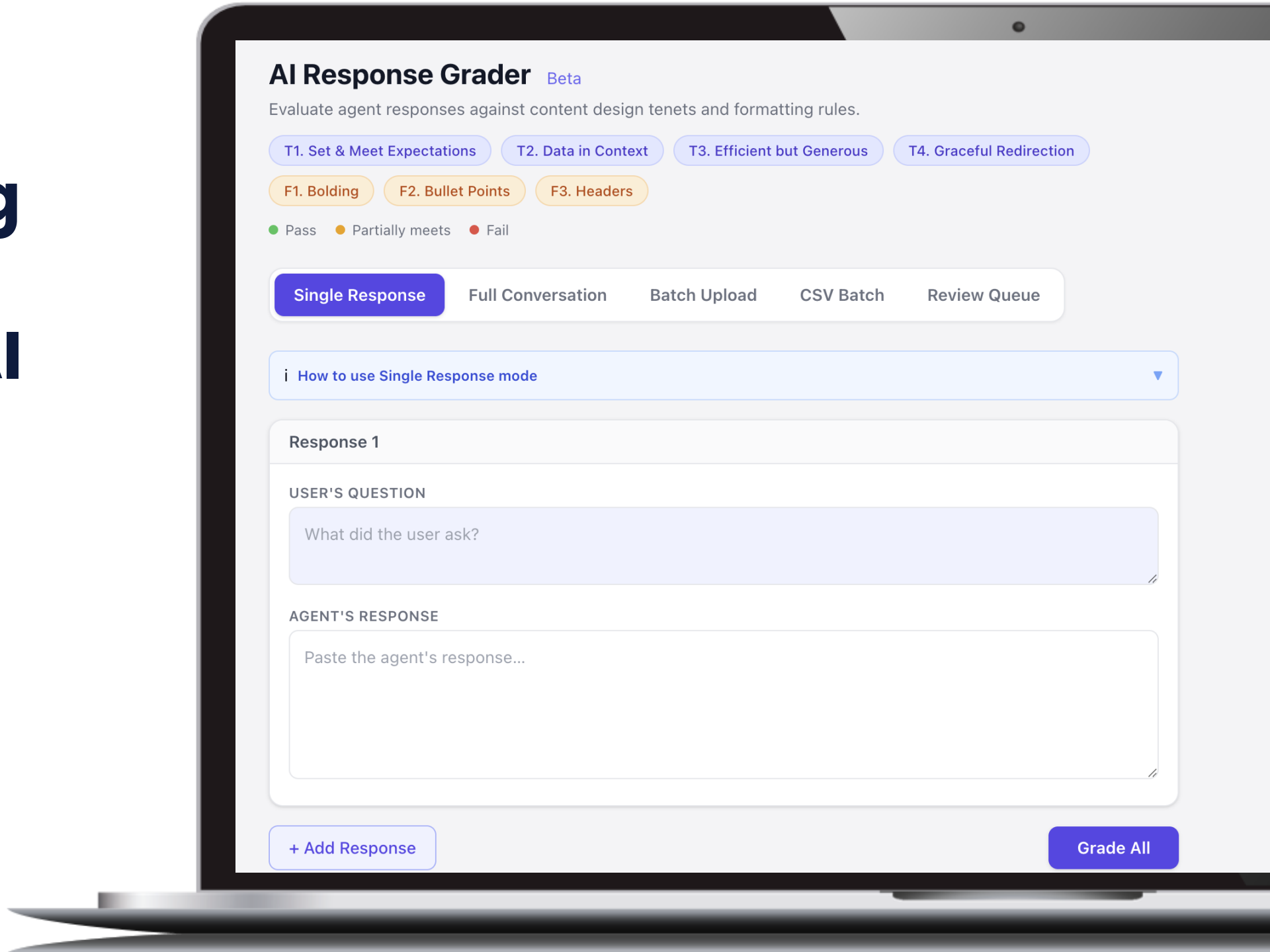
Leads with the key point; all details without AI-ish verbosity.

## Personalized

Understands the constrains of the user and their client

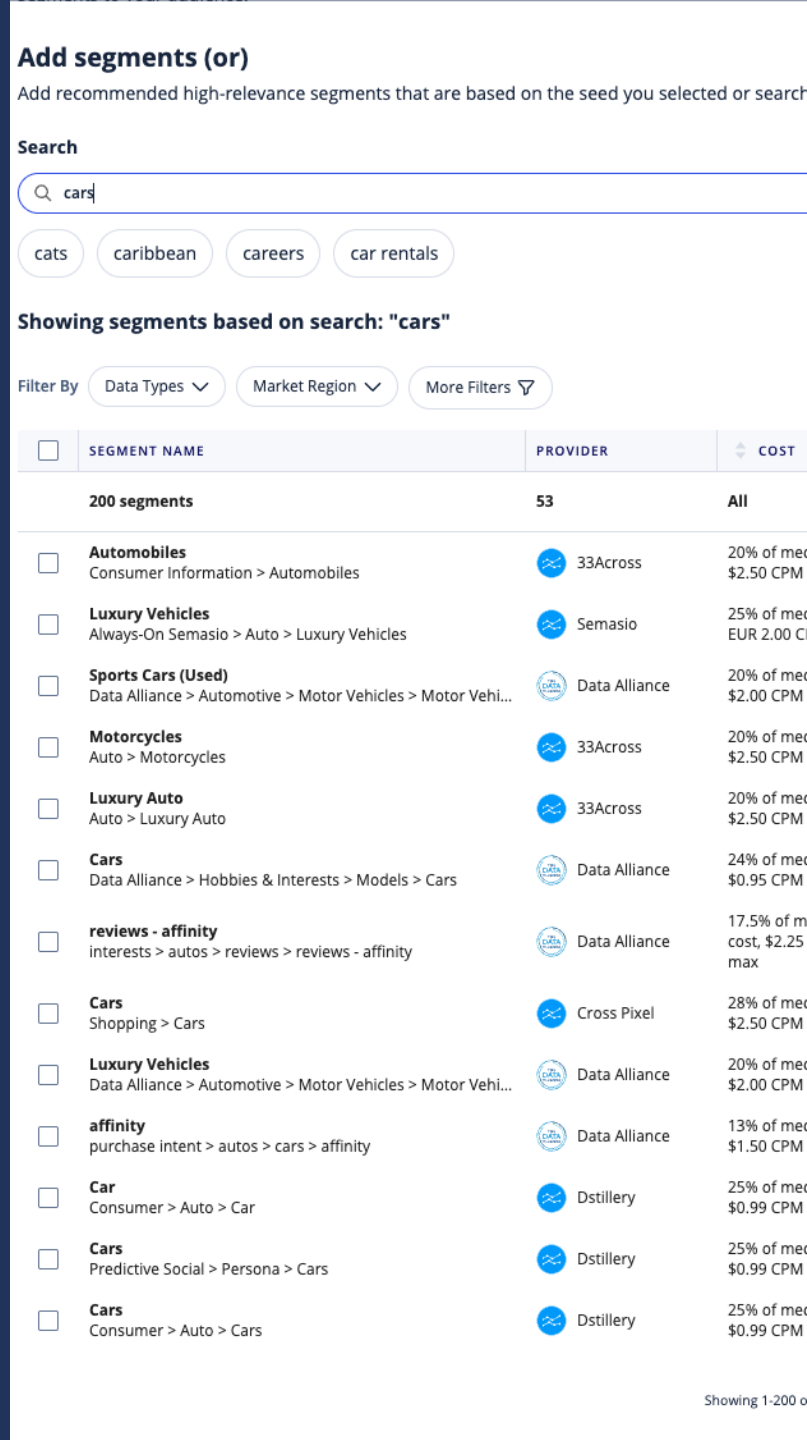
PHASE 02

# Leveraging AI to help build out AI



PHASE 03

# Refining & Systematically Evaluating



PHASE 03 · EVALUATION

# A universal framework for quality AI agents

# AI Evaluation Flow

Two parallel feedback loops that continuously improve the model

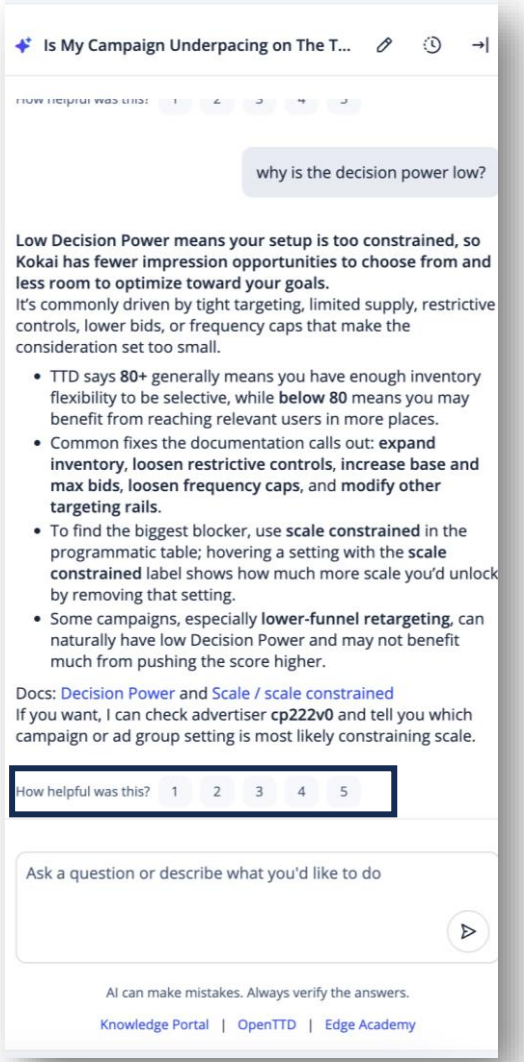
**IN-CHAT FEEDBACK**

### Prompt-Level
Rate each question 1 – 5

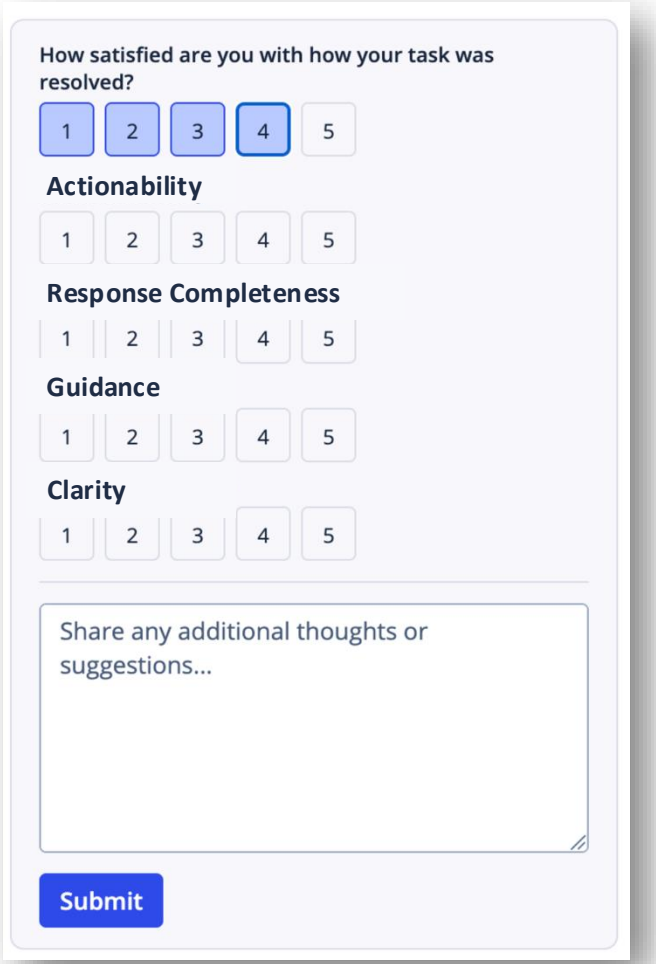
### Action Resolution Rate
Satisfaction with how AI resolved task

### Written
Open ended feedback

## Prompt – Level Rating




## Action Resolution Rate



# AI Evaluation Flow

Two parallel feedback loops that continuously improve the model

## IN-CHAT FEEDBACK

**Prompt-Level**
Rate each question 1 – 5

**Action Resolution Rate**
Satisfaction with how AI resolved task

**Written**
Open ended feedback

## DATA & METRICS

**Traces Collected**
Interaction logs stored per session

**NLP Topic Modeling**
Cluster written feedback by theme

**Scores Collected**
Distribution of Scores

**Key Improvement Areas**
Ranked by frequency & severity

## STRUCTURED EVALUATION

**Benchmark Questions**
Critical questions AI must answer

**AI Grader Tool**
Improve formatting and clarity

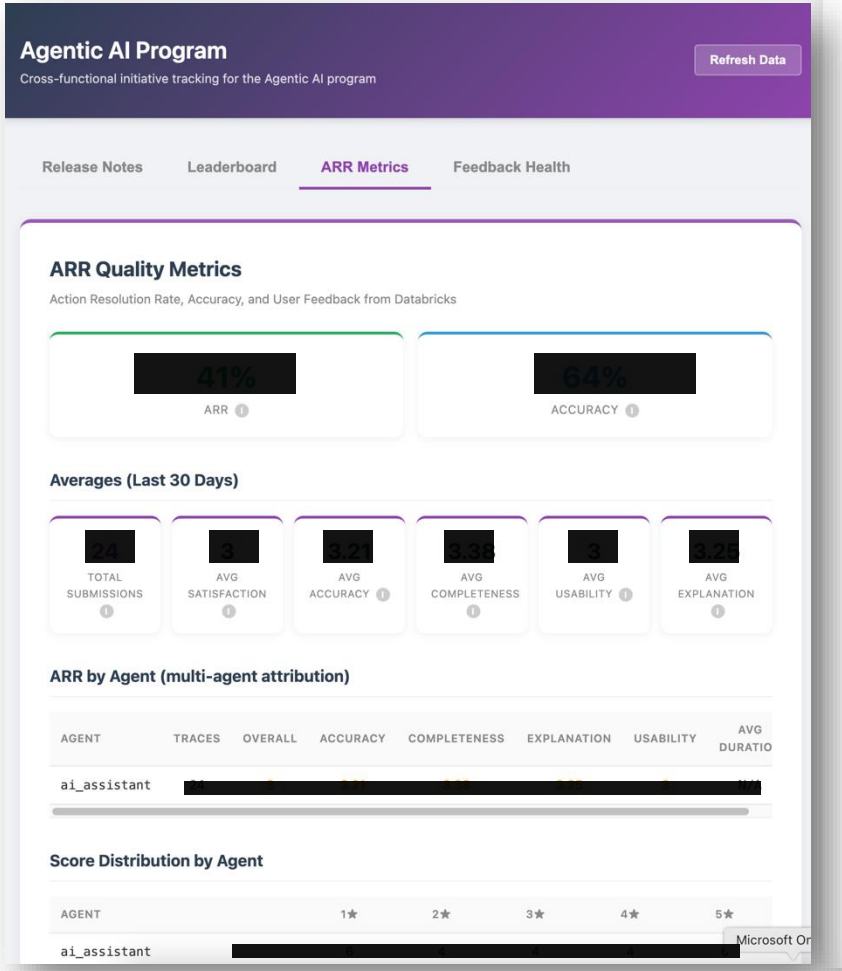
**Review Edge Cases**
Provide solutions and alternatives

