Rezonait

REZONAIT RESEARCH REPORT · SPRING 2026

State of AI Visibility: Sustainable Fashion

AI visibility, accuracy, and hallucination analysis across 17 sustainable fashion brands on OpenAI ChatGPT and Google Gemini.

Published
May 13, 2026

rezonait.com
AI visibility for mission-driven brands

EXECUTIVE SUMMARY

The findings, in brief

Consumers are shifting to AI-powered product discovery at a pace that has surprised even the researchers tracking it. When consumers ask an AI assistant to recommend a sustainable fashion brand, the answer they receive depends not only on what the brand has published about itself but on whichever AI platform the consumer uses and on whether that platform can accurately identify and reproduce complex sustainability claims.

This report presents findings from our audit of 17 sustainable fashion brands across the two most popular frontier LLMs used by consumers: OpenAI's GPT-4o and Google's Gemini 3 Flash. Each brand was tested using a layered query methodology that simulates how real consumers discover, evaluate, and compare brands through AI-powered conversations, and how those conversations affect conversion. The results reveal significant gaps in how AI represents the brands, many of which depend heavily on accurate, nuanced storytelling.

Summary of key findings

The platform your customer uses changes your brand's story. OpenAI's GPT-4o achieved a mean accuracy score of 0.889 across all brands tested. Google's Gemini scored 0.653 using the same prompts. That 24-point gap means a consumer using one platform gets a materially different, and often materially less accurate, representation of the same brand.

Gemini fabricates brand facts and claims at 3.6 times the rate of OpenAI. Across the 17 brands in the study, Gemini produced 585 unique hallucinations to OpenAI's 163. These are not vague inaccuracies. They include fabricated certifications, invented program names, and false claims about supply chain practices. Girlfriend Collective, for example, received zero hallucinations from OpenAI and 20 from Gemini.

Brand size and digital footprint predict AI visibility far more than brand quality. Tier 1 benchmark brands averaged 0.675 visibility on OpenAI. Tier 2 mid-market brands (the $5M–50M range most representative of mission-driven sustainable brands) averaged 0.396. Emerging and niche brands averaged 0.049. The Tier 3 emerging/niche brands, those with the most to gain from accurate AI representation, are the least likely to appear.

Brands disappear during the purchase conversation. When AI interactions move from initial discovery to a multi-turn purchase journey, some brands that appeared in early recommendations dropped out entirely. Others never entered the conversation at all. On Gemini, 6 of the 17 brands never appeared in the first turn of a simulated purchase journey.

CONTEXT

The shift to AI-powered discovery

The shift from traditional search to AI-powered product discovery has been building for two years. But in the span of a few months, from late 2025 through early 2026, established analysts and research outlets released findings that, taken together, suggest this is no longer an emerging trend. It is a significant, long-term shift in consumer behavior.

The scale of what is moving

Adobe, which tracks over a trillion visits to U.S. retail sites, measured a 4,700% year-over-year increase in traffic to retail sites from generative AI sources by mid-2025.[1] Gartner, approaching the rise in AI search from the demand side, predicted that traditional search engine volume would drop 25% by 2026 as consumers shift to AI chatbots and virtual agents.[2] Their mid-2025 consumer survey confirmed the movement: 51% of U.S. consumers reported that their research habits had already changed because of generative AI.

From browsing to buying

Early skeptics made a reasonable argument: people might browse through ChatGPT, but they would still buy through traditional channels. While that is partially true, Adobe's holiday season data showed that the 49% gap in conversion rate between non-AI and AI sources dwindled down to 23% by July 2025. Furthermore, of those that used AI for their shopping, 85% noted an improved shopping experience.[1]

Salesforce, drawing on data from over 1.5 billion shoppers globally, found that AI influenced 20% of all retail sales during the 2025 holiday season.[3] Shoppers referred through AI-powered search channels converted nine times more often than those arriving from social media.

