

The Funnel's New Gatekeeper: 48-Participant Behavioral Study of How Buyers Shortlist in AI Mode vs. Standard Google Search

Report Highlights:

- AI Mode turns shortlisting into acceptance: across the study, 74% of AI Mode shortlists came straight from the AI output, with 64% of participants clicking nothing.
- Trust worked differently across the two experiences: Search users more often cross-checked different SERP elements and outside sources, while AI Mode users leaned more on the AI's framing and familiar brands.
- AI Mode narrows the options, so brands that don't appear often don't get considered. Unfamiliar brands that appear often get passed over.
- When users leave AI Mode to external websites, they mainly do so to verify a candidate they have accepted. In standard Search, users click through to discover, compare, and build the shortlist.

This is the third usability research study in which [Citation Labs](#) participated to understand how people behave in Google Search and Google AI Mode.

This is not about what they say they would do.

It's about what they do.

Citation Lab sponsored the first two, which were published on Kevin Indig's Growth Memo. The series began with "[The First-Ever UX Study of Google's AI Overviews](#)." The second was "[What Our AI Mode User Behavior Study](#)" Reveals about the Future of Search, which examined how brands appear and perform in AI-generated results.

This study focuses on major purchases.

Working directly with [Eric Van Buskirk](#) and his team at Clickstream Solutions, Citation Labs set out to understand how people build a candidate set when they are close to a buying decision, and how that process differs between AI Mode sessions and Google's standard Search experience.

Note: *The comparison here is between AI Mode and Google's standard Search experience, where AI Overviews may still shape part of the path.*

EXECUTIVE OVERVIEW

This usability study shows that dedicated Google AI Mode intensifies shortlist acceptance behavior during bottom-of-the-funnel purchase research compared with Google's standard Search experience.

Using a remote, unmoderated think-aloud usability design, we observed 48 U.S.-based participants completing 185 major-purchase tasks.

We analyzed screen recordings and participants' speak-aloud audio. We also asked survey questions as they worked on tasks. In total, participants spent 23 hours on the study, averaging about ½ hour each. There were over 275 pages of transcripts from the audio.

Key Findings

How shortlists form

- 74% of AI Mode shortlists were generated directly from the AI's output, with no external verification or independent construction.
- 64% of AI Mode participants clicked nothing at all during their task. They read the AI text and declared the finalist choices.
- In the Search sessions, 89% clicked on a SERP element. More than half built shortlists from multiple sources, both in and out of the results page, independently, even though AI Overviews were still identifiable in many of those sessions.

How trust works differently

- In AI Mode, the AI's product description drove the decision in 48% of tasks. Framing, structure, and language functioned as the trust signal.
- In the Search sessions, that figure was 6%. Participants more often relied on corroboration across multiple independent sources (although some still used the AIO as a decision aid or entry point).
- Being ranked first in the sectioned answer layer of the AI response strongly predicted being chosen: 74% of AI Mode participants selected the top-ranked item.
- Ranking in position 1 of an AI Mode 'recommendation list' may be more valuable than ranking #1 in traditional search ever was, regardless of whether AI recommendation rank results in a click.

What participants did not do

- AI Mode showed fewer visible brand options than standard Search, but participants rarely noticed or complained that the choice set felt limited.

Brand consequences

- Unfamiliar brands were eliminated on name recognition alone—not on price, coverage, or ratings.
- Brands absent from the AI's output often were not evaluated at all.

Where the AI's pricing information broke down

- For physical products with shopping panels (e.g., washers, dryers, TVs), 85% of participants understood the pricing clearly.

The prompting posture effect

- 83% of AI Mode tasks used natural-language prompts; 15% used keywords
- Natural-language query prompting prompted the AI to return a complete list of recommendations, which participants then accepted.

Methodology

Study Design

This was a remote, unmoderated usability study to evaluate how U.S.-based consumers shortlisted major purchases in AI Mode and Google's standard Search experience. A think-aloud protocol was employed, in which participants spoke their thoughts aloud as they completed tasks. This design allowed us to capture both behavioral and attitudinal data.

Tasks

1. Suppose you or someone close to you is buying a high-end television and soundbar. Narrow this down today to two products you'd recommend for their consideration.
2. A friend's washer and dryer were damaged. Recommend two replacement sets with a budget of \$2,000.
3. Imagine your friend added a teen driver to their insurance, and their renewal price jumped. They need less expensive alternatives.
4. Suppose you're tasked with buying a high-end laptop for a small business. Shortlist two products.

Cognitive Load & Protocol Controls

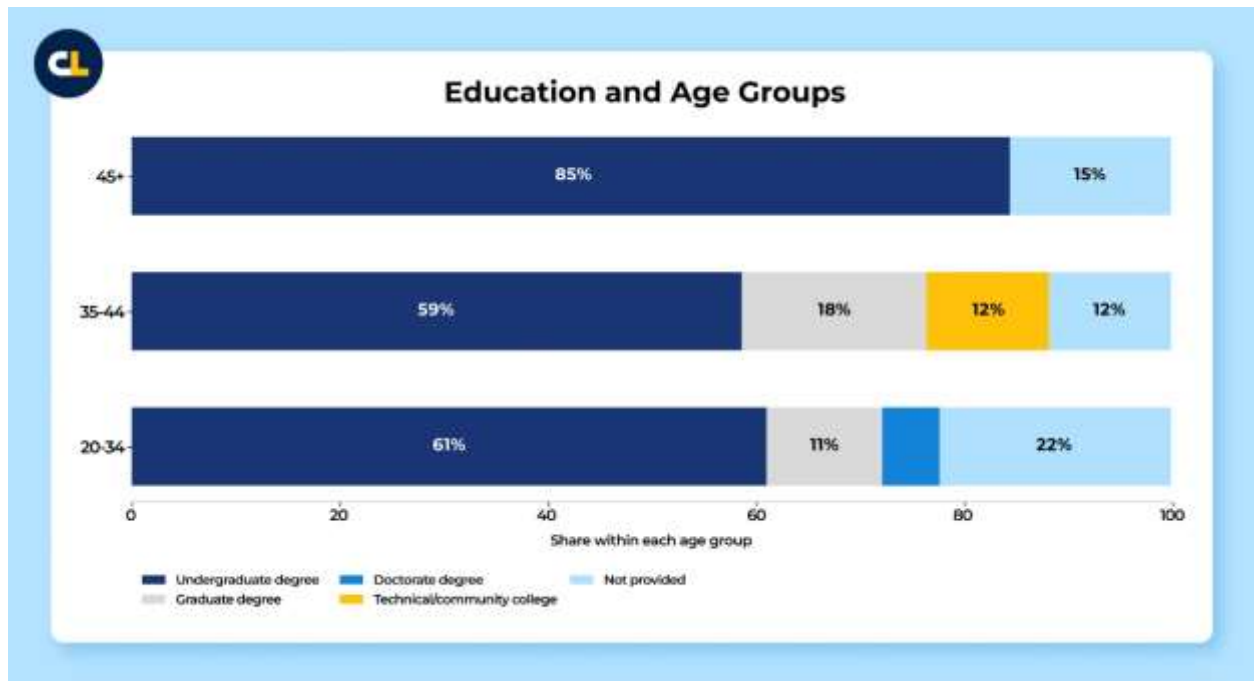
- Participants were instructed to shortlist 2-3 final options, not make final decisions
- Pre-task instructions designed to frame the task as narrowing, not deciding
- Participants were asked to spend about 18- 25 minutes completing the study.