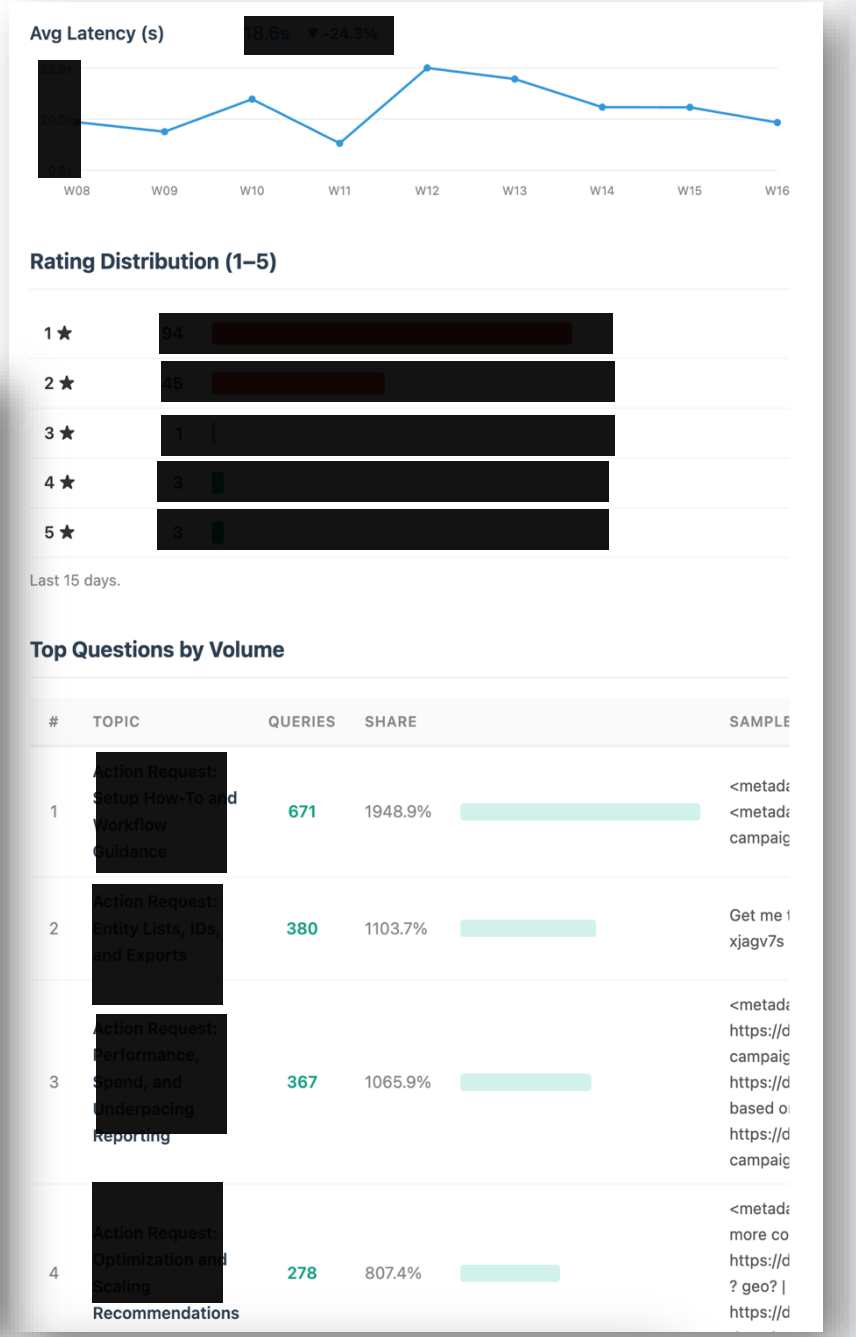
## Model Improvement

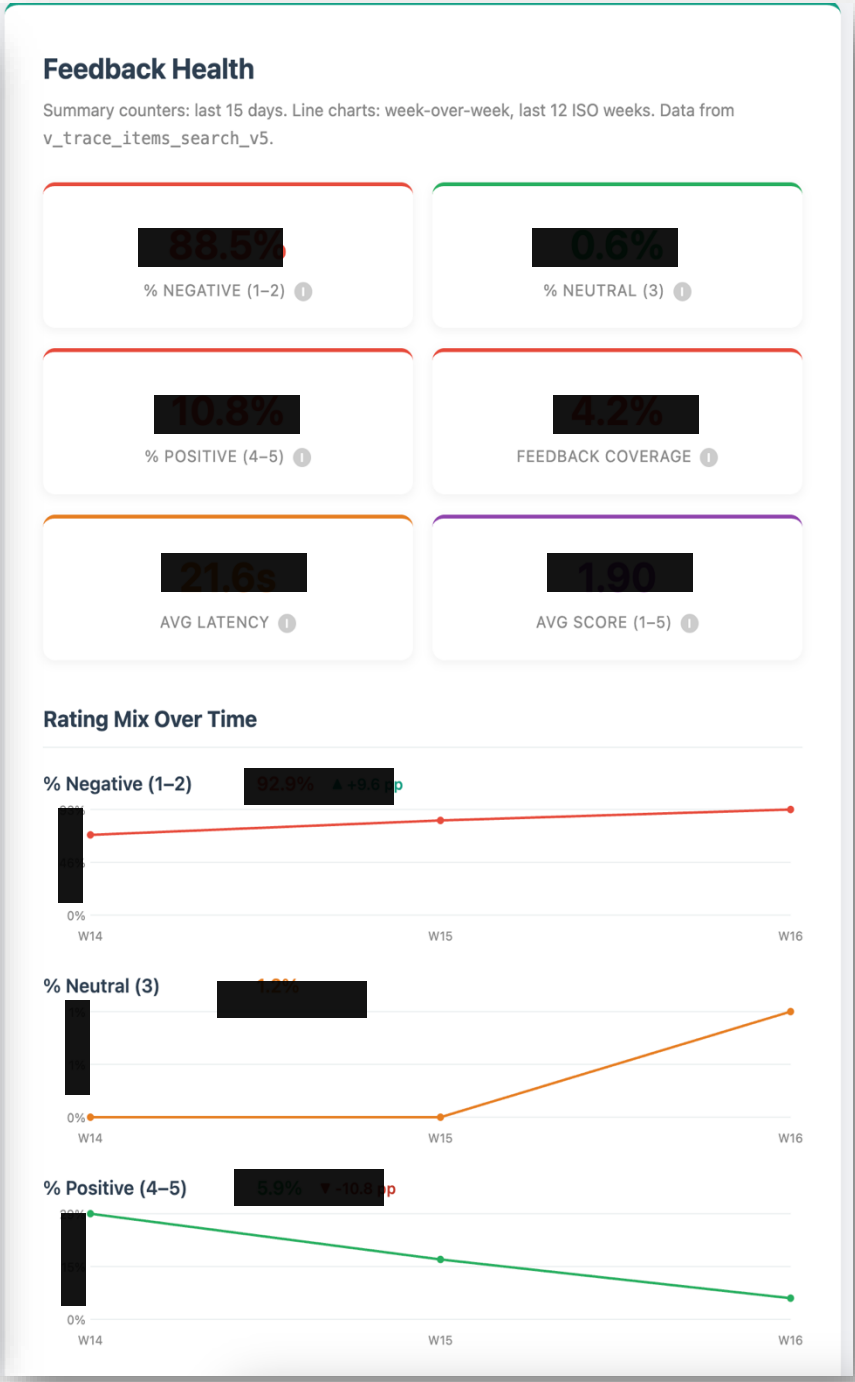
Signals from both tracks shape retraining, prompt tuning & evaluation criteria

PHASE 03 · EVALUATION

# Metrics Tracking Success

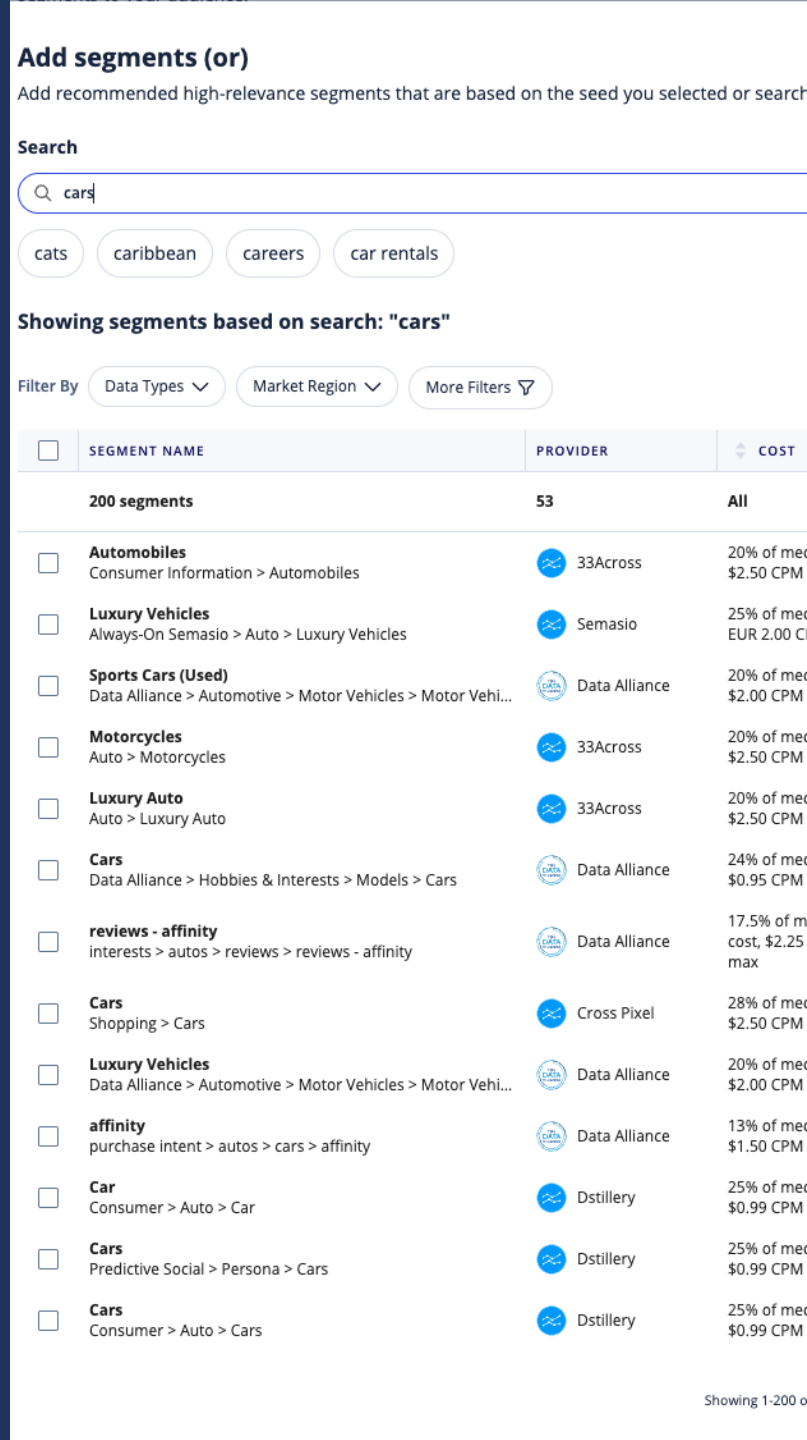


PHASE 4

# Impact measurement



PHASE 02 · DEFINITION

# Time = $

# Is AI helping traders save time?

PHASE 4

# Is AI saving traders time? Time = $

**The trap to avoid** AI adopters aren't random — they're more motivated, more tech-savvy, and likely faster at baseline. A naive “AI users vs non-AI users” comparison would overstate impact by attributing self-selection to the tool.

THE DESIGN

01

## Within-user pre/post

Each trader is their own control. Time-on-task measured before vs after their AI adoption timestamp. Removes stable differences in skill and motivation.

02

## Matched alpha testers to control group

Propensity score matching on role, tenure, baseline sessions, pre-period usage. Controls for observable differences between adopters and non-adopters.

03

## Difference-in-Differences

Did adopters improve more than matched non-adopters over the same window? Isolates AI effect from company-wide productivity trends.

PHASE 4

# Metrics Tracked

### Time to Resolve

Minutes from login to resolution action (i.e. change in campaign setting)

### Steps to Root Cause

Number of diagnostic checks before isolating the issue

### Time to Campaign Spend

Days under pacing campaigns took to spend in full again after troubleshooting

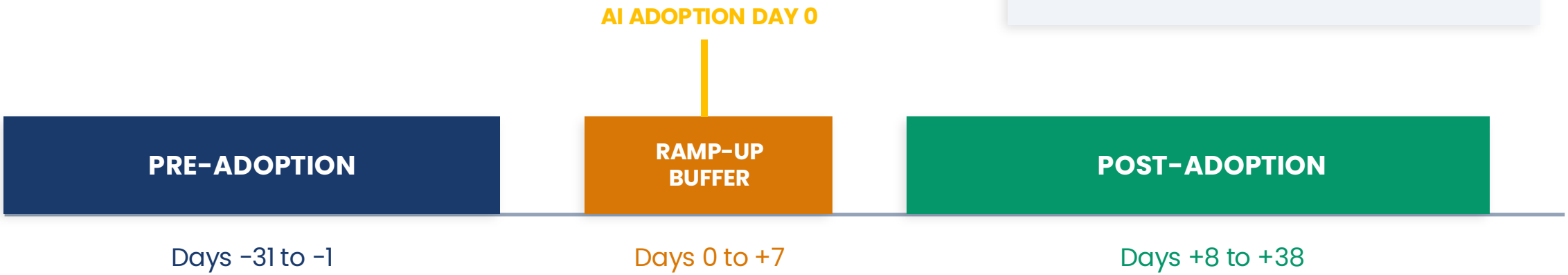
PHASE 4

# Within User Pre/Post

**~900 users**

**-12min**

On average less time to resolve per session (within-user, post vs. pre)