Boston Consulting Group, in a January 2026 report surveying more than 9,000 consumers across nine countries, found that shopping-related generative AI use grew 35% in under a year, with more than 60% of consumers expressing high trust in AI-generated recommendations.[4]

The McKinsey and Business of Fashion *State of Fashion 2026* report put a specific face on these numbers: between June and August 2025, ChatGPT accounted for 16% of Zara's inbound website traffic.[5] That is a major global retailer seeing a significant share of its product discovery happening through a channel that barely existed eighteen months earlier.

What this means for brands that rely on trust

The research paints a clear picture for consumer brands broadly. But for brands that compete on certifications, supply chain transparency, and nuanced sustainability claims, this shift carries a specific risk that the broader research does not address.

When a consumer asks an AI assistant to recommend sustainable running shoes, the model synthesizes information from across the web and delivers a curated answer. It does not present ten blue links for the consumer to evaluate. It presents a conclusion. And the brands that appear in that conclusion, and how they are described, depend entirely on what the AI can find, interpret, and verify about them.

A brand with a simple value proposition is relatively easy for an AI to summarize accurately. A brand with a complex story, involving B-Corp certification, regenerative materials sourcing, and fair labor standards across a multi-tier supply chain, presents a much harder challenge. The more nuanced the differentiation, the more likely it is to be flattened or omitted entirely when an AI constructs its answer.

The following findings quantify exactly how that flattening plays out across 17 sustainable fashion brands and two major AI platforms.

KEY FINDINGS

What the audit revealed

Brand size predicts AI visibility far more than brand quality

The brands in this study were organized into three tiers by market presence. Tier 1 included three established benchmarks with extensive digital footprints. Tier 2 comprised ten mid-market brands with smaller digital footprints. Tier 3 included four emerging or niche brands with minimal digital footprints.