Participants

A curated panel of 48 participants made decisions about major purchases across four product categories using Google AI Mode and Google's standard Search experience." They were recruited from [Prolific](#), an online research platform that connects researchers with a pool of vetted participants.

Each session lasted approximately 20–30 minutes, and participants received a monetary incentive for their time. Twenty-one participants were not included in the final 48: Seven had technical issues (for example, their long video didn't upload properly). 14 were excluded because their responses did not demonstrate adequate thoughtfulness or engagement.

Overall, participants were engaged and followed task instructions conscientiously. More than half received engagement ratings of 4 or 5 out of 5 from annotators, and the think-aloud protocols yielded rich qualitative material: participants narrated their reasoning, flagged uncertainties, and, in many cases, explicitly reflected on why they made the choices they did. This level of participant candor strengthens confidence in the behavioral findings.



Age Group Overview					
Age Group	Count	% Of 48 Total Participants	Mean Age	Min Age	Max Age
20-34	18	37.5%	27.8	23	33
35-44	17	35.4%	39.3	35	44
45+	13	27.1%	50.3	45	60

Education Level by Age Group				
Education Level	20-34	35-44	45+	Total
Undergraduate degree (BA/BSc/other)	11	10	11	32
Graduate degree (MA/MSc/MPhil/other)	2	3	0	5
Doctorate degree (PhD/other)	1	0	0	1
Technical/community college	0	2	0	2
Not provided	4	2	2	8
Total	18	17	13	48

Task Protocol

For each task, participants:

1. Read a task prompt.
2. Navigated to [Google.com/AI Mode](https://www.google.com/ai-mode) or [Google.com](https://www.google.com) (we specified which to use).
3. Entered keywords/queries they constructed.
4. Spoke their reasoning and reactions aloud.
5. Declared a final answer before closing the task.

Data Capture

The study was set up and conducted via [UXtweak](https://uxtweak.com), an enterprise-level Usability platform. It enables researchers to run studies and capture session recordings, provides tools to analyze user behavior, identifies friction points, and evaluates how users interact with digital products.

Annotation & Coding

Trained analysts systematically annotated each desktop video recording. They captured both quantitative and qualitative markers:

Each session was broken down into specific tasks so we could see what people did and how they did it. We tracked micro- and macro-behaviors, from hovers over links within the AI text to the results shown in the right-hand panel. We converted the audio to transcripts and analyzed those beyond the video coding.

The annotations turned complex, real-world interactions into a clear set of deeply analyzed data points. This allowed us to measure how people asked questions, which results caught their attention, how often they dug deeper into Google UI elements, and how smoothly the experience felt when moving outside Google. In short, it gave us a structured way to understand the outcomes and the user journey that led to them.

Statistical Confidence and Sample Size

With 48 participants, our confidence intervals are wider than they would be in a large-scale survey. Our findings are best understood as directional: they consistently point to patterns unlikely to be due to chance. However, the precise percentages should not be treated as population-level estimates.

A confidence interval is a statistical range that tells us how much a finding might vary if we ran the same study with a different group. Narrower intervals indicate more precision; wider intervals mean the true effect could be somewhat larger or smaller than what we observed.

Several factors strengthen confidence in these findings. Each of the 48 participants completed four purchasing tasks, yielding approximately 185 task-level observations after data cleaning—a substantially larger analytical base than the participant count alone would suggest.

More importantly, this study observed what people actually do rather than just what they say they would do. Behavioral patterns of this kind are more stable and more replicable than self-reported attitudes or intentions.

These findings provide a well-grounded set of directional signals to support your strategic decisions.

Core Hypotheses

BOFU Shortlist Logic: Does shortlisting follow different logic in AI Mode vs. Google's standard Search experience?

1. Standard Search experience = more progressive comparison across multiple sources, even when AI Overviews appear.
2. Dedicated AI Mode = more early acceptance of a model-generated shortlist, then refinement or justification within that environment.

Other Hypotheses:

- **Prompting Posture:** Keyword-style queries produce different behavior than natural-language prompts? (Keyword = verification-first; Prompt = delegation to the AI.)
- **Trust vs. Rank:** Trust comes from AI framing and brand recognition. Rank position is a bit overrated because marketers rely too heavily on click data to understand search behavior.
- **Consensus vs. Discovery:** Participants accept the AI's brand set and don't try to break out of it?
- **Pricing Opacity:** Participants often don't understand the real price: they accept AI-framed estimates as accurate.

FINDINGS

1. BOFU Shortlisting: Purchase Funnel Closes Differently in AI Mode and Search

We hypothesized that shortlisting during bottom-of-the-funnel major purchases would follow a different logic in AI Mode than in Google's standard Search experience.

In AI Mode, we expected users to accept an AI-generated shortlist early, then spend the rest of the task refining, verifying (even if just via UI elements like merchant cards), or negotiating within it. In the standard Search experience, we expected users to build their shortlist more progressively across multiple sources, even though AI Overviews might still influence parts of that process.

How participants were coded

Term	Definition
AI Adopted	Participants' final shortlist came directly from the AI output, with no external check or independent construction.
AI Verified	Participants took the AI's candidates but verified at least one externally before committing.
User Built	Participants assembled a shortlist from multiple sources, largely independent of the AI output.
Hybrid	A mix of AI-generated and self-found candidates.
Follow-up query	An additional prompt typed after the AI's initial response, within the same session.
Rank override	A participant engaged with or chose a lower-ranked item over a higher-ranked one in the AI's output.

2. The Default State in AI Mode Is Acceptance, Not Construction

Across all 149 AI Mode tasks, **88% of shortlists** came directly from the AI's output with no external check or independent construction (coded AI Adopted). Among the remaining participants, most still ended up with candidates the AI had originally suggested. **Only 8 of 147 codeable AI Mode tasks** produced a genuinely self-built shortlist.