### Symmetric Windows

Equal-length pre and post periods eliminate exposure bias

### Buffer Period

Exclude 7-day ramp-up to avoid underestimating AI impact

### Sufficient Activity

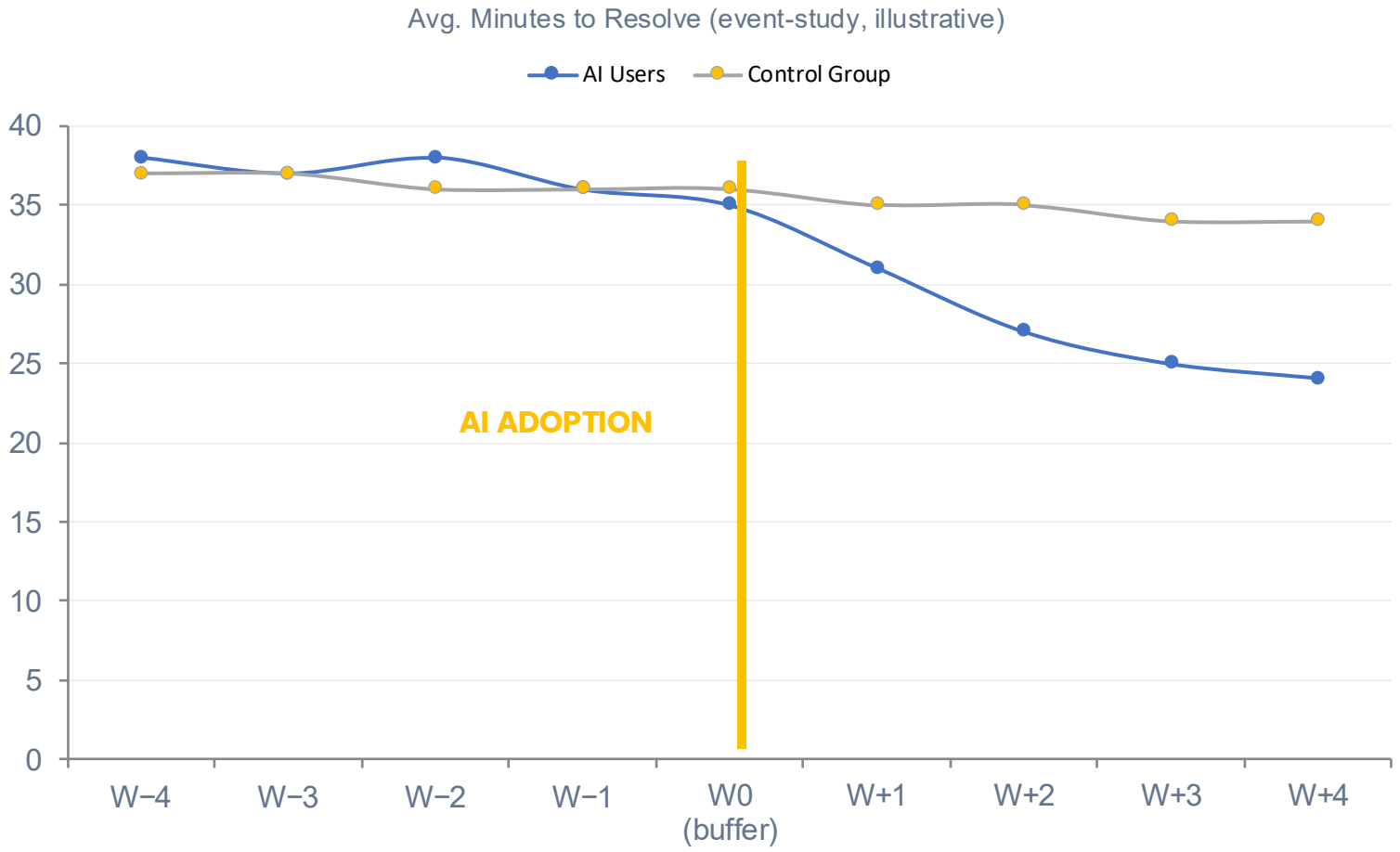
Minimum 15 tasks per window per user for stable estimates

### Full-Week Cycles

Windows span complete weeks to control for weekday patterns

PHASE 4

# Trends Before & After AI Adoption




**✓ No pre-trend**

AI users and controls move in parallel before adoption — validates the parallel trends assumption

**✓ Clear ramp-up**

W0–W+1 shows partial improvement, consistent with a 7-day learning buffer design

**✓ Sustained effect**

Gains persist through W+4, ruling out a novelty-driven "Hawthorne" spike

Note: W0 represents the 7-day buffer window (excluded from primary analysis). The divergence between AI users and control widens consistently after W+1, confirming the AI tool is driving the effect.

PHASE 4

# Difference-in-Differences: Isolating the AI Effect

Within-user analysis alone can't rule out: everyone getting faster over time, easier tasks appearing later, or system-wide improvements.

| | BEFORE | AFTER | CHANGE |
|---|---|---|---|
| **AI Users** | A | B | **B – A** |
| **Control** | C | D | **D – C** |
| **DiD Estimate** | | | **(B-A) - (D-C)** |

**DiD removes:**

- General learning / skill gains over time
- Macroeconomic or market shifts
- Platform or product improvements
- Seasonal demand patterns

- Within-user says: *"You got faster."*
- DiD says *"You got faster than everyone else did — even accounting for the tide that lifted all boats."*

**Combined approach: within-user pre/post (directional signal) + DiD with matched control (causal estimate)**

PHASE 4

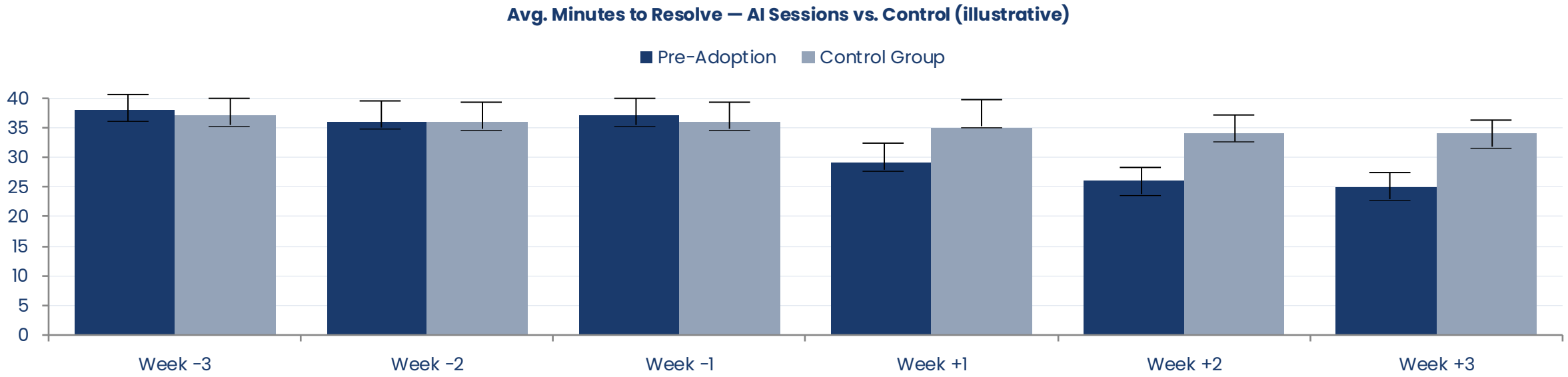
# Findings: Time to Resolve & Steps to Root Cause

**32%**

Less steps to root cause treatment vs. control group trend

**44%**

Time reduction to resolve vs. control group trend



PHASE 04 · ITERATION

# Where we are: early alpha, feedback loop closed

CURRENT STAGE

## Closed Beta ~2000 users

- Feedback coverage ~20%
- Bi-weekly trader “bug bash” quality sessions
- Dashboard instrumented for ARR, feedback coverage, latency, follow-up rate
- Time-saved research comparing AI vs baseline troubleshooting workflows

HOW LEARNING COMPOUNDS

1. Trader asks real question in tool
2. Feedback captured inline, trace logged
3. Failure classified (scoping · actionability · jargon · link)
4. Prioritized by volume × severity
5. Fix shipped · golden set re-tested

PHASE 04 · ITERATION

# Next steps...

Beta

## External clients

Agentic

## Automation

Notifications

## Post Chat Experience

# 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

**Behavirol 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**

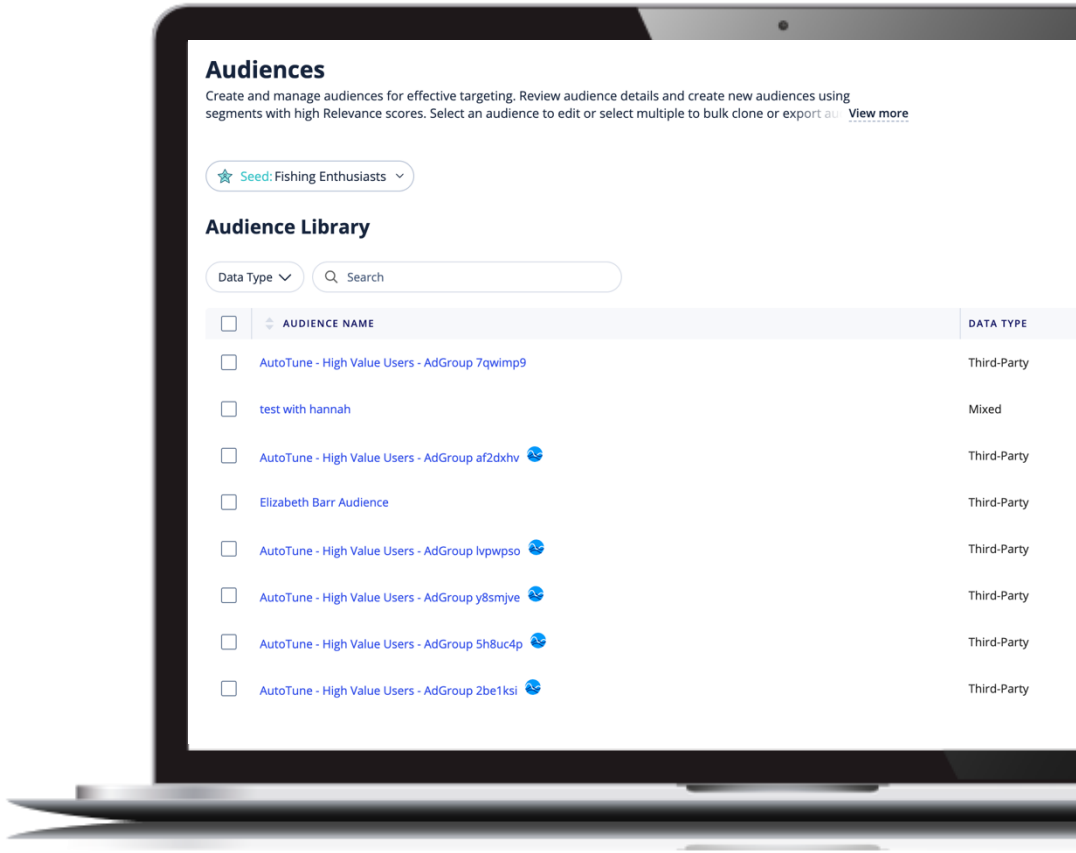
Run a randomized A/B experiment in Audience Builder

**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

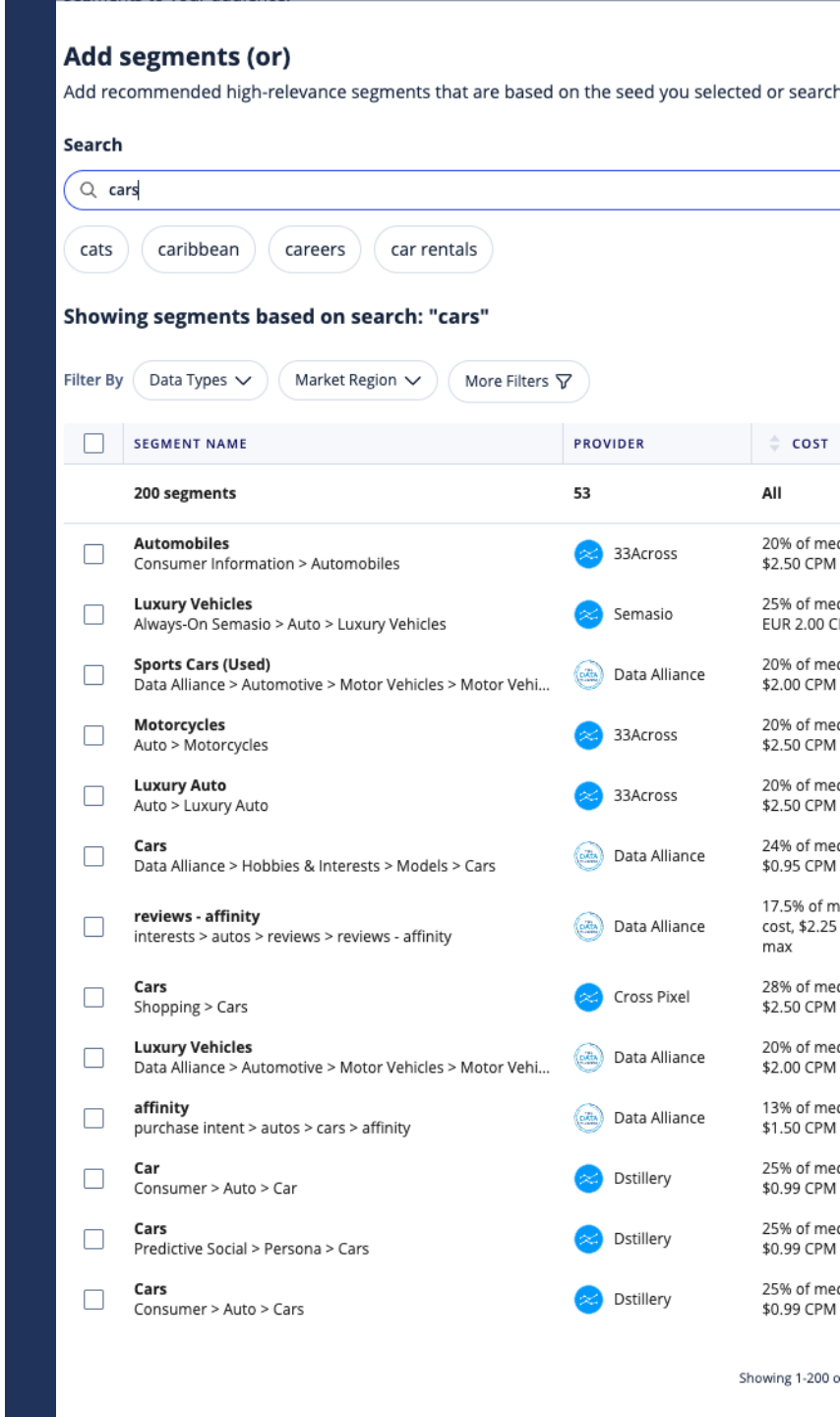
To To behaviorally test position bias we ran a randomized A/B experiment in Audience Builder:

| 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

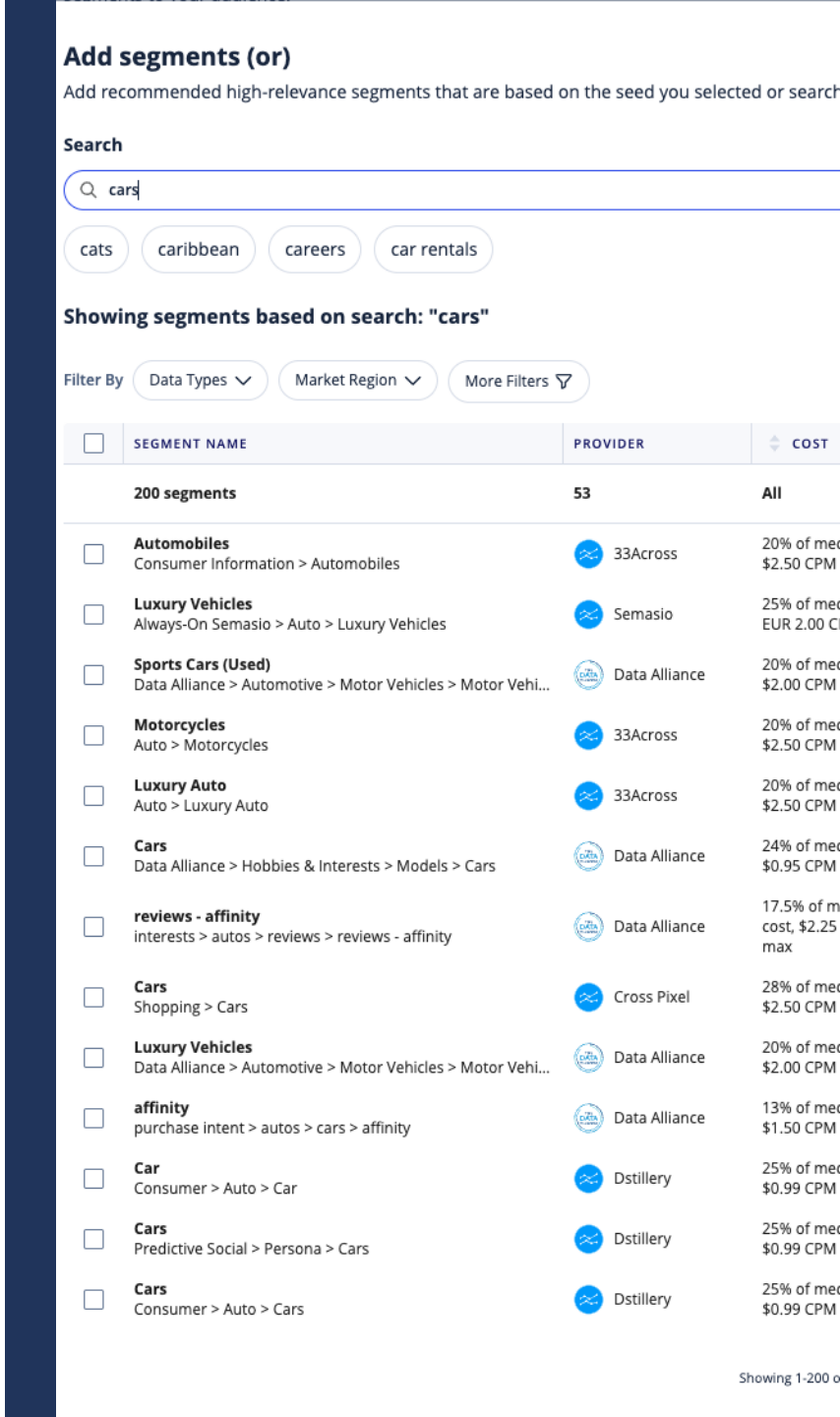
- **Randomization**: Search Query Event Level
- **Ranked by Relevance Score** : We don't change the segment pool, only the order they are displayed.