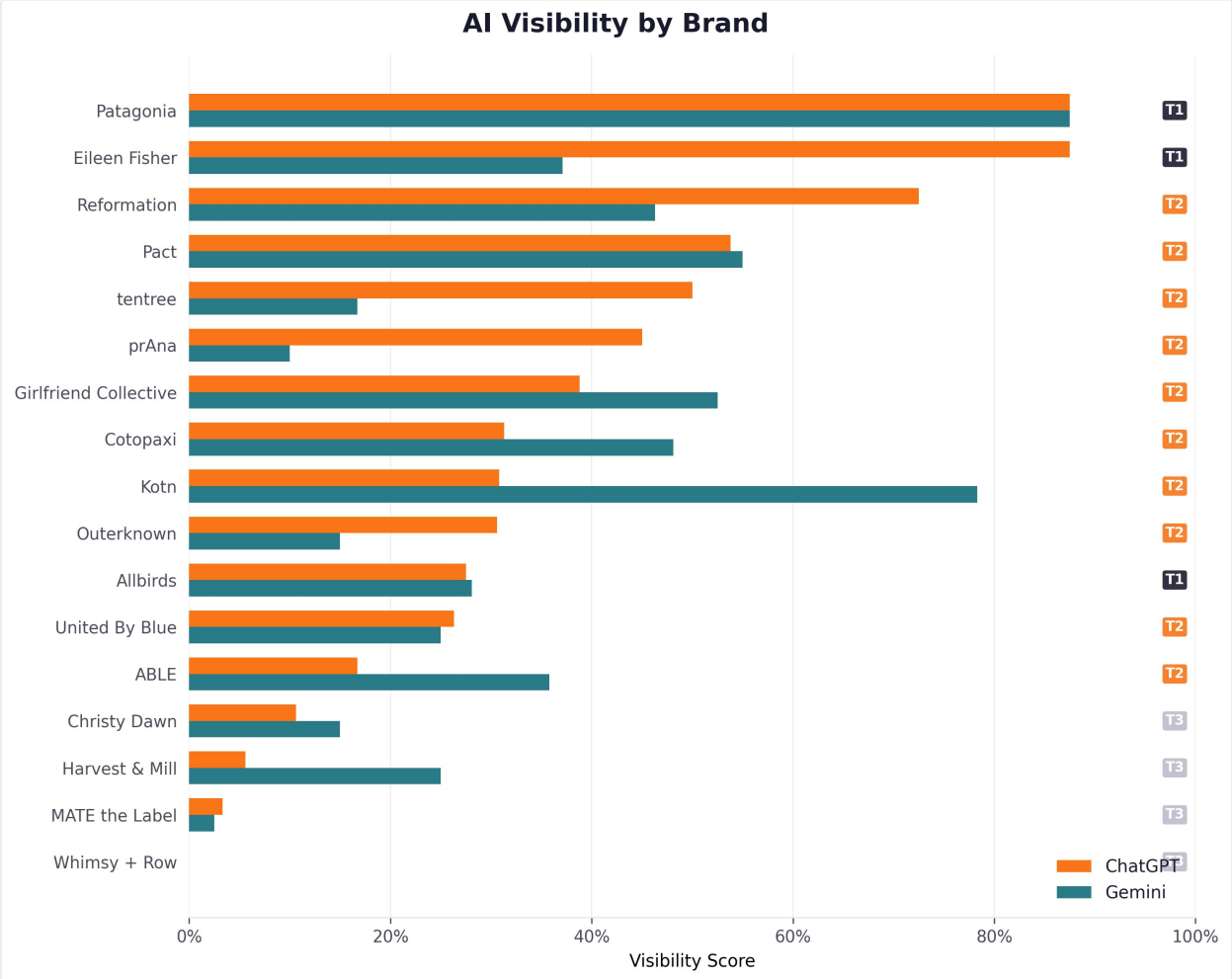
Tier 1 — Benchmark brands	Tier 2 — Mid-market brands	Tier 3 — Emerging/niche brands
Allbirds*	ABLE	Christy Dawn
Eileen Fisher	Cotopaxi	Harvest & Mill
Patagonia	Girlfriend Collective	MATE the Label
	Kotn	Whimsy + Row
	Outerknown	
	Pact	
	prAna	
	Reformation	
	tentree	
	United By Blue	

*** NOTE ON ALLBIRDS** On April 15, 2026, shortly before this report's publication, Allbirds announced it would sell its footwear brand and rebrand as "NewBird AI," an AI compute infrastructure company. The brand that once defined sustainable shoes, particularly in the startup community, is exiting the industry entirely. We have retained Allbirds in this analysis as a Tier 1 benchmark because AI models will continue to reference and recommend the brand long after the company has moved on, which is itself a useful illustration of how slowly AI knowledge updates relative to market reality.

On OpenAI, Tier 1 brands averaged 0.675 visibility, Tier 2 averaged 0.396, and Tier 3 averaged 0.049. The drop from Tier 2 to Tier 3 is particularly stark: it represents a near-complete disappearance from AI-powered discovery. Patagonia and Eileen Fisher each scored 0.875

visibility on OpenAI. Whimsy + Row scored 0.000 on both platforms, appearing only when asked about by name. MATE the Label scored 0.033 on OpenAI and 0.025 on Gemini. Harvest & Mill scored 0.056 on OpenAI.

The mid-market tier tells the most relevant story for the brands this research is designed to serve. Within Tier 2, visibility ranged from 0.725 (Reformation) to 0.167 (ABLE) on OpenAI, and from 0.783 (Kotn) to 0.100 (prAna) on Gemini. The spread within this tier is wide enough that two brands with similar revenue, similar certifications, and similar sustainability commitments can have entirely different levels of AI representation.