AI Adopted captures the outcome, not a single pathway to it. Of the 117 AI Adopted participants, roughly **85% showed no internal verification behavior**. About **15% (17 participants) verified within AI mode**. The most common behavior was opening an inline product card or merchant pop-up to check a specific price, spec, or review without leaving the page. Participants would click into a card, read what they needed, and return to the AI's list.

Other participants used follow-up prompts as verification tools, asking the AI directly for prices, requesting alternative options, or re-prompting with additional constraints.

The most striking single behavior: **64% of AI Mode participants clicked nothing** at all during the task. They read the AI-generated text, often viewing inline product snippets, and declared their finalists without clicking any links, cards, or citations. For insurance tasks specifically, **that figure reached 85%**. Participants who did click still largely adopted the AI's recommendations.

Best for General Productivity & Travel

These models are favored by small business owners who need a balance of portability and power. <#>

Lenovo ThinkPad X1 Carbon Gen 13 14" Ultrabook 21n
 \$2,576.99
 CDW & more
 4.5 ★★★★★ (1.8K)

- **Expert Insight:** Named the "Best Business Laptop" for 2026 by Forbes Vetted due to its ultra-light carbon-fiber chassis and robust security.
- **Key Specs:** Features the Intel Core Ultra 7 (Series 2) platform with AI Boost, up to 64GB RAM, and a 2.8K OLED display.
- **Why it's high-end:** It undergoes MIL-STD-810H durability testing and includes a fingerprint reader and discrete TPM 2.0 for data protection.

14-inch MacBook Pro Apple M5 chip CPU and GPU
 \$1,899.00
 Best Buy & more
 4.8 ★★★★★ (1.8K)

- **Expert Insight:** Rated 4.8/5 by users for its exceptional "Liquid Retina XDR" display and cooling system.
- **Key Specs:** Equipped with the M5 chip, featuring a 10-core CPU and up to 21 hours of battery life.
- **Why it's high-end:** It offers industry-leading performance for creative tasks and superior build quality using premium aluminum materials. www.forbes.com #7

Main answer block with an icon link. A link in the second "ranking" product in this section is to related citations on the right sidebar.

These reviews highlight top-tier business laptops that balance performance, portability, and advanced security features for small businesses:

The 8 Best Laptops of 2026 - RTINGS.com

In addition to increased performance, there are a few other...

www.rtings.com

Best Laptops 2026 - Forbes Vetted

The Best Laptops, Based On Hours Of Testing. By Steven John, Writer...

www.forbes.com

Would you like a more detailed **specification comparison** between the Lenovo ThinkPad X1 Carbon and the MacBook Pro for your specific business tasks?

Example of scrollable product carousel in main answer block

In Search, the pattern was more distributed. **Nearly 89% of Search participants** clicked on something, and more than half built their own shortlist from multiple sources. The mechanics of how a decision gets made differ substantially across the two surfaces.

Table 1.1a Shortlist origin: AI Mode vs. Search (laptop + insurance tasks)



Shortlist origin	AI Mode (n=56)	Search (n=36)
AI Adopted	88%	21%
USER BUILT	5%	56%
AI VERIFIED	5%	18%
HYBRID	2%	6%

Laptop+insurance comparison subset. Full study AI Mode shortlist origin (n=181): AI Adopted 69%, USER BUILT 15%, AI VERIFIED 12%, HYBRID 5%.

Actionable Insights:

- If your brand doesn't appear in Google's AI shortlist, buyers often won't evaluate it because they won't see it. And they won't know to go looking for it.
- [Optimize for inclusion in AI-generated shortlists](#) rather than only for click-through rate. A brand that appears in the AI answer block has a better chance of entering consideration than a high-ranking brand in standard search results that doesn't appear in the AI output.

Don't assume standard SEO performance [carries over into AI Mode](#). No-click data shows participants often don't reach most of what standard Search surfaces.

Table 1.1b No-click rate by task (AI Mode, all tasks, n=149)

Task	No-click rate	n
Insurance	85%	27
Television	67%	46
Laptop	69%	29
Washer/dryer	47%	47
All AI Mode	64%	149

Actionable Insights:

- For physical product categories like TVs and appliances, the no-click rate is lower, indicating there is still some browsing behavior to capture.
- Start monitoring whether your brand appears in AI Mode responses for your [top BOFU queries](#). This needs to become part of normal measurement.
- High no-click rates mean many sessions can end on Google without a visit to your site. Treat that as [a real shift in buyer behavior](#).
- For categories with lower no-click rates, product pages and comparison pages can still drive traffic. Make those pages fast, clear on specs, and easy to navigate.

Participant quotes: acceptance without deliberation

*"Given that the first paragraph says Lenovo or Apple... going with that."
P672489, laptop, AI Mode. Position one in the AI response was the entire decision.*

*"I liked it more than anything else I've ever used for product searching — it made it a lot quicker to find the options."
P699b5c, 25 years old, laptop, AI Mode. Speed experienced as a feature, not a shortcut.*

*"I already suspected that these would be amongst the top one to five recommendations, and the fact that it confirmed that makes me much more confident."
P699f2f, 38 years old, laptop, AI Mode. Verification used to amplify confidence rather than to check.*

*"I'd have to go blind and this is the information I can rely on, so let's hope it's right."
P6990045, laptop, AI Mode. Acceptance here is not trust — it is the absence of an alternative. Participant may feel a bit overwhelmed by shortlisting with minimum time.*

"Travelers and USAA actually tell me how much, whereas State Farm and GEICO give percentages. Just knowing the exact amount makes me want to pick Travelers or USAA right off the bat."

3. How Search Participants Worked Differently

Search participants whose final shortlist was self-built followed a recognizable multi-step process.

The clearest split is between participants who actively engaged with the AIO and those who noted it briefly and moved on.

With Active AIO engagement as the primary surface for the final choice, the behavioral pattern is nearly indistinguishable from AI Mode.

When participants actively engaged with and accepted from AIO, the functional behavior is the same as AI Mode:

- Shortlist formation code is ACCEPT in all 9 active-engagement task cases.
- Outcomes are AI ADOPTED (5 tasks) or AI VERIFIED (3 tasks), with one HYBRID.
- Trust mechanisms are AI TRUST or AI FRAMING in the clear-acceptance cases.
- No external verification beyond the AIO itself in the pure ADOPTED cases.

When participants briefly engaged with AIO, they typically acknowledged or reacted positively to it, but ultimately built their shortlist from organic results, sponsored listings, or external sites.