| 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

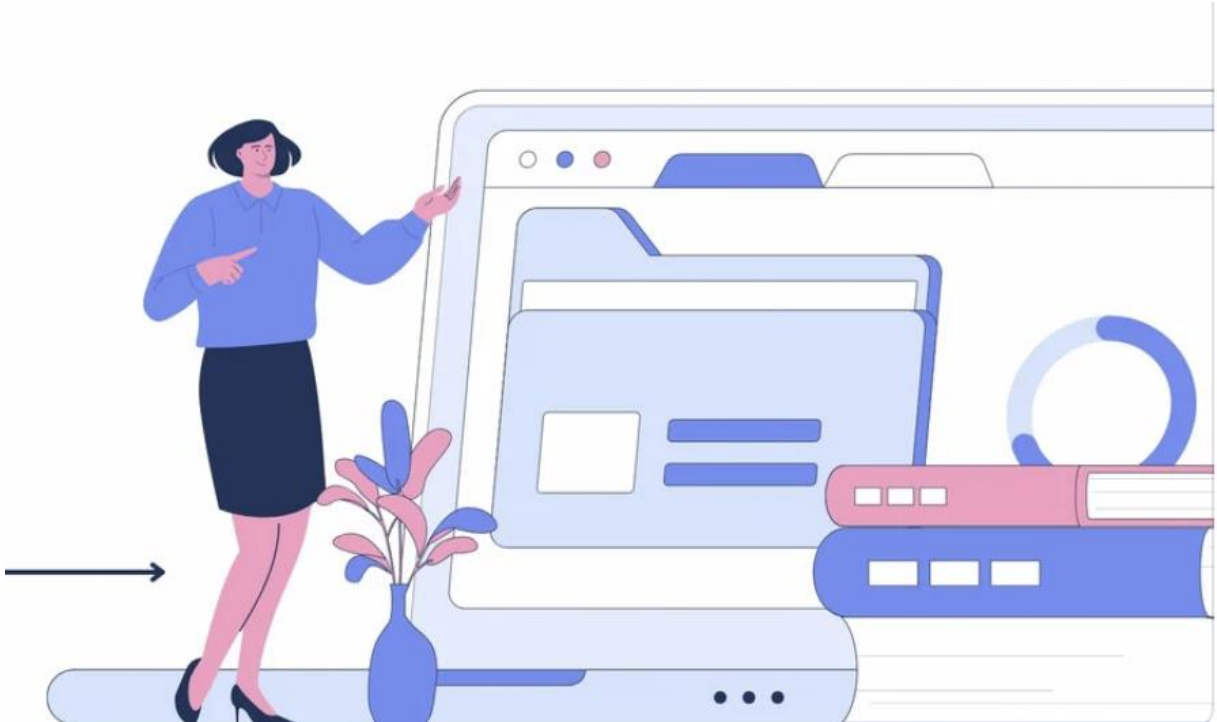
Position bias hypothesis proven

# 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

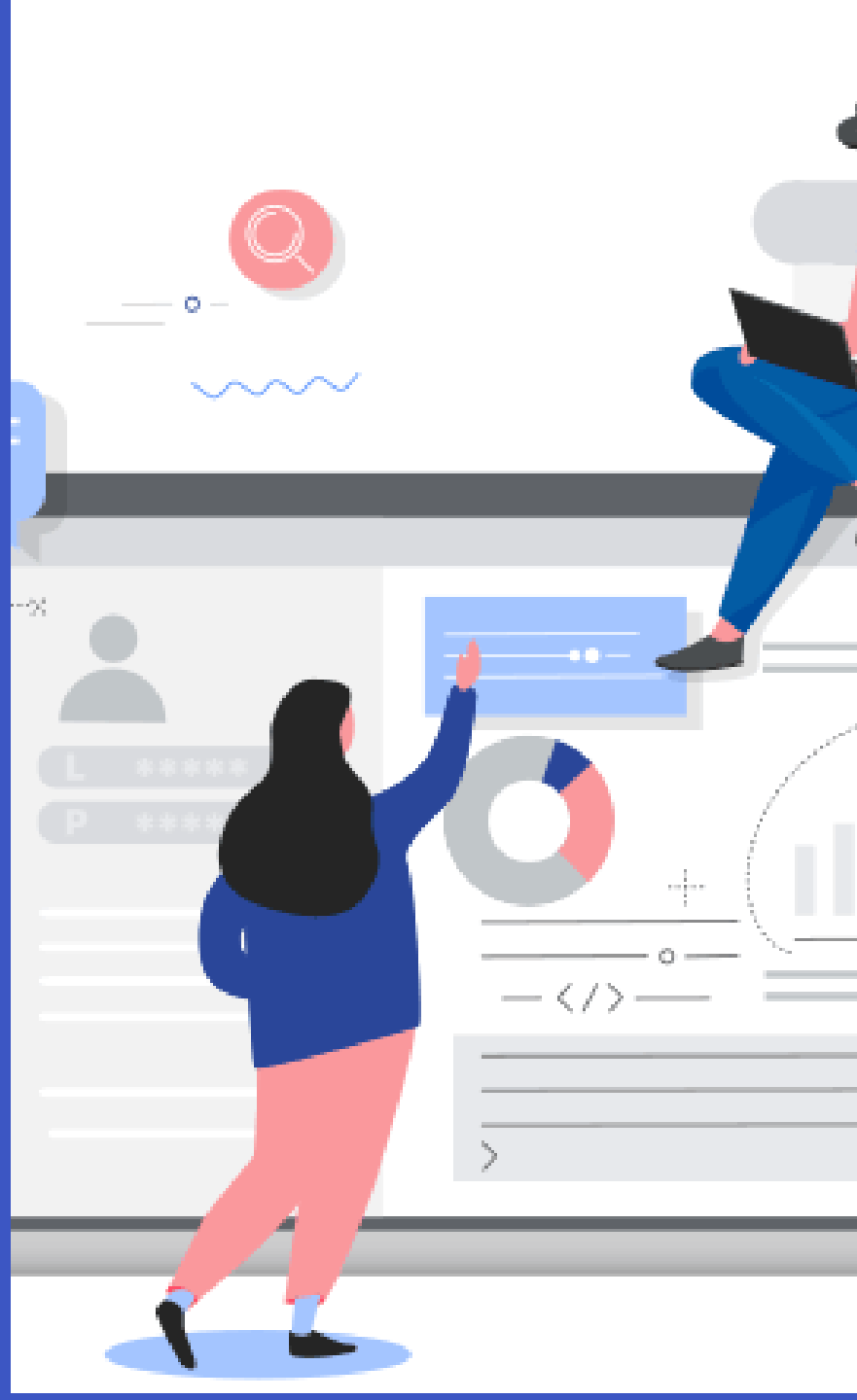


*“If you can’t measure it, you can’t improve it.”*

*– 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?

**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**

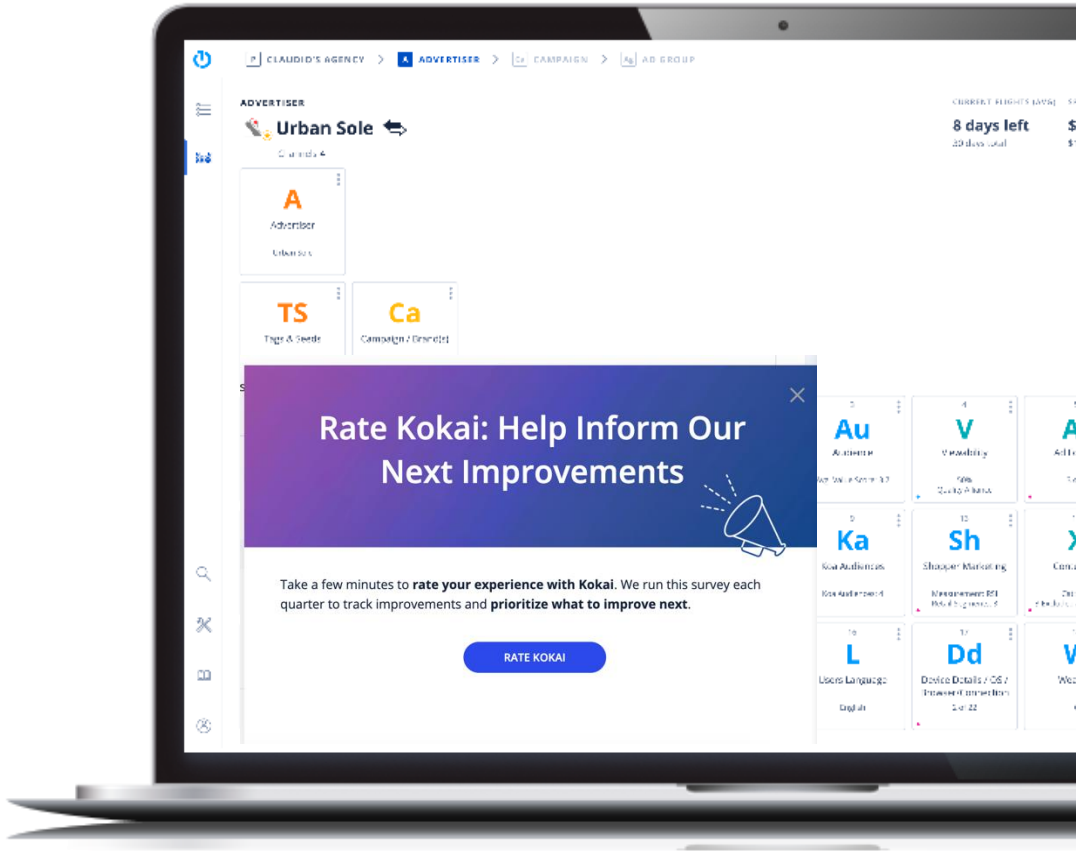
The ethical and responsible practices of a company such as transparency, integrity, fairness, and adherence to ethical standards.

**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

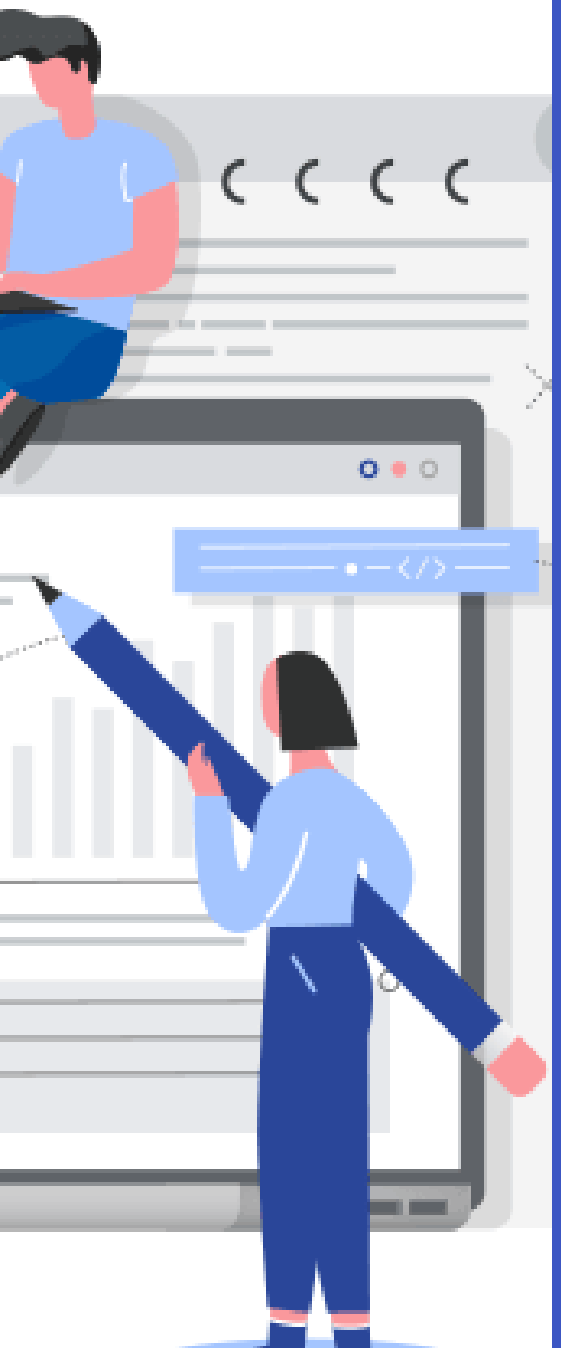
- Data cleaning
- Distributions
- Poststratification weights