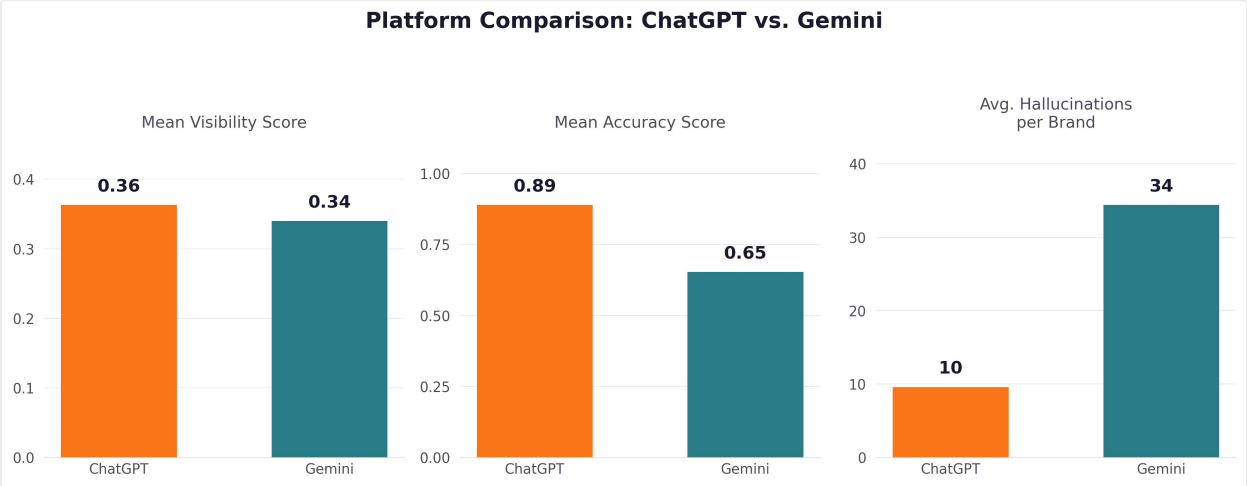


Several factors likely contribute. Larger brands generate vastly more web content — press coverage, reviews, Reddit threads, Wikipedia entries — which means AI models have simply encountered them more often during training. Media coverage compounds this: mainstream publications that are well-represented in training data cover Tier 1 brands regularly, while smaller brands may only appear in niche outlets. Perhaps most consequentially for this

audience, the detailed certifications and supply chain documentation that differentiate smaller sustainable brands often live in PDF reports and certification databases that AI models process poorly, if at all.

Which AI your customer uses materially changes your brand's story

Across all 17 brands tested, OpenAI's GPT-4o achieved a mean accuracy score of 0.889 on a 0-to-1 scale. Google's Gemini scored 0.653. Every single brand in the sample received a higher accuracy score on OpenAI than on Gemini.