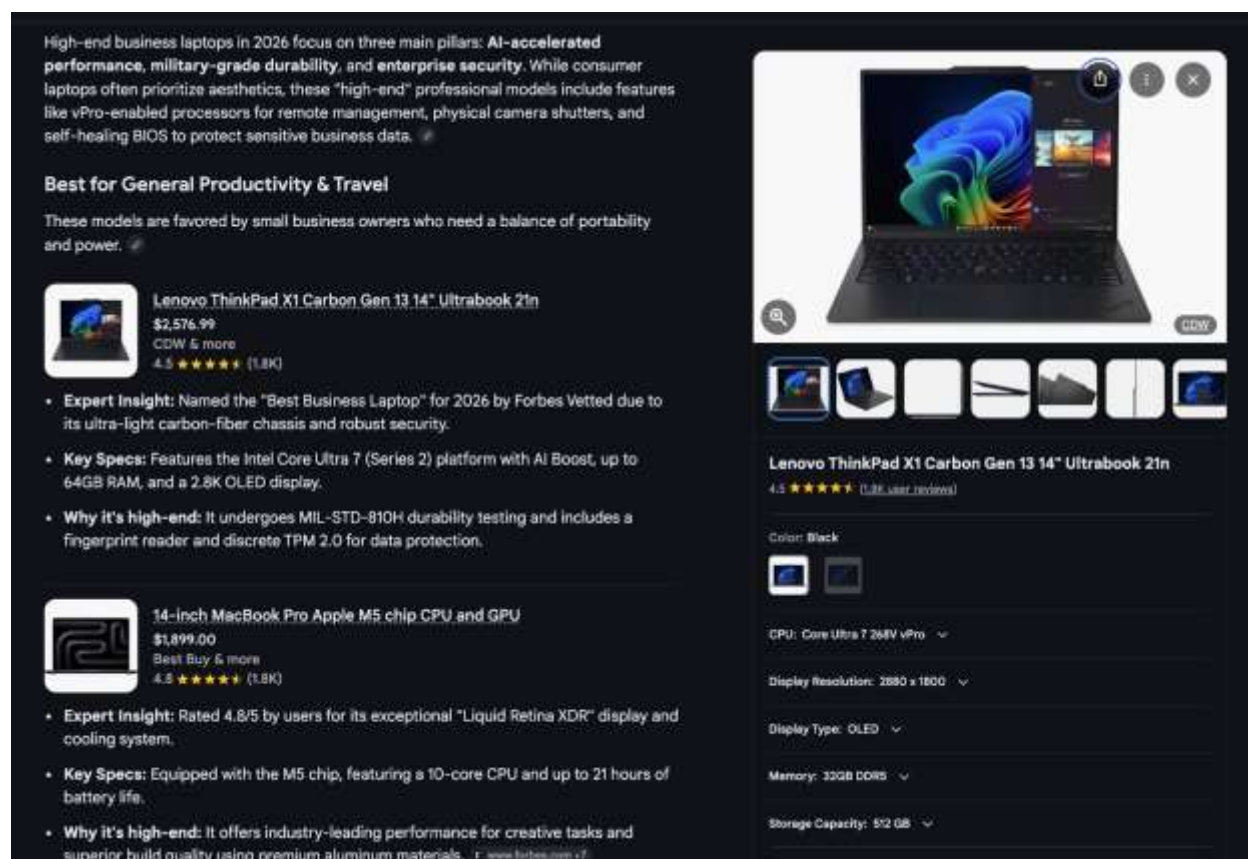
One participant (insurance, Search) clicked out to Progressive and Geico independently, read both landing pages, consulted an Experian article, and then arrived at a shortlist. Another (laptop, Search) applied hardware filters and then flagged a review score discrepancy:

"It shows 4.6 out of five stars for the reviews, but when you actually click the link: not reviewed yet."

laptop, Search. Active skepticism of aggregated data — a behavior absent from AI Mode transcripts.

4. How Behavior Differed by Shortlist Origin

Participants whose final shortlist came directly from the AI (coded AI Adopted) finished fastest. Most showed no internal verification behavior before accepting. **About 15% opened product cards or clicked inline links** in AI Mode first, but all remained in AI Mode.



Expanded view of AI mode text with the Product card clicked, opened

Participants who built their own shortlist from scratch took on average 89 seconds longer and consulted more than twice as many sources. This pattern held consistently across all four product categories.

Follow-up queries, additional prompts typed after the initial AI response, told a different story. Participants who accepted the AI's list issued the fewest follow-ups, with more than half issuing none at all.

Those using a mixed approach issued the most, suggesting follow-ups were how they negotiated changes to the candidate set without leaving AI Mode.

Participants who built their own list asked almost no follow-up questions; they stopped engaging with the AI and went elsewhere. If the differences were solely due to the “accept” group trying to finish faster, we would not see this.

Among participants who issued follow-ups, the most common type was a refinement query that narrowed by specs, price, or other constraints. Clarification requests came next, followed by delegation (asking the AI to simply decide) and reframing (shifting to a different category or seller).

Table 1.2a Behavior by shortlist origin type (AI Mode tasks, n=147 codeable)

	Accepted AI list(n=117)	Verified externally(n=15)	Built own list(n=8)	Mixed approach(n=7)
Sources consulted before committing				
Mean	1.5	2.7	3.4	2.9
Time on task (seconds)				
Mean	171	215	260	224
Median	150	193	257	180
Follow-up queries				
Mean per task	0.9	1.3	0.3	1.4
Tasks with none	56%	47%	88%	14%

Actionable Insights:

- Treat follow-up queries as the narrowing stage in AI mode. If you win on price, specs, or fit, make that information [easy for AI to retrieve](#) and easy for buyers to confirm.
- Don't assume every buyer will visit your site before deciding. Some will ask AI to choose for them, so your product needs to show up with a clear reason to pick it.

Table 1.2b Follow-up query types (AI Mode tasks with identifiable follow-ups, n=69)

Type	Description	% of tasks
Refine	Narrow by specs, price, or constraints	42%
Clarify	Ask why a recommendation was made	35%
Delegate	Ask the AI to simply choose	30%
Reframe	Shift to a different category or seller	26%

Percentages sum above 100% because some tasks included more than one follow-up type.

Actionable Insights:

- Don't ignore buyers who leave AI Mode to verify elsewhere. Strong editorial coverage, review presence, and comparison listings still shape that decision.
- Build key pages for buyers who arrive mid-decision. Comparison pages, spec sheets, and pricing pages should [help someone confirm a shortlist](#), not start from scratch.