## 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

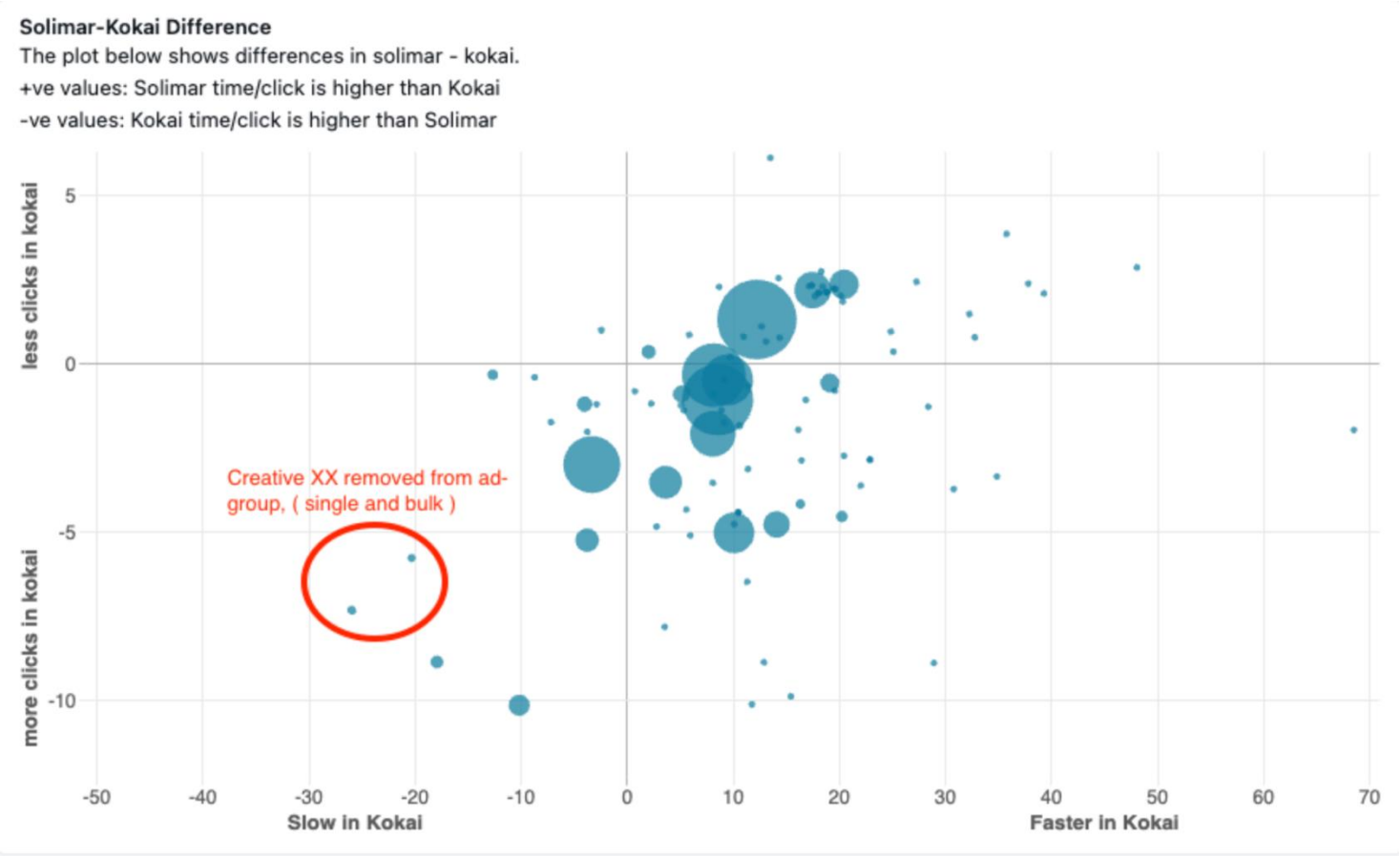
**22%**
of sessions that contain Campaign Overview also contain Campaign Settings


**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<br>Rating of the individual factors comprising Trust. | | How to Improve<br>(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. |