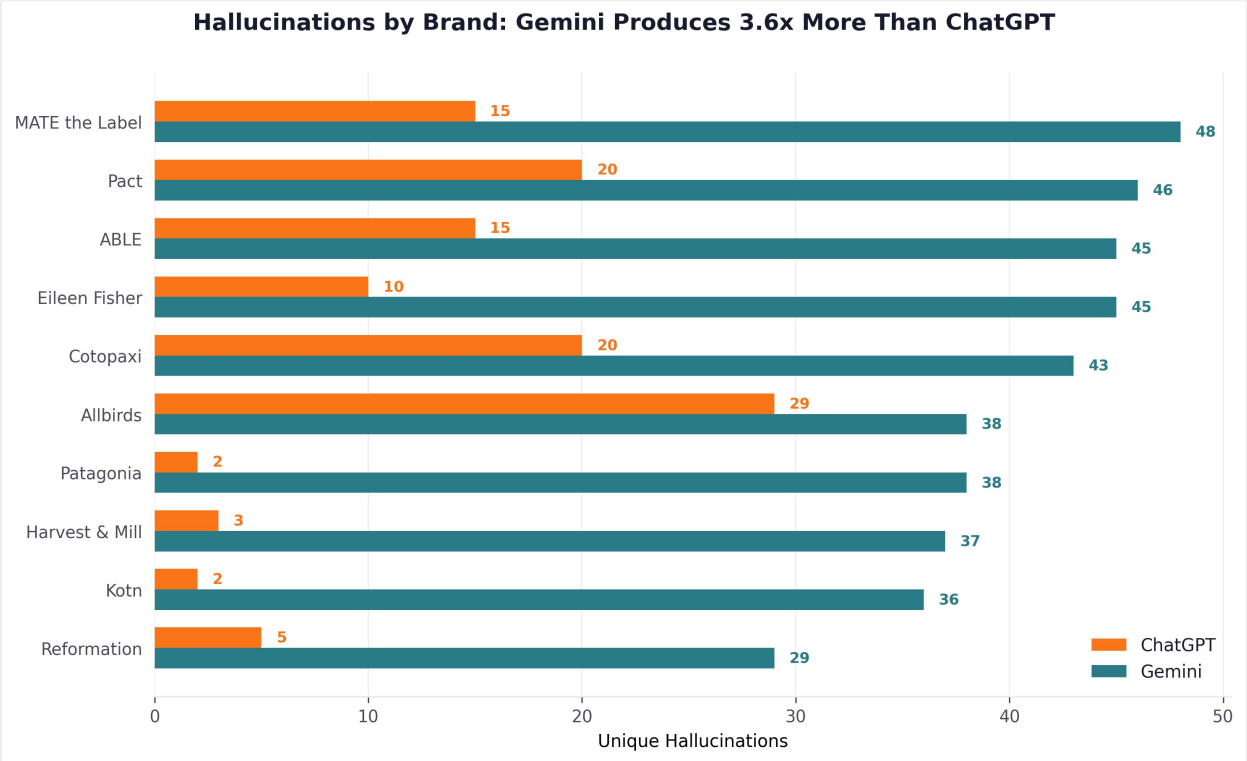
Eight brands scored above 0.90 accuracy on OpenAI, meaning the AI's claims about their certifications, materials, and supply chain practices were almost entirely correct. These included Girlfriend Collective (1.000), Kotn (0.993), Patagonia (0.987), Harvest & Mill (0.958), Reformation (0.947), Eileen Fisher (0.942), tentree (0.939), and Whimsy + Row (0.950). On Gemini, no brand exceeded 0.801.

The gap is not uniform. Harvest & Mill showed the largest accuracy divergence in the sample: 0.958 on OpenAI versus 0.567 on Gemini, a spread of nearly 0.39 points. MATE the Label recorded the lowest Gemini accuracy of any brand at 0.329, meaning that more than two-thirds of the AI's claims about the brand were inaccurate or fabricated. For a brand whose identity depends on specific certifications and sourcing practices, this level of misrepresentation is not a minor inconvenience. It is a clear risk for marketing and brand leaders.

Metric	OpenAI (GPT-4o)	Gemini 3 Flash	Gap
Mean Accuracy	0.889	0.653	0.236
Median Accuracy	0.885	0.670	0.215
Brands ≥ 0.90	8 of 17	0 of 17	—
Brands < 0.50	0 of 17	1 of 17	—

Gemini fabricates brand claims at 3.6 times the rate of OpenAI

Hallucinations in this context are specific, verifiable false claims: a certification the brand does not hold, a program that does not exist, a supply chain practice the brand has never adopted. Across the 17-brand sample, Gemini produced 585 unique hallucinations to OpenAI's 163, a ratio of 3.6 to 1.