5. Reading Behavior: Mostly Skimming, Even Among Deep Readers

Across all AI Mode tasks, **70% of participants skimmed**. They glanced for takeaways without reading carefully. **Only 18% scanned the page for something specific**, and **fewer than 7% read through comprehensively**, meaning they engaged with the majority of the response carefully, even if their reading pattern was non-linear or not section-focused.

What makes this finding notable is that among those who read deeply, **88% still ended up with an AI-adapted shortlist**. Deep reading in AI Mode did not result in the rejection of AI recommendations. Instead, it produced more confident acceptance.

Compared with Search, the skim rate was similar across both surfaces (**66% AI Mode, 62% Search, no meaningful difference**). Scan behavior was more common in Search (**27% vs. 11%**), while multi-pattern reading appeared more often in AI Mode tasks.

The AI answer block's mixed content, including text, tables, shopping panels, and inline ratings, created natural attention shifts within a single area rather than across the page.

"This graph really drew my attention."

P694418, laptop, AI Mode. The AI pre-processed the comparison; shallow reading was sufficient to close the decision.

"Ooh, they don't like it. XPS three stars out of five, not good."

P6942bd, laptop, AI Mode (coded SKIM + SCAN). One of the more active reading participants was still entirely within the AI response.

Table 1.4 Reading pattern distribution (AI Mode, n=147 valid tasks)

Reading pattern	% of AI Mode tasks
SKIM (glanced for takeaways)	70%
SCAN (searched for something specific)	18%
Comprehensive	7%
Multi-pattern (skim + scan or deeper)	~9%
Deep down page	5%
For those that are deep/mixed readers: they are still AI Adopted	75%

Actionable Insights:

- Most buyers will skim AI outputs. If the opening description gets your product wrong or frames it weakly, many users won't stick around long enough to fix that impression.
- More detail [won't save weak positioning](#). In AI Mode, the clearer and sharper the description, the better your chance of staying in the set.
- Format matters. If you want your content to [survive inside AI answers](#), use a structure that's easy to extract and easy to scan.

Search traffic still behaves more like lookup behavior. Make specs, differentiators, and key answers [easy to find fast](#).

6. Position in the AI Output Strongly Predicted Final Choice

Across all AI Mode tasks, **74% of participants chose** the item ranked first in the AI's response as their top pick.

Defining "rank" in AI Mode is tricky. We chose to disregard where a section sat on the page in favor of the position within a section. Across our entire study, here are the section types we saw. A section typically contains two to five citations.

The mean rank of the final choice was **1.35**. Only **10% of participants chose** something ranked third or lower.



This word cloud shows the language Google commonly uses in its section headers, weighted by how often each term appears. We stripped out words specific only to specific tasks and focused on adjectives. There’s a focus on “high-end” because we were looking at major purchases. There are other terms that help us understand what framed the final short-list choices.

About **26% of participants overrode rank order**, typically based on brand recognition. But overriding rank did not mean rejecting the AI’s output: **81% of rank-override participants** still ended up with an AI-adapted shortlist. They chose a different item from the AI’s list, not an item from outside it (e.g.a card from a shopping carousel).

In Search, rank override rates were statistically similar (**29% vs. 26% in the laptop+insurance comparison**), but the consequence differed: it typically meant navigating to a different source entirely rather than choosing a different item from the same list.

Table 1.5 Rank of participants' first choice

Measure	AI Mode (n=137)
Choose rank 1	74%
Mean rank of first choice	1.54
Choose rank 3+	10%
Rank override rate	26%
Of overrides: still AI Adopted	81%

Actionable Insights:

- Treat placement near the top of the AI response as [a major advantage](#). If your brand is not near the top, you have less room to win the click or the choice.
- Pay attention to how Google frames the options it surfaces. If AI answers group products under labels like “best,” “recommended,” or “high-end,” your product descriptions and structured data should support the same framing.
- Don’t think only in terms of rank #1. [Being in the AI shortlist](#) at all is far better than being absent, because even buyers who skip the top pick often stay inside the set AI provided.
- If you’re a challenger brand, recognition still matters. Once you make the shortlist, familiarity and clear positioning can help you beat a higher-ranked option.
- Don’t read AI Mode ranking the same way you read search rankings. A top placement may lead to a quick confirmation step rather than a longer discovery journey.

7. How Participants Decided to Trust a Recommendation

Participants' videos were coded for the primary signal that drove their final choice. The two most common were nearly tied: AI framing (the way the AI presented and described a product drove the decision) and brand recognition (the participant already had a preference for that brand, and the AI confirmed it).

The split between these two signals tracked closely with the product category. For televisions and laptops, where most participants already had brand preferences, brand recognition dominated. For insurance and washer/dryer purchases, where participants had less prior knowledge, AI framing dominated. When you do not have a prior view, the AI's description becomes the trust signal.

The comparison with Search sharpened this. AI framing drove decisions in **48% of AI Mode tasks but only 6% of Search tasks**. Multi-source convergence appeared more often in Search (**37% of tasks**) than in **AI Mode (4%)**. In Search, participants more often built trust through corroboration. In AI Mode, the AI's framing carried more of that work..

Table 1.6 Trust mechanism by task and surface (multi-coded, % of tasks)

Trust signal	All AI Mode	TV	W&D	Laptop	Insurance	Search (l+i)
AI framing	37%	24%	36%	38%	59%	6%
Brand recognition	34%	37%	23%	55%	26%	23%
AI trust (general)	26%	28%	30%	21%	22%	11%
Source trust	17%	17%	26%	17%	4%	23%
Multi-source convergence	5%	4%	6%	0%	7%	37%

Actionable Insights:

- The way AI mode describes your product now shapes trust. [You can influence answers](#) with clear product data, strong manufacturer copy, and consistent language across crawlable sources.
- In categories where buyers don't bring strong brand recognition, AI framing carries more weight. That makes upstream brand familiarity more important before the search starts.
- In categories with strong brand recognition, brand equity still helps. Keep investing in the familiarity and positioning that make your brand easier to choose.
- [Citations matter](#), not just brand mentions. When AI draws on credible sources, those sources can reinforce the recommendation.

8. Task-Level Variation Exists, But Does Not Change the Pattern

The four product categories differed in the extent to which participants delegated to the AI. Insurance participants delegated most heavily: **85% clicked nothing**, and most ended up with an AI-adapted shortlist origin, meaning their final candidates came directly from the AI output.

Washer/dryer participants were the most likely to verify or build their own shortlist, probably because the category requires matching against specific constraints, capacity, stacking compatibility, and dimensions that the AI's summary did not always resolve.