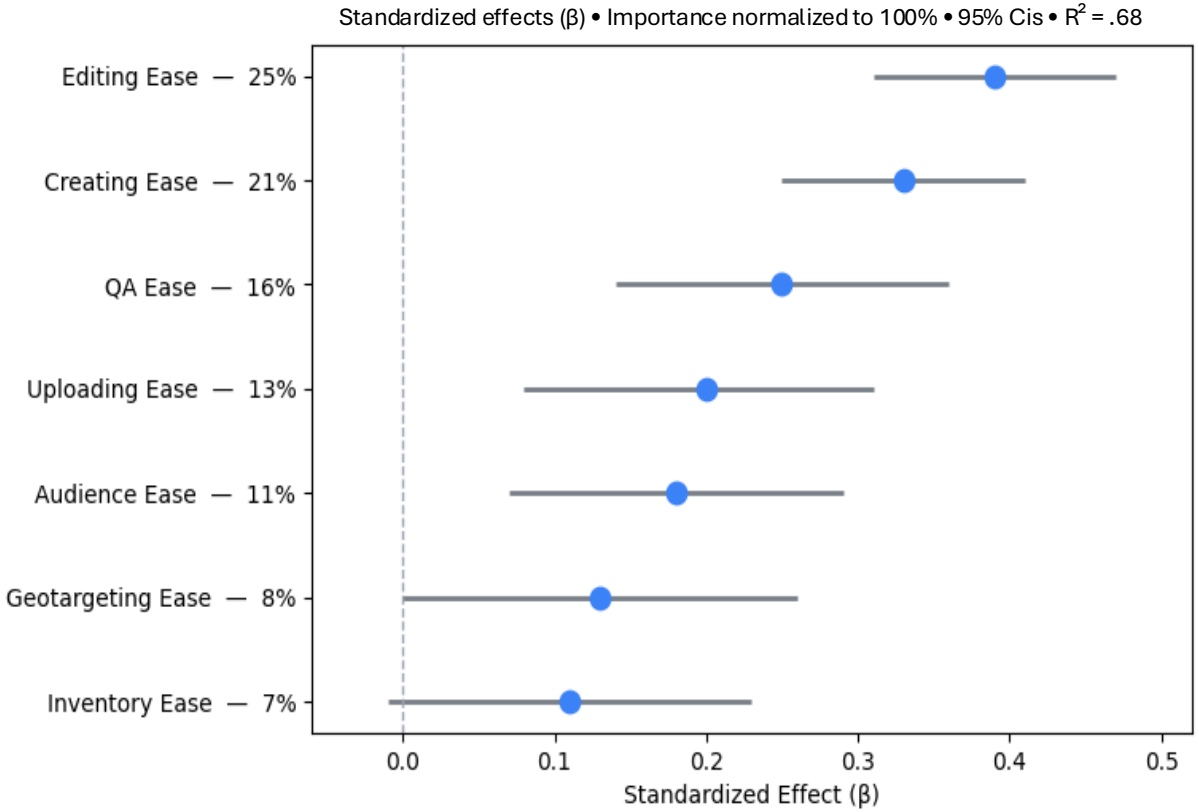
**Trust Index**

**74.8**

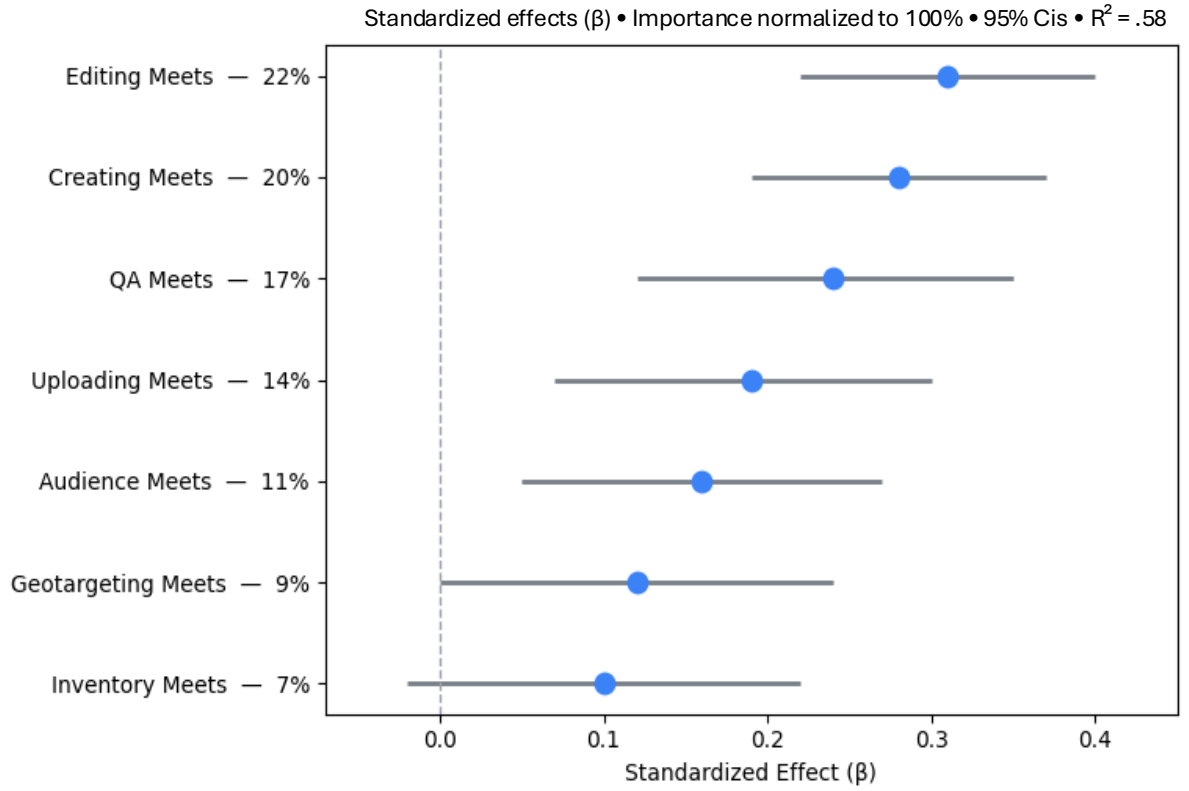
Out of 100

# Driver Maps

## What Drives Platform Ease?

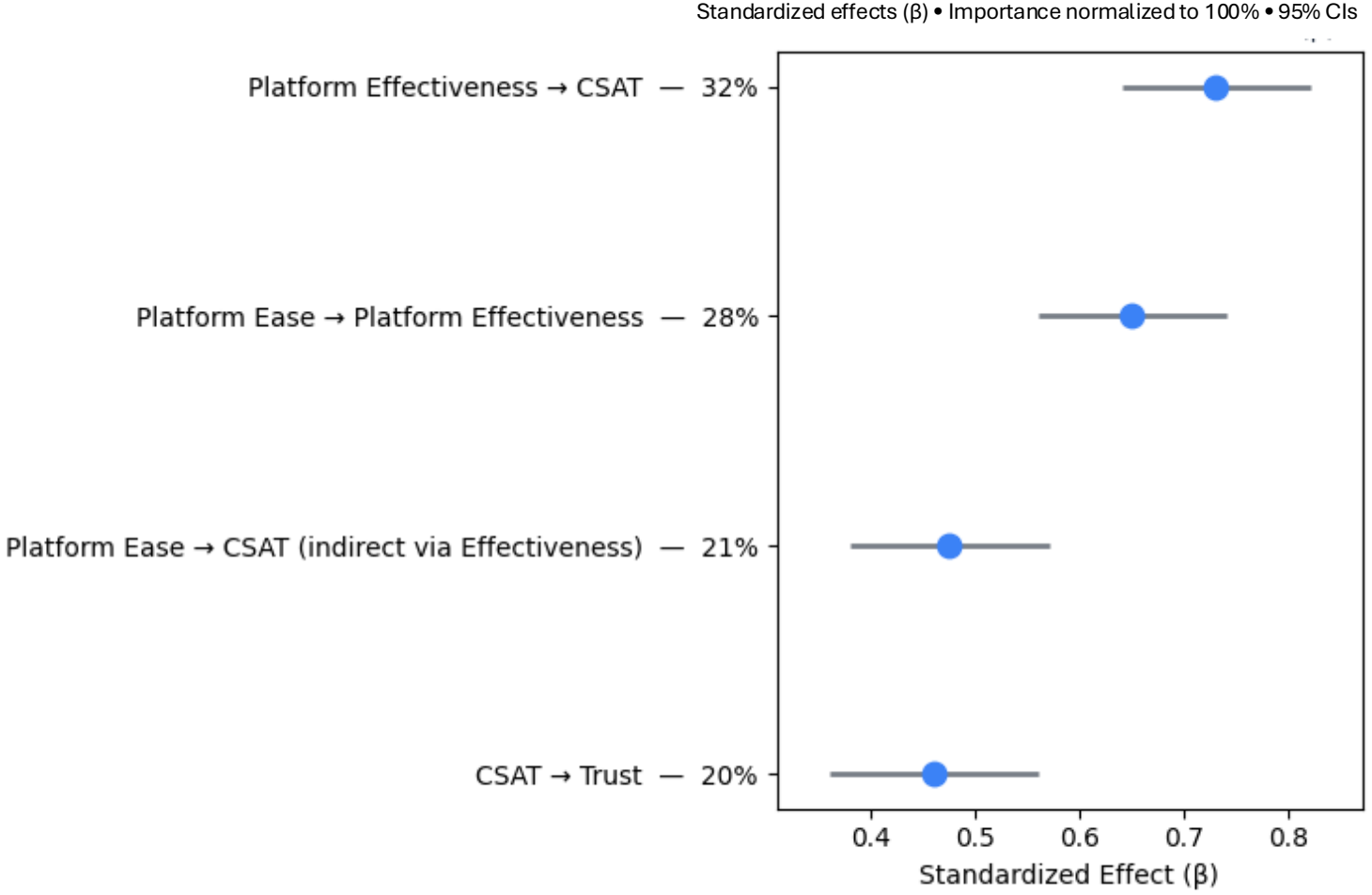



## What Drives Platform Effectiveness?

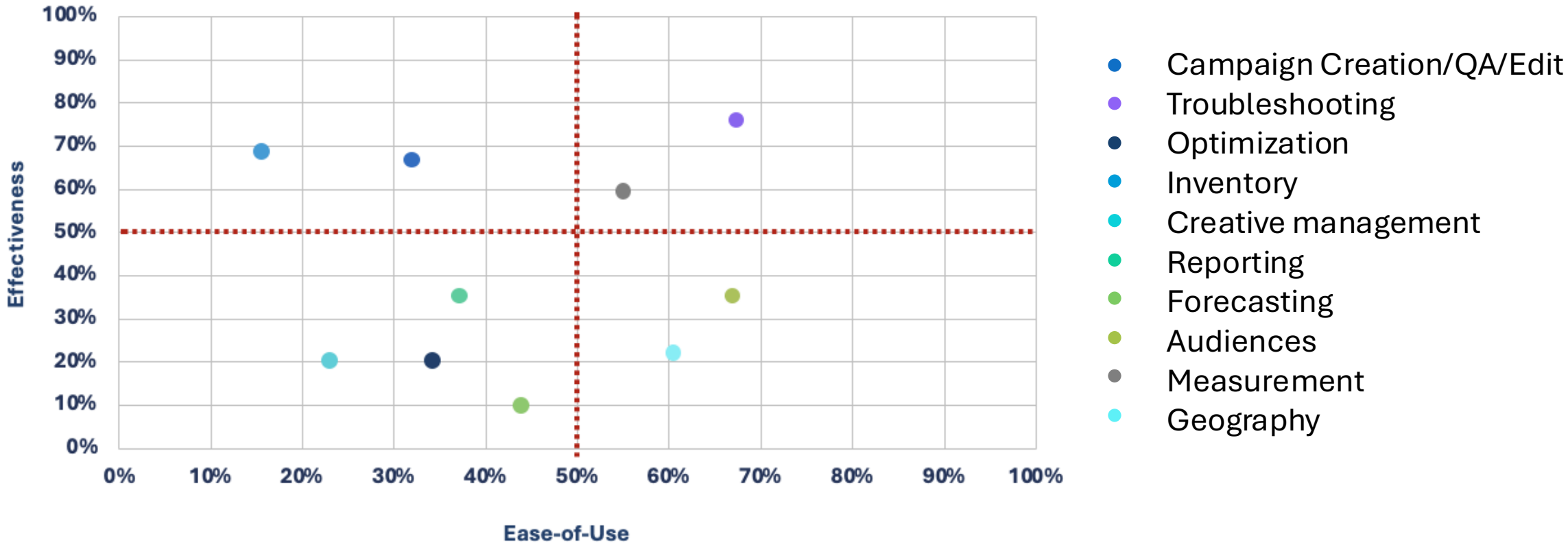


# Driver Maps

## Platform Chain:
## Ease – Effectiveness- CSAT- Trust



# Product Area Comparisons



# Hold Co vs Indie

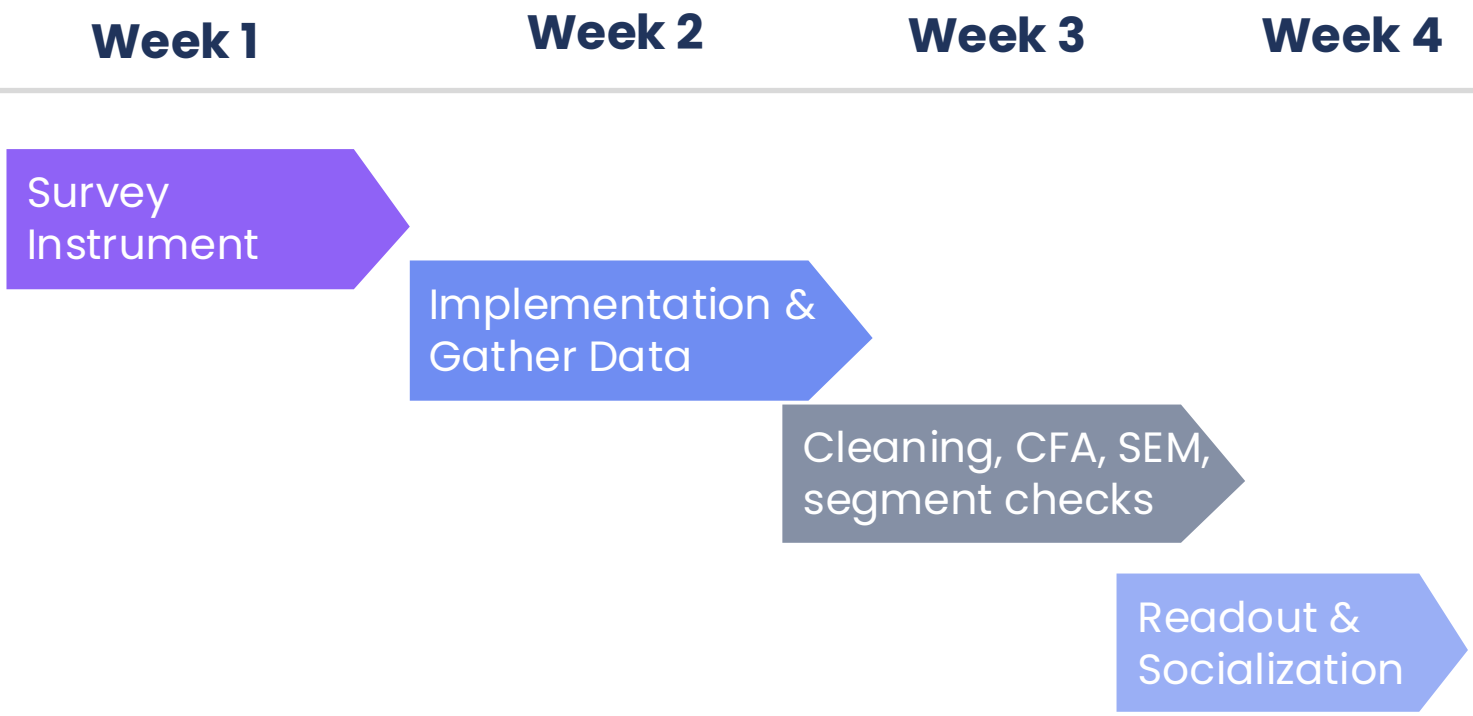
| | Hold Co | Indie Agencies |
|---|---|---|
| **Top CUJ** | Monitor, Report, Troubleshoot, Create Campaigns | Monitor, Report, Troubleshoot, Create Campaigns.<br>Manage inventory more than Hold Co*. |
| **CSAT** | Slightly higher than overall CSAT | Slightly lower than overall CSAT |
| **Trust in TTD** | 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

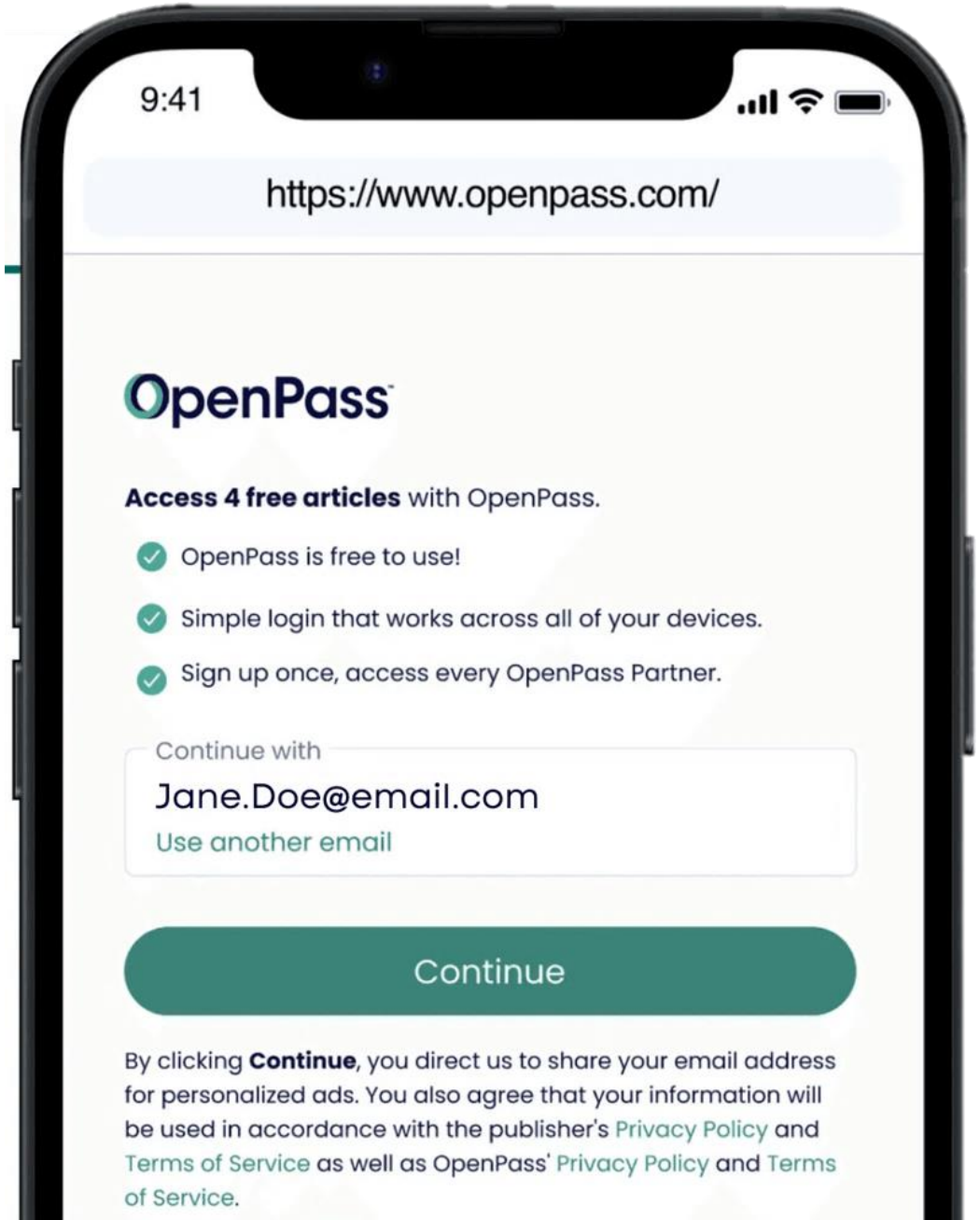
- Product Area PMs
- Data Science
- Qual UXR
- Content team
- Client Services
- Senior Leadership

# 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,

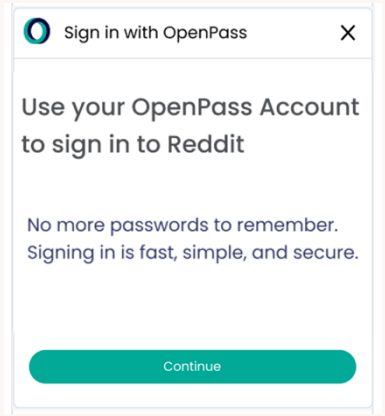
NOT ENOUGH TIME!