The nature of these fabrications matters as much as the volume. Gemini consistently invented specific, authoritative-sounding details that varied across iterations. For Cotopaxi, it fabricated an elaborate narrative about Climate Neutral certification that the brand does not

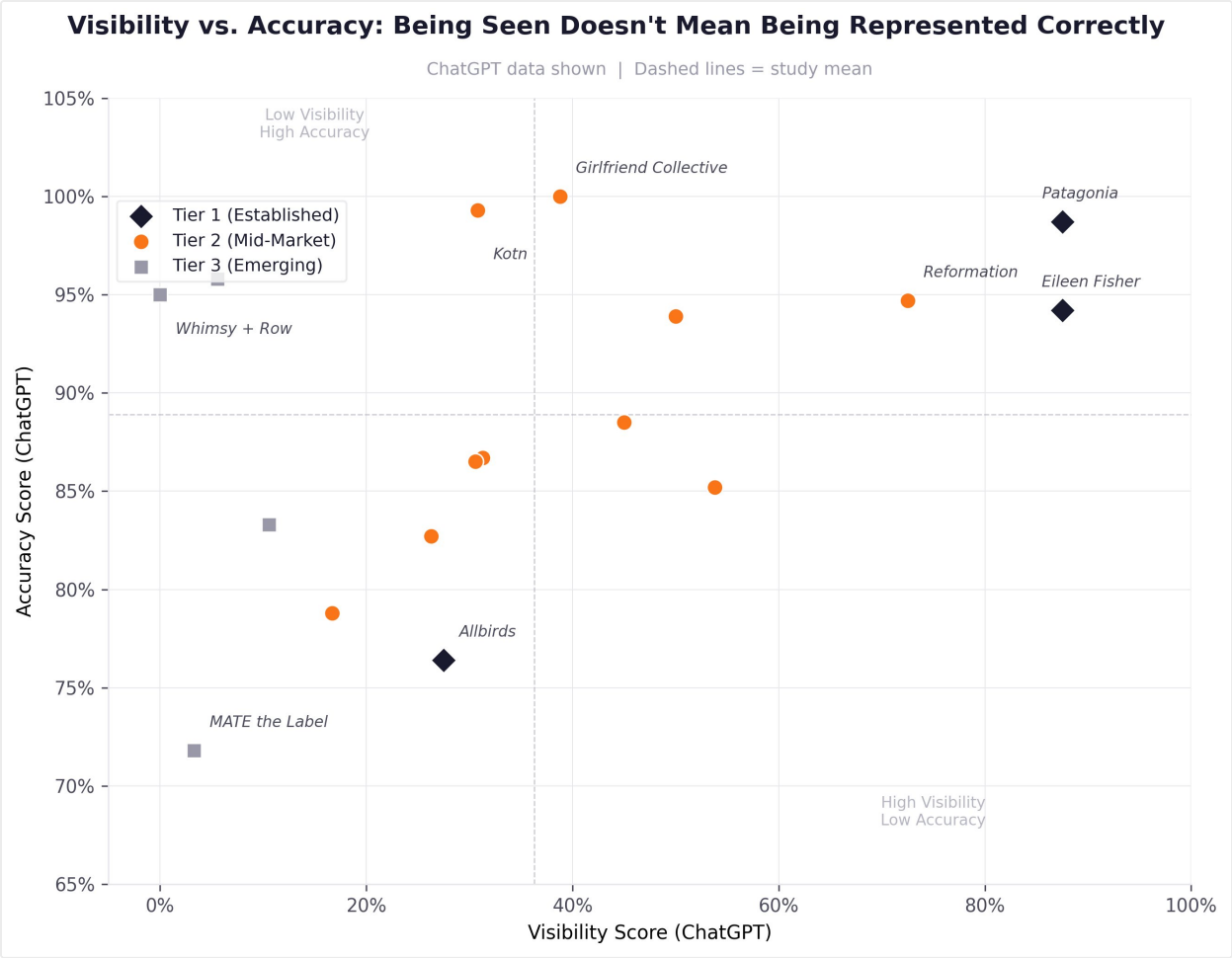
hold. For United By Blue, it invented corporate acquisition details, complete with fabricated parent company names. For Kotn, it attributed GOTS certification and B-Corp status that the brand has never claimed. These are not vague overstatements. The AI presents these as confident assertions of facts that do not exist.