9. AI Mode Produced Narrower Brand Outcomes Across All Categories

Across all 142 valid AI Mode tasks, brand outcomes were highly concentrated. In insurance, three brands, Geico, Travelers, and State Farm, accounted for **roughly 64%** of all AI Mode final choices. In Search (laptop and insurance tasks only), those same three brands **accounted for about 30%**, with Progressive, Erie, Mercury, and Liberty Mutual each appearing multiple times.

The pattern held across product categories. Among laptops, Lenovo ThinkPad and Apple MacBook variants together accounted for most AI Mode choices, with no other brand appearing

more than once. Search distributed more broadly: HP EliteBook variants appeared three times, ASUS once. For televisions, Samsung and LG dominated AI Mode outcomes. For washer/dryer, LG variants and Electrolux led.

The implication is structural: brands absent from the AI-generated shortlist may never enter consideration at the bottom of the funnel because they were never surfaced in the first place.

Table 1.8 Top final choices by category (AI Mode, valid tasks)

Category	Top choice	2nd	3rd	Top-3 share
Insurance (n=25)	Geico (52%)	Travelers (48%)	State Farm (28%)	~64%
Laptop (n=29)	Lenovo ThinkPad (~52%)	Apple MacBook (~34%)	Dell XPS (~7%)	~93%
Television (n=42)	LG (~21%)	Samsung (~14%)	Sony (~10%)	~45%
Washer/dryer (n=46)	LG variants (~33%)	Electrolux (~11%)	Whirlpool (~4%)	~48%

Percentages approximate. LG variants consolidated across naming inconsistencies in raw data.

Actionable Insights:

- Treat AI Mode as a winner-take-most environment. If your brand is [not in the small set Google surfaces](#), you are unlikely to win meaningful share from BOFU sessions there.
- Don't read this only as a conversion problem. In many cases, buyers may never meaningfully encounter the brands outside that shortlist.
- If you're a challenger or regional brand, don't expect one page to fix this. You need [sufficient editorial coverage](#), review presence, and structured visibility so your brand shows up consistently in the AI set.
- If you already appear [near the top of the AI output](#) in your category, defend that position. Keep product data, pricing, and review signals up to date and consistent.
- DTC brands with strong sites and weak third-party coverage are exposed. AI draws on consensus and authority, so off-site presence matters as much as on-site presence.

If you're outside the [top tier of AI visibility](#) in your category, don't expect AI Mode to send much traffic. Put more effort into Search, comparison sites, and the external sources that still influence buyers before and after AI.

10. When Participants Left AI Mode, They Were Verifying, Not Exploring

Across all AI Mode tasks, **23% involved at least one visit to an external website**. In Search, that figure was **67%** for the laptop and insurance tasks. When AI Mode participants did leave, their destinations reflected verification behavior: retailer sites (Best Buy appeared in 10 of 34 AI

Mode tasks with external visits), manufacturer sites (Dell, Lenovo, Apple, Samsung), and price-comparison utilities for insurance (LendingTree, WalletHub, Bankrate).

In the AI Mode sessions that did include an external visit, participants usually went to retailer sites, manufacturer pages, or comparison utilities to check details on options they were already considering. Best Buy appeared in 10 of 34 AI Mode tasks with external visits. Other destinations included Dell, Lenovo, Apple, Samsung, LendingTree, WalletHub, and Bankrate.

Note: We did not systematically code whether every external destination had first appeared in the AI answer or the Search results. The pattern we can support is behavioral: when participants left AI Mode, they usually did to confirm a candidate rather than to widen the field.

One Search participant noted unprompted: “it’s pulling in a lot of aggregator sites or news sites rather than a lot of company website results.” This illustrates the broader Search pattern: participants encountered a more mixed set of source types and used those sources to compare options, not just confirm one.

The behavioral contrast is straightforward: in AI Mode, external visits usually supported verification of a candidate already under consideration. In standard Search, external visits more often supported discovery and comparison.

The exit behavior difference was directional: in AI Mode, participants usually left to verify a candidate already in contention. In Search, participants more often left to discover or compare candidates.

11. Price Clarity: A Potential AI Mode Advantage

Price clarity was coded in **76 of 149 AI Mode tasks** (the remainder had no price-related evaluation visible in the session). Among those 76 coded tasks, **67% were rated CLEAR**, the participant accurately understood what they would pay. But that headline figure obscures a sharp split across product categories.

For washer and dryer tasks, 28 of 33 coded participants (**85% were rated CLEAR**). AI Mode’s shopping panels displayed explicit prices, retailer names, and sale amounts directly alongside product recommendations. Participants could read the actual cost without clicking anywhere. Quotes from those sessions reflect this: “\$1,299 at Lowe’s, that’s definitely in our budget”; “Best Buy \$1,079, Lowe’s \$1,100, free delivery, that’s nice”; “it’s on sale for \$1,672.” The price was surfaced; the participant just read it.

Insurance told the opposite story. **Of 16 coded insurance tasks, 10 (63%)** were rated overconfident/rash. Participants accepted AI-quoted rates or discount descriptions without checking whether the figures applied to their friend’s actual state, age, driving record, or current insurer.

The AI presented plausible-looking numbers, and participants took them as settled. One annotator noted: “accepts rate figures without verifying state or profile.” Another participant said the AI had premiums “listed in bold” and that they would be “really useful to share as a comparison”, treating formatted AI output as verified data when it was a general estimate.

Laptop was a third pattern: the most common problem was simply that prices were not shown. Four participants were rated CONFUSED, with quotes like “It’s not giving me prices, which is not helpful” and “I noticed it did not list the price just right here in the results; using an external

website was necessary.” When AI Mode surfaced laptop prices, participants engaged with them clearly.

The category split maps directly onto the AI Mode interface. Shopping panels show explicit, retailer-confirmed prices for physical products. Insurance recommendations in AI Mode are text-based summaries of rate ranges and discount categories; the AI cannot quote an actual premium without knowing the user’s specific details, and most participants did not notice that gap. The format looked authoritative enough that they did not ask.

Search had too few coded price clarity tasks (7 of 36) for a reliable comparison. Within AI Mode, shopping panels improved price clarity for physical products, while text-based insurance summaries created pricing risk.

Price Clarity Actionable Insights

1. Structured pricing data is your most direct lever for appearing clearly in AI Mode.

The washer and dryer results showed that when AI Mode surfaces explicit pricing, retailer name, sale price, budget fit, participants evaluated it confidently and moved to a decision.

That price visibility came directly from Google’s Shopping Graph, which is fed by two sources: structured data markup on your product pages and your Google Merchant Center feed.