# A/B Survey

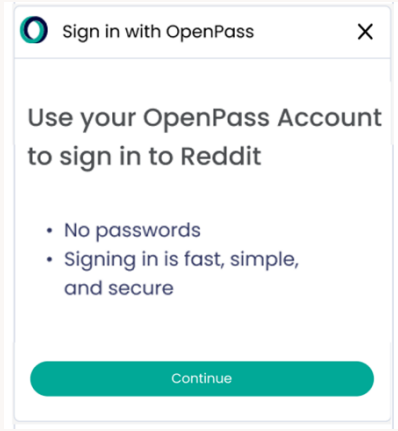
# Conjoint Attributes and Levels

## 1. Text Format

### Text

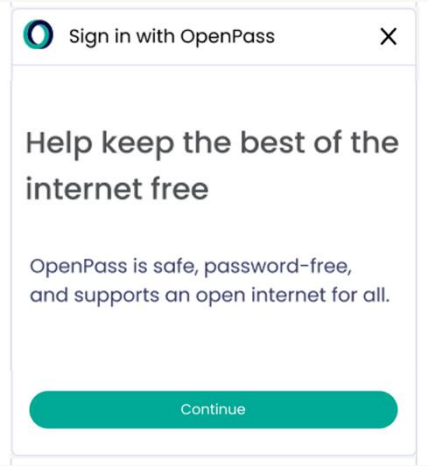



### Bullet

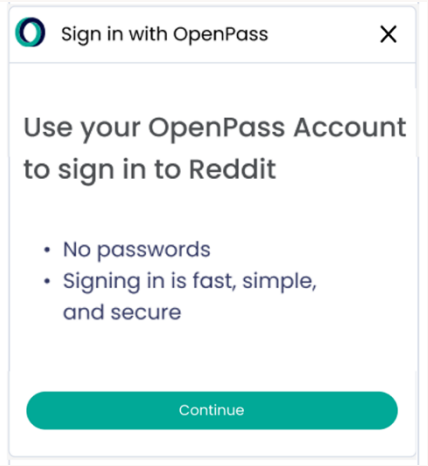



## 2. Content

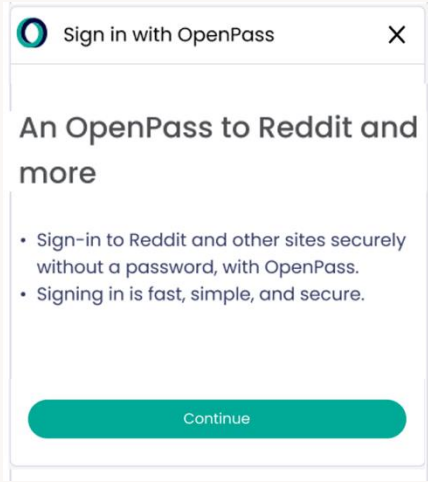
### Free



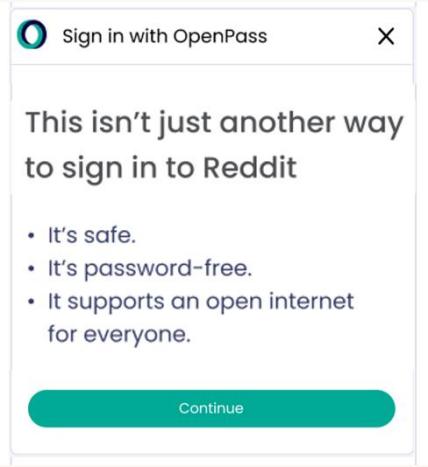

### Neutral



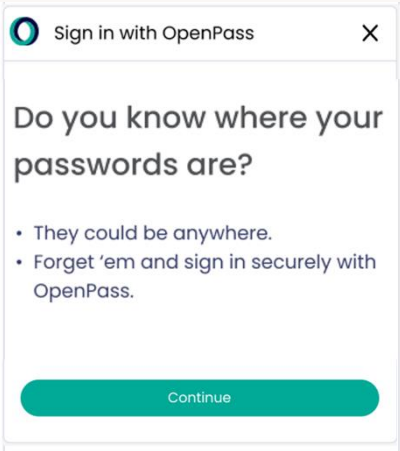

### More



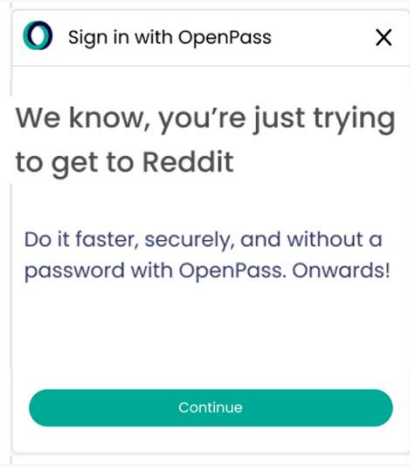

### Intrigue




### Fear



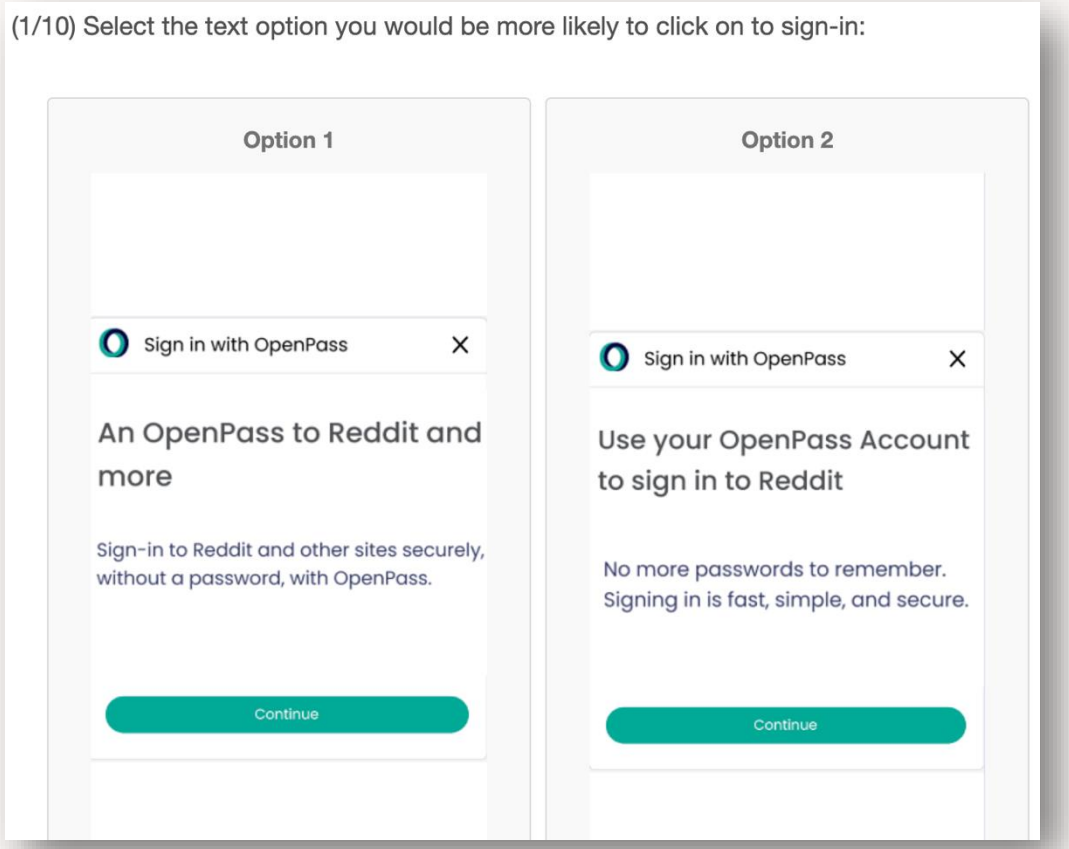

### Humor



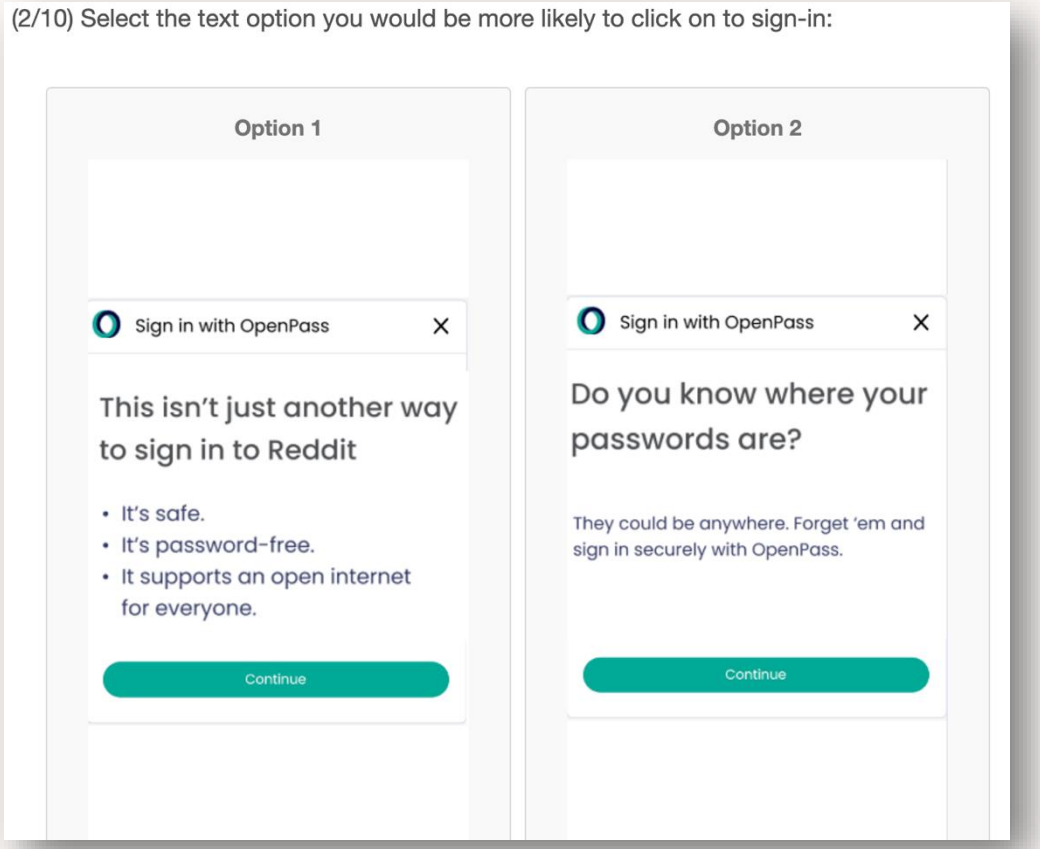
# Conjoint Methodology

**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:

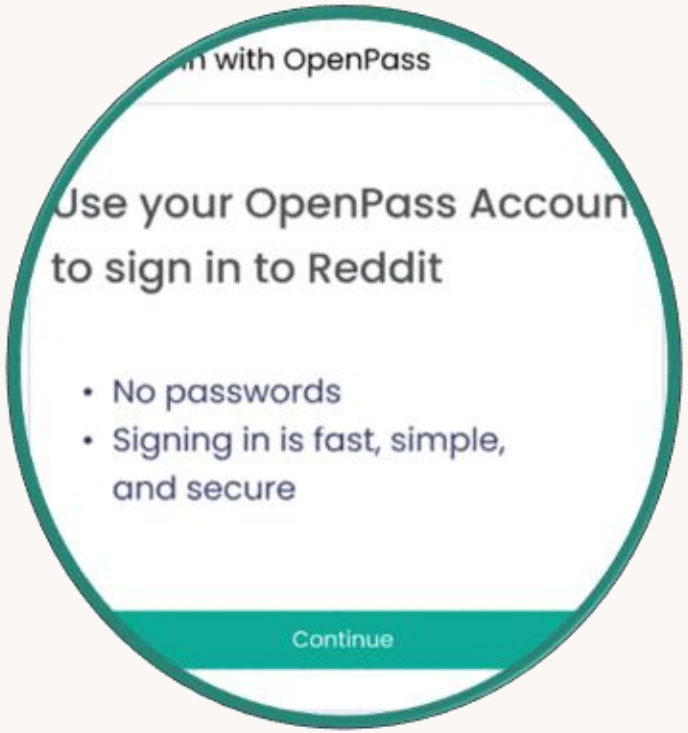


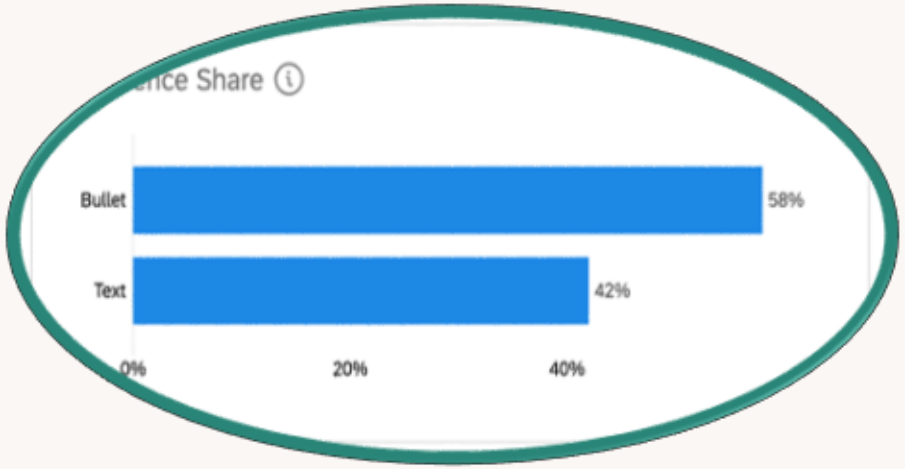

## Task 2:

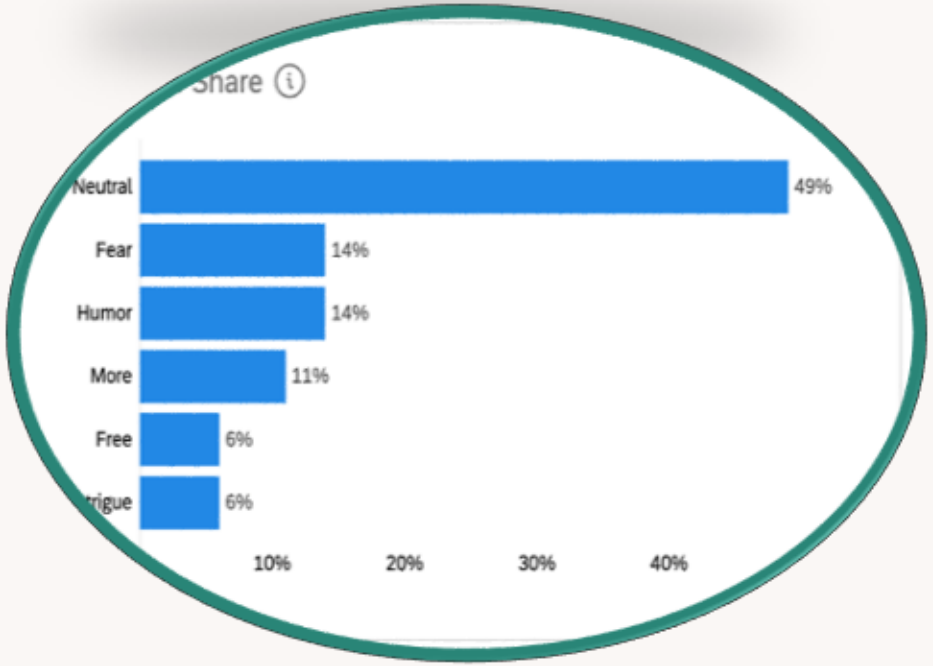


# Conjoint Methodology

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.

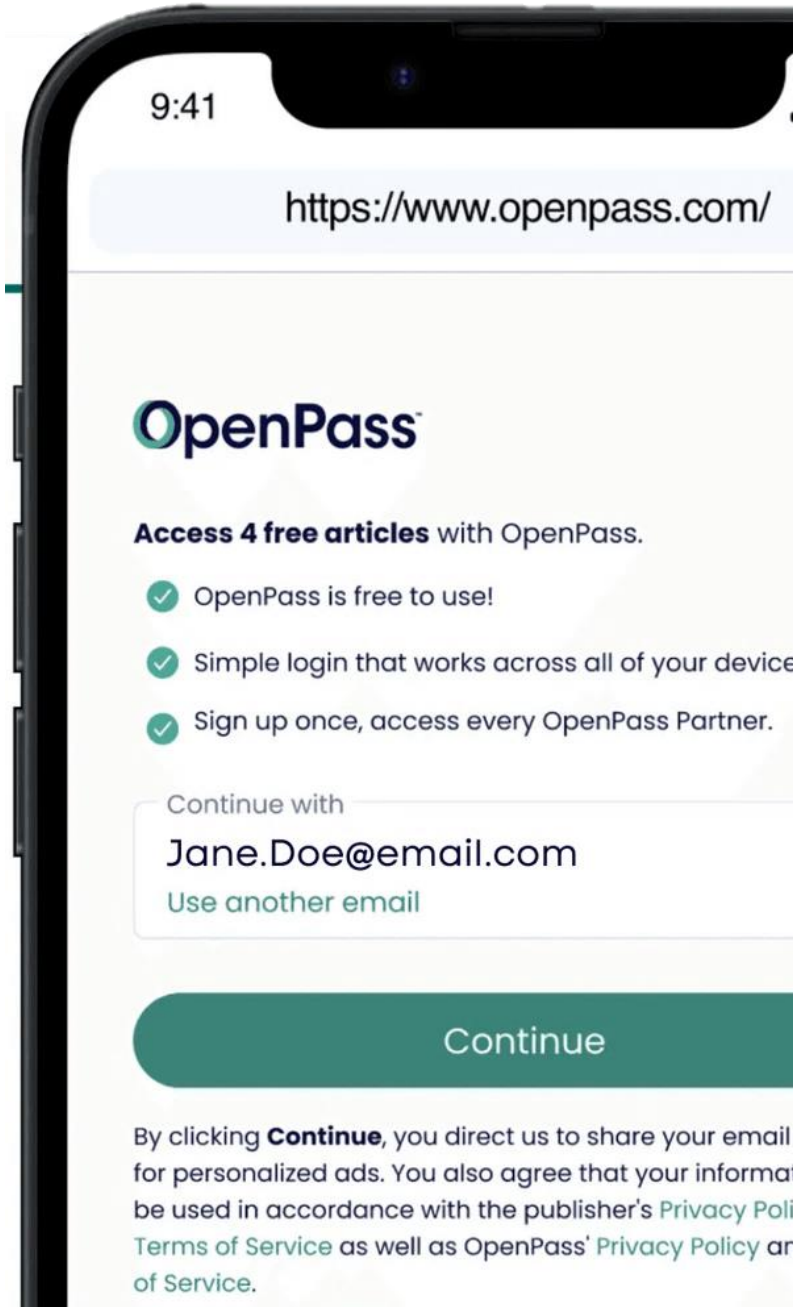
# Research Approach

- **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

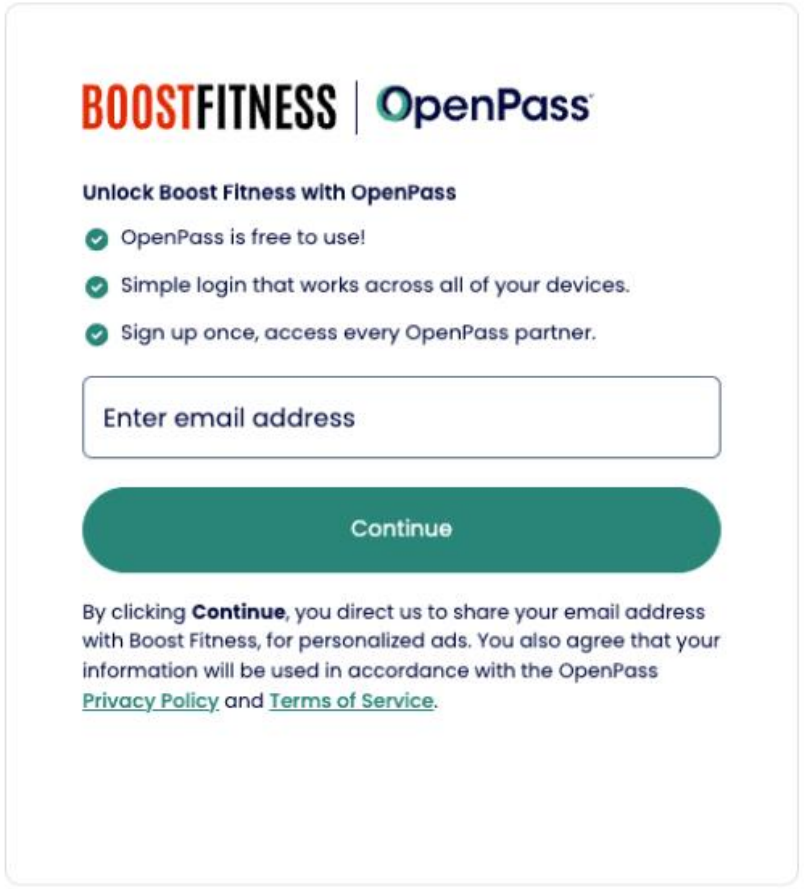
- Participants completed **two tasks**:
  - New user sign-in
  - Returning user sign-in