The contrast with OpenAI is sharpest at the extremes. Girlfriend Collective received zero hallucinations from OpenAI across all queries and iterations, while Gemini produced 20. Patagonia, the best-documented brand in the sample, received only 2 hallucinations from OpenAI but 38 from Gemini. Having extensive public documentation did not protect brands from fabrication on Gemini; it only appeared to help on OpenAI.

Across the 17 brands in this study, hallucinations followed recognizable patterns rather than appearing randomly. The most common was certification stacking: AI models attributed certifications held by other brands in the sustainable fashion space to brands that don't hold them, as though drawing from a shared pool of industry credentials. The second was what we call *superlative inflation* — taking a claim that is partially true and presenting it as absolute. A brand that uses organic cotton in most of its product line becomes a brand that uses "100% organic cotton across all products." The third was staleness: presenting outdated information as current, such as citing certifications a brand once held but has since let lapse. These were not random errors. They were systematic patterns that any brand in this space should expect to encounter.

Being seen and being correctly described are different problems

An intuitive assumption might be that highly visible brands are also accurately described, and that invisible brands are simply unknown to AI. The data does not support this.

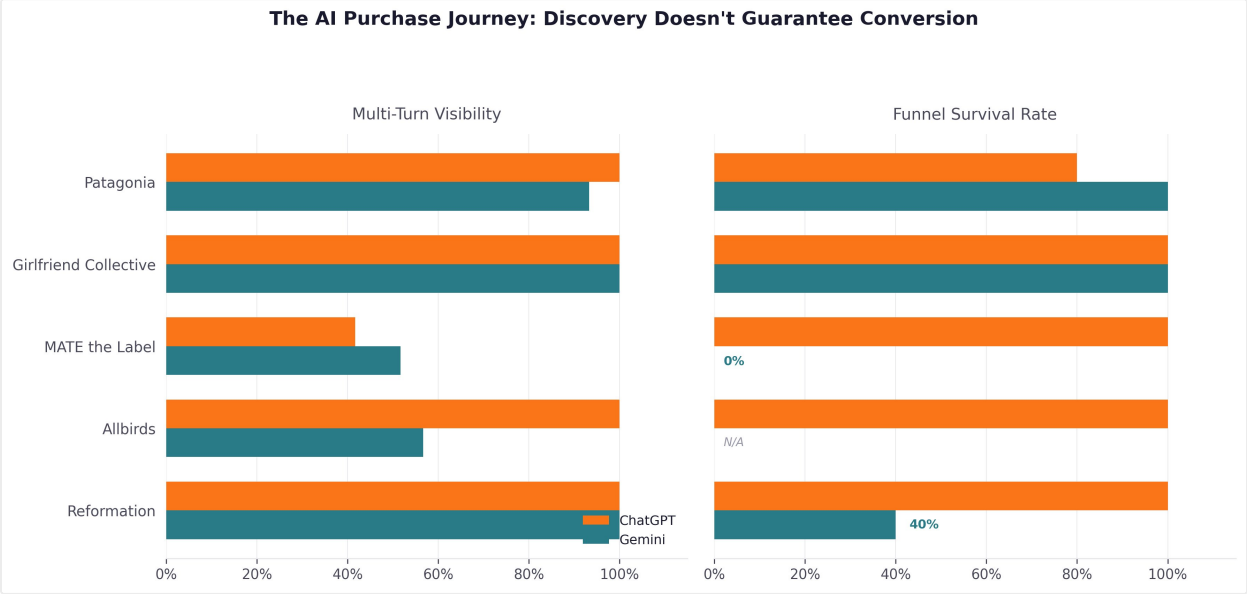


Allbirds is a Tier 1 brand with broad consumer recognition, yet it recorded the highest hallucination count of any brand on OpenAI (29) and relatively low visibility (0.275). It is well-known but frequently misrepresented. Kotn, a Tier 2 brand, showed the largest visibility gap between platforms: 0.783 on Gemini versus 0.308 on OpenAI. Gemini sees Kotn far more often, but when it does, its accuracy drops to 0.724 compared to OpenAI's 0.993. Kotn is more visible on the less accurate platform and less visible on the more accurate one.