Both matter.

2. Overconfidence is an opportunity for content and disclosure for services and insurance

The insurance findings show the opposite problem to the washer/dryer case. The AI surfaced average rate ranges and discount categories. Participants treated those as quotes. This is a known failure mode for AI in services categories, and it has direct consequences for any brand whose pricing is context-dependent: insurance, financial products, SaaS with variable pricing, healthcare, and legal services.

The actionable response is not primarily technical; it is editorial. Brands in these categories should ensure their landing pages, product descriptions, and FAQ content explicitly state what determines the price: “Your rate depends on your state, your driving record, and your current insurer.”

When AI Mode summarizes a page that clearly frames price as conditional, that conditional framing is more likely to appear in the AI’s output than if the page only lists rate ranges. Structure FAQ content using a machine-readable question-and-answer format, with questions like “How is my rate calculated?” and “What affects my premium?” give the AI explicit material to cite when framing your pricing.

There is also a downstream risk worth naming. Participants in this study who were overconfident about pricing were using that confidence to narrow their shortlist. A participant who believes Travelers costs roughly \$4,800 per year and Geico costs \$5,000 is making a real elimination decision based on numbers that may not apply to them at all.

If your brand’s pricing is systematically misrepresented in that AI summary, either too high or too low, it is affecting consideration, not just comprehension.

That is a brand problem, not just a UX problem.

3. Retailers: price absence is a missed shortlist moment.

For laptop tasks, the dominant clarity problem was that price did not appear. Participants who wanted to check whether a model fit their budget had to leave AI Mode to do it.

Several did not bother. The participant who said “it’s not giving me prices, this is not helpful” is a shortlist-abandonment risk: a participant who cannot quickly confirm a price will often fall back on brand familiarity rather than making an informed comparison.

Brands that ensure their current retail pricing is accessible via structured data and Merchant Center feeds give themselves a better chance of surviving that budget check in AI Mode, rather than losing the participant to a competitor whose price is visible.

4. Price framing in your content affects AI output framing.

One of the most clearly observed dynamics in the insurance tasks was that the AI presented some insurers with dollar amounts and others with percentage discounts. That formatting difference directly drove participant choices.

Presenting a specific annual premium estimate (“most drivers in this profile pay between \$X and \$Y per year”) rather than only percentage savings (“save up to 15%”) gives the AI a concrete number to surface. Concrete numbers are what participants acted on.

5. Validate what AI Mode is currently surfacing about your pricing.

Most brands have not audited how AI Mode represents their pricing. A practical starting point: query your own category in AI Mode the way a participant would (“best car insurance for a family with a teen driver”) and read exactly what the AI says about your pricing and your competitors’.

Then run the same query with your structured data validated using Google’s Rich Results Test and your Merchant Center feed audited for pricing accuracy. The gap between what your product data says and what AI Mode surfaces is the gap you need to close.

12. How People Ask Shapes What They Accept

Keyword-style queries and natural-language prompts produce different downstream behaviors.

The hypothesis is that keyword users treat AI Mode as a verification tool, while prompt users delegate judgment and accept the model's output with fewer comparisons.

How a question is phrased becomes a structural driver of which brands enter the candidate set.

13. Prompting Posture Diverged Sharply Across Surfaces

Across all 149 AI Mode tasks, **83% used natural-language prompts**, full sentences asking the AI to assess, recommend, or decide. **Only 15% entered keyword-style queries**. In Search, the split was nearly reversed: **53% used keywords and 27% used prompts**.

This matters because prompting style shapes what happens next. When participants phrase their query as a natural-language request, the AI returns a complete list of recommendations. When they use keywords, they tend to treat the AI output as one input among several rather than as a decision. The interface rewards prompting with a more complete answer.[15]

Note: The participants who did Search tasks did so after two AI Mode tasks. So, their frame of mind may have biased them toward continuing with a prompt style.

Table 2.1 Query type by surface (all tasks)

Query type	AI Mode (n=149)	Search (n=36)
Natural-language prompt	83%	27%
Keyword-style	15%	53%
Mixed	2%	20%

Full study (n=185): 71% prompt, 23% keyword, 7% mixed.

Actionable Insights:

- Write product content for [natural-language retrieval](#). Use the way buyers describe needs, use cases, and constraints in full sentences, not just keyword fragments.
- Plan for more buyers to [search by asking questions](#), not typing keywords. That shifts more shortlist formation into the interface before a buyer ever reaches your site.
- Don't abandon keyword search. Some buyers still search by specs, model names, and direct comparisons, and those visits still have value.
- [Build pages for both behaviors](#). Broad, question-based queries need content AI can summarize cleanly, while comparison and spec-driven queries need pages that help buyers verify details fast.
- Review where your current traffic comes from and what kind of query drove it. Don't assume today's mix of search behavior will hold as AI Mode usage grows.

"It's nice sometimes that you can, like, talk to AI and not ask a question, and they can like read between the lines. Like if I Google this search or I Google exactly what I said, I don't think that as helpful things would pop up."

30-year-old, insurance task, AI Mode. Explicitly choosing natural language because the interface rewards it.

"It seems like AI is telling me it's the cheapest."

P672489, washer/dryer, AI Mode. Built an elaborate scenario across multiple follow-ups and still framed the final decision as AI authority.

On the Search side, keyword behavior was consistent across nearly every participant. A 28-year-old performing the laptop task went directly to Best Buy and applied hardware filters, 64GB RAM, 16-inch minimum screen, without delegating any part of the problem to the search environment.

14. Ranking First Is Not the Same as Being Believed

Rank position within an AI response does not fully determine which product or brand a participant trusts or chooses. Brand recognition, source credibility, and the AI's own framing language can each outweigh positional rank. This changes what optimization means in AI Mode.

We expected the “rank override” to be more common. However, it’s also true that most sections in AI Mode only have 2-5 listings. People mentioned what they found expert, authoritative, or trustworthy a LOT during our [study last spring](#).

It was one of the most important findings from watching 70 people as they did a desktop video of their screen and spoke out loud during search sessions. [16] Text and sentiment analysis of the transcripts revealed.

16. Trust Mechanisms Differed By Kind, Not Just Frequency

The most significant difference in trust mechanisms between surfaces was not how often participants trusted something; it was **what** they trusted. Multi-source convergence, building confidence by seeing consistent signals across several independent sources, **appeared in 37% of Search tasks and 4% of AI Mode tasks.**

The practical implication for brands: in Search, appearing consistently across multiple independent sources builds trust. In AI Mode, the way the AI describes a product, its framing, structure, and language, carries more weight. A brand can be chosen or passed over based on how the AI explains it, regardless of ranking.