- 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

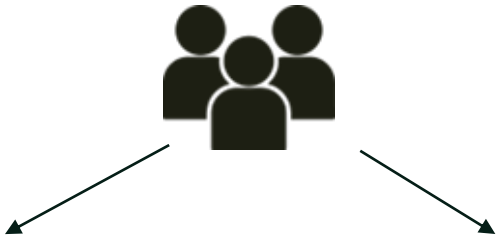


# A/B Test

OpenPass Sign-In

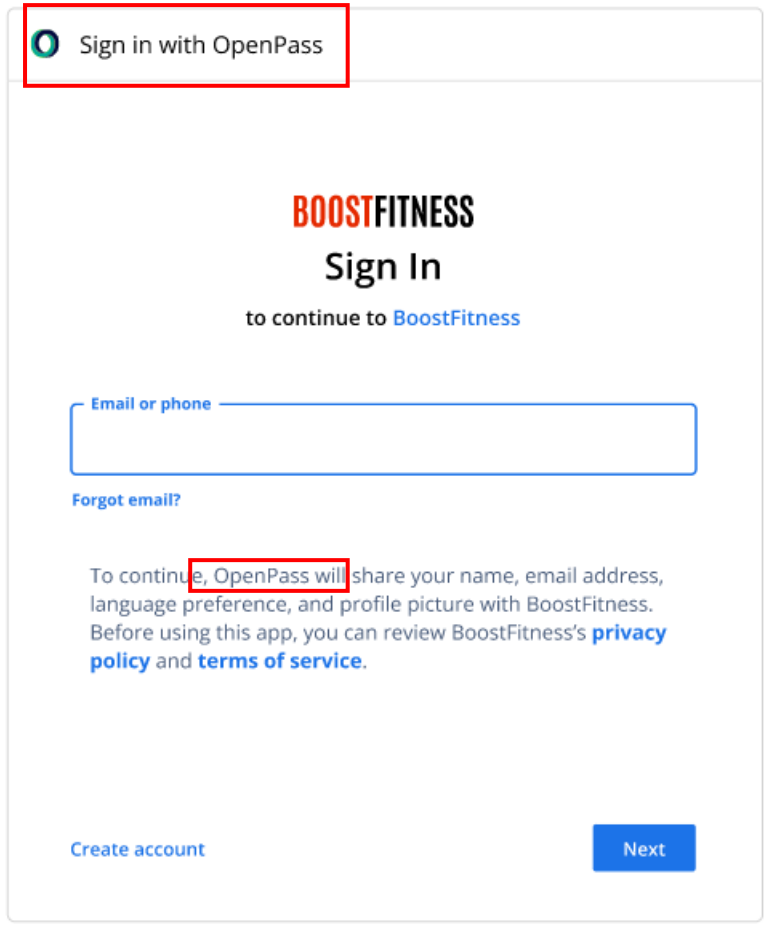



**300 consumers**

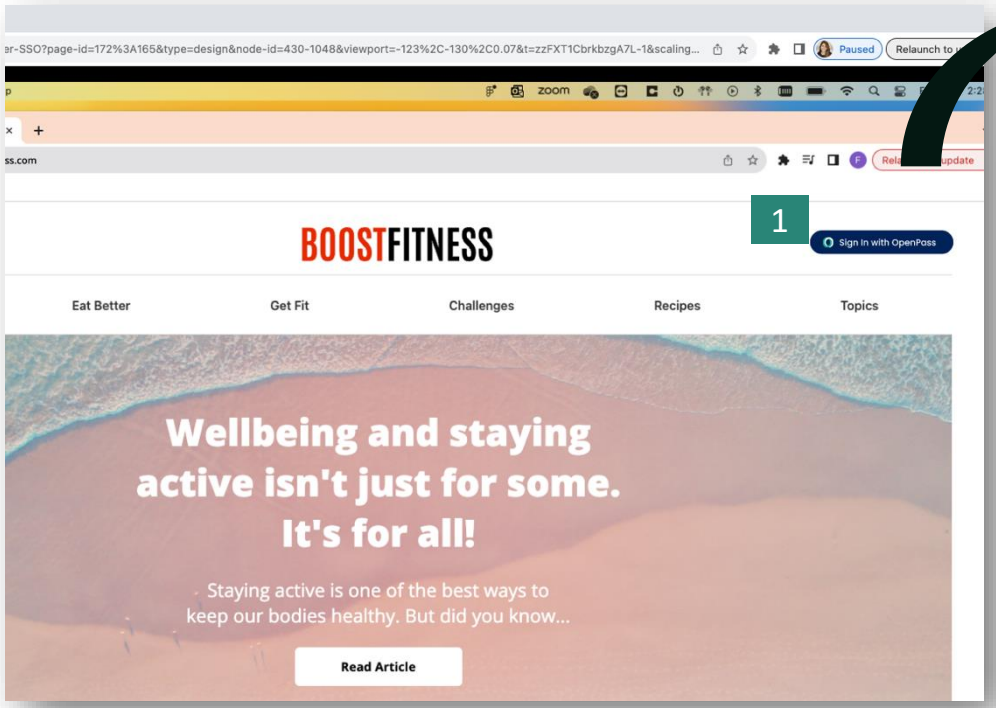


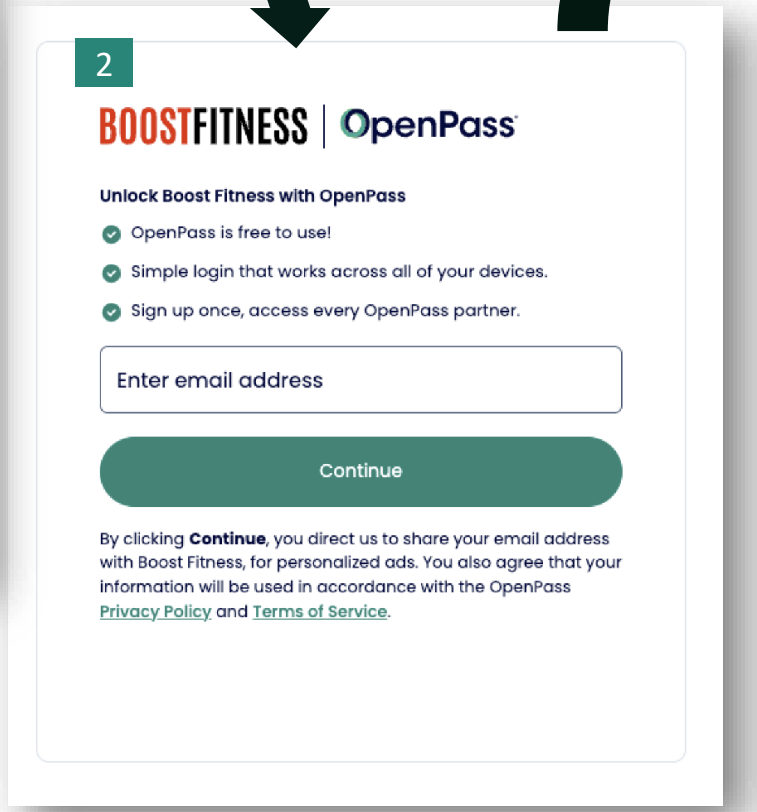
*Control for bias towards Google's brand recognition.
Fair test and comparison of the sign-in experiences.

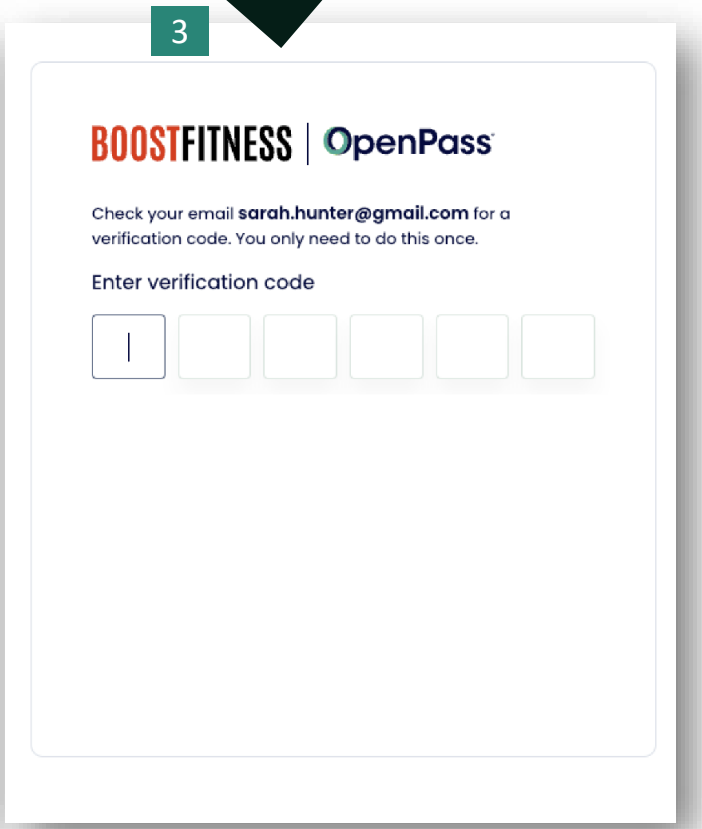
White Label Sign-In*



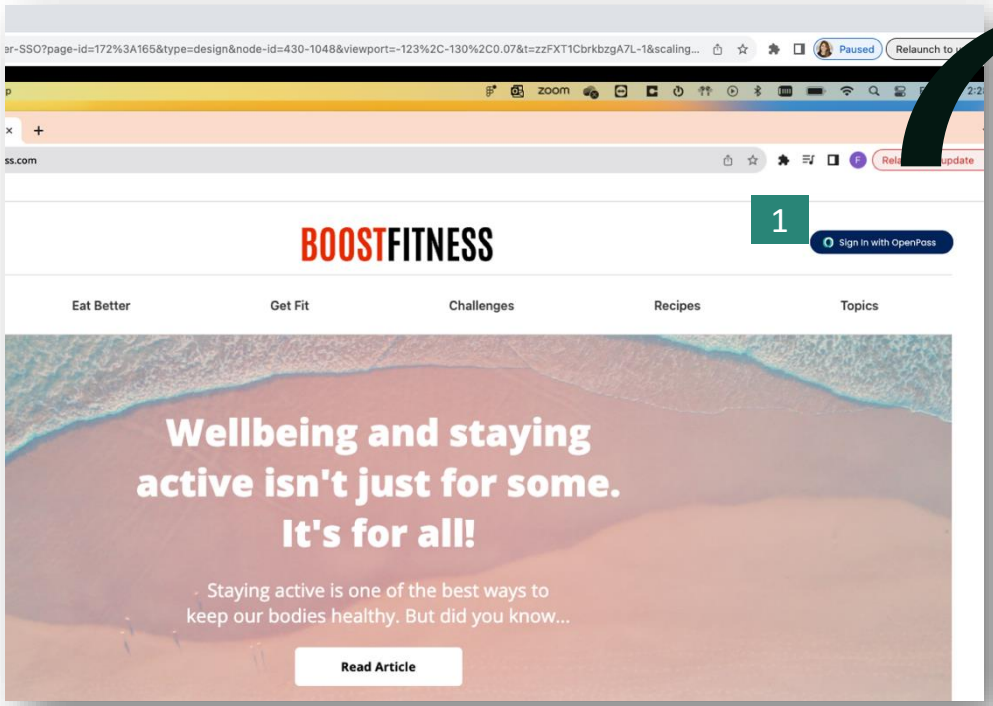
# OpenPass Workflow

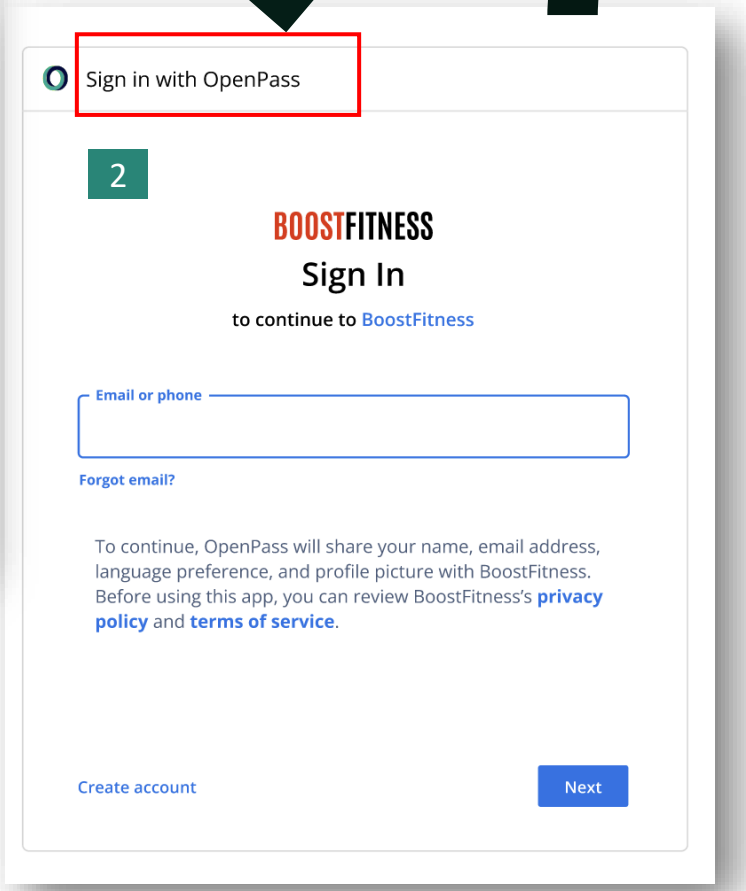


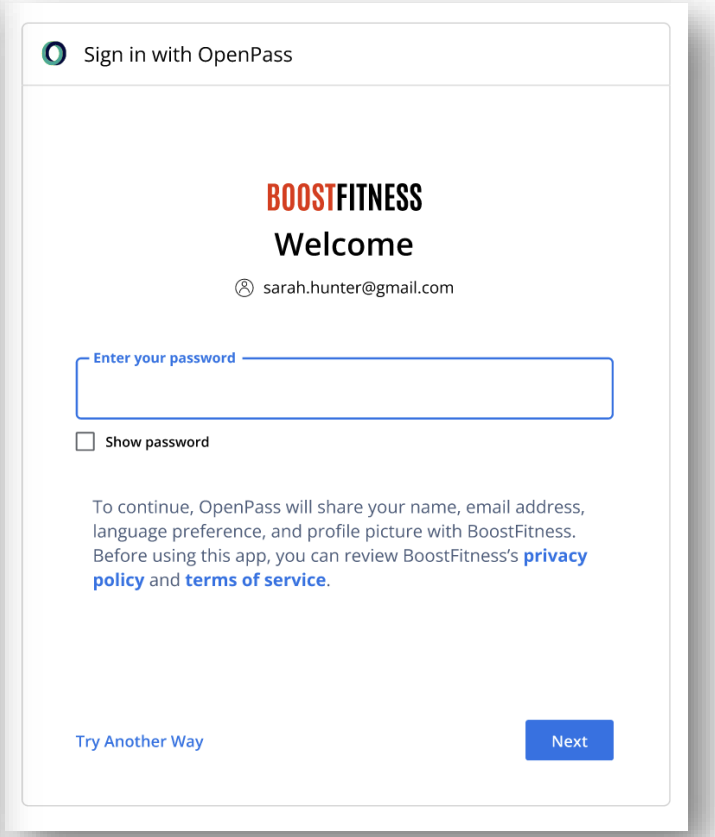



# OpenPass Workflow



# 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)