"Travelers and USAA actually tell me how much, whereas State Farm and GEICO give percentages. Just knowing the exact amount makes me want to pick Travelers or USAA right off the bat."

insurance, AI Mode. Output format drove the decision, not rank and not prior brand knowledge.

17. Brand Absence in AI Mode Creates Structural Invisibility

AI Mode compresses the comparison space into a short, model-curated set of brands that favors established and familiar names.

For brands outside that set, the problem is not that buyers saw them and passed. In many cases, buyers never encountered them in a meaningful way. No participant exhibited behavior that would have surfaced a brand absent from the AI shortlist.

Participants also treated the AI’s formatting as a quality signal. One participant eliminated Erie partly because “there’s not even a link there,” while another said, “I’m already eager to believe these are good recommendations because they mention LG and Samsung, two brands I consider very reliable.” In both cases, the AI’s presentation shaped credibility before the participant ever reached an external source.

That creates a structural disadvantage for challenger, regional, and direct-to-consumer brands. When Erie Insurance appeared in AI Mode results, multiple participants still dismissed it based on name recognition alone, not on price, coverage, or ratings. Presence matters, but recognition still shapes whether presence turns into consideration

XOFU Visibility Aligned With Participant Brand Selection

[XOFU’s visibility data](#) points in the same direction.

Comparing the brands participants chose most often with the brands that had the strongest XOFU visibility scores showed strong alignment in television and washer/dryer tasks (RBO = 0.812 and 0.853) and moderate alignment with the laptop task (RBO = 0.754).

Table 1.9 Brand Rank in AI Mode vs Xofu Visibility Tracking

Brand	Rank (our results)	Count (our results)	Rank (Xofu results)
LG	1	32	1
Samsung	2	13	2
Electrolux	3	10	6
Whirlpool	4	8	4
Amana	5	1	NA
GE	5	1	3

In plain terms, across most categories, the brands participants chose were also the ones already most visible in the recommendation environment.

Driver insurance was the exception (RBO = 0.236). That weaker alignment likely reflects the category itself: participant choices included regional insurers and comparison-style providers that did not always appear in XOFU’s ranked visibility set, which limited direct comparability.

Even with that exception, the broader pattern holds: AI Mode narrowed decisions around a smaller, more visible set of brands.

Brand Absence Actionable Insights

1. If your brand is not in the AI's output, it's invisible

Start by checking whether your brand appears for the [BOFU prompts](#) that matter in your category.

Look at which brands show up, how they are ordered, and how they are described, because that shortlist now shapes who gets considered at all.

2. How the AI presents your brand matters as much as whether it appears

Work on the inputs that shape AI's description of your brand: product data, pricing clarity, use-case language, review coverage, and crawlable third-party mentions.

Presence helps, but weak framing can still cost you the decision.

3. Challenger brands need recognition before the AI moment

If buyers do not know your brand, showing up once in AI Mode may not be enough.

Build the editorial, review, PR, and awareness signals that make your brand easier to recognize and easier for Google to treat as part of the credible set.

18. Participants Were Not Significantly Frustrated by the Narrower Option Set

Narrowness frustration, in which participants explicitly signaled that the option set was too limited, occurred in **15% of AI Mode tasks and 11% of Search tasks**. The difference was not statistically meaningful. Despite accepting the AI's shortlist at dramatically higher rates, AI Mode participants were not more likely to push back on the size of the candidate set.

This distinction matters for interpreting the adoption rate. If high adoption were driven by participants feeling they had no choice, we would expect a materially higher frustration rate. We did not see that pattern.

The relationship between frustration and adoption is more complicated than the aggregate suggests. Frustration and adoption co-occurred in AI Mode in ways they did not in Search.

"The AI information is kind of just pointing to teen drivers, teen drivers, teen drivers — even though I'm asking specifically for a family with teen driver. See, it's a different thing, that's why it's confusing."

P69610585, insurance, AI Mode. The most skeptical AI Mode participant in the comparison set. Still chose Geico and Travelers — the consensus AI result.

"See, it's gonna want tons of information and we don't like that."

P6969539, insurance, Search. Abandoned NerdWallet's quote tool and pivoted to Forbes. Frustration in Search produced a different destination.

"This is where without a lot of additional information, the process can be a little bit — not overwhelming, but there's too many options."

P6991dd6, 40 years old, laptop, Search. Responded by introducing constraints and filters rather than accepting a narrower set.

19. Participants Delegated Without Feeling Constrained

The combination of near-total shortlist adoption in AI Mode and a narrowness frustration rate statistically indistinguishable from Search is one of the more practically significant findings in this dataset.

Having fewer surfaced choices did not produce materially higher narrowness frustration than Search, at least not at rates we could measure.. Whether this reflects genuine satisfaction with the AI's output, reduced awareness of the set of options, or the cognitive ease of the AI format cannot be determined from behavioral data alone.

20. Task Differences Worth Tracking in the Full Dataset

Insurance and washer/dryer tasks showed the clearest behavioral contrast within AI Mode. Insurance participants delegated most heavily: 85% clicked nothing, and AI framing was the dominant trust signal.

Washer/dryer participants were the most likely to verify or extend beyond the AI's output, with the highest proportion of AI VERIFIED and USER BUILT outcomes.

Source trust was highest for washer/dryer, possibly because appliance decisions involve long-term ownership risk that participants felt required external validation.

Conclusion: AI Mode favors consensus over discovery

This study shows that AI Mode pushed buyer shortlisting away from multi-source comparison and toward earlier acceptance.

In Search, people more often built a shortlist by clicking, cross-checking, and comparing across multiple sources. In AI Mode, they more often accepted the shortlist Google generated, refined it inside the interface, and left mainly to verify a choice already in progress.

That changes what visibility means.

Brands missing from the AI output often did not enter consideration, while brands that did appear benefited from the AI's framing, familiar names, and a narrower field of candidates. The selection mechanism moved upstream toward Google's recommendation layer.

For marketers, that means the work no longer starts with earning the click.

It starts earlier: [being present in the shortlist](#), being framed clearly, and being reinforced by the sources and signals AI pulls into its synthesis. Search still leaves more room for discovery. AI Mode narrows that process earlier.

AI Mode increasingly governs who gets discovered in the first place.

If your brand is missing from AI shortlists, [Xofu](#) helps you [track BOFU visibility](#). This lets you see which competitors appear and understand how AI systems frame the choice.

If you need help closing those gaps, [Citation Labs](#) turns the findings into citation optimization and focused campaigns designed to improve AI visibility and measure impact over time.