This decoupling of visibility and accuracy means that a brand cannot assume its AI representation is sound simply because it appears in recommendations. Visibility without accuracy is, in some respects, worse than invisibility: it means consumers are receiving confident but incorrect information about the brand's practices.

Brands disappear during the purchase conversation

The audit tested each brand’s persistence across a simulated multi-turn purchase journey: a three-turn conversation that begins with a broad category question, narrows to specific differentiators, and ends with a direct brand comparison. This structure mirrors how consumers actually use conversational AI, moving from exploration to evaluation to decision.



The results expose a dimension of AI visibility that single-query testing cannot capture. On OpenAI, three brands (ABLE, United By Blue, and Whimsy + Row) never appeared in the first turn of the purchase journey. On Gemini, six brands never entered the funnel: Outerknown, prAna, United By Blue, Harvest & Mill, Whimsy + Row, and Allbirds.

Three brands illustrate the range of outcomes. Girlfriend Collective maintained perfect funnel survival on both platforms: it appeared in Turn 1 and persisted through the comparison stage. United By Blue appeared on neither platform’s first turn, meaning that a consumer following a natural purchase conversation would never encounter the brand unless they already knew to ask for it by name. Whimsy + Row registered zero organic visibility on both platforms across every query type, including the multi-turn journey. It is, for AI purposes, a brand that does not exist unless the consumer already knows it exists.

AI is leaving out significant parts of every brand’s story

Completeness measures how many of a brand’s key differentiators the AI captures when it does mention the brand. Across the entire sample, completeness scores ranged from 0.197 to 0.691. No brand on either platform exceeded 0.70. The overall mean was 0.439 on OpenAI and 0.441 on Gemini.

This means that even when AI mentions a brand accurately, it typically conveys fewer than half of the characteristics that make that brand distinct. For brands that have invested in specific certifications, unique supply chain structures, or differentiated business models, the AI is reducing them to one or two generic descriptors. Pact's per-product carbon footprint data, Christy Dawn's farm-to-closet regenerative cotton model, Eileen Fisher's circular design program: these are the differentiators that justify a premium and earn consumer trust, and AI is consistently omitting them.

This finding is structural, not brand-specific. It affects Tier 1 and Tier 3 brands alike. Patagonia, the most visible and most accurately described brand in the sample, still scored only 0.466 on completeness (OpenAI). The AI knows Patagonia exists, describes it correctly, but tells less than half its story.

Even brands that scored well on visibility and accuracy rarely exceeded 50% completeness. AI models tend to reach for category-level descriptors — "uses organic materials," "ethically made," "eco-friendly" — rather than surfacing the specific differentiators that distinguish one brand from another. The result is a flattening effect in which brands that have invested in building genuinely differentiated sustainability programs are described in the same generic language as brands that have not.

A brand's discoverability depends on the question being asked

Each brand was tested across four layers of progressively specific queries: broad discovery ("best sustainable fashion brands"), subcategory ("best brands for organic cotton basics"), niche (queries targeting specific differentiators without naming the brand), and verification (asking about the brand by name). The first layer at which a brand appears is its discovery depth.

Nine of 17 brands appeared at the broadest query level on OpenAI. On Gemini, eight did. But these aggregate numbers mask significant platform divergence at the brand level. tentree, for example, surfaces at the broadest level on OpenAI but requires a niche-level query to appear on Gemini. prAna shows the opposite pattern. This means that a brand's AI discoverability is not a fixed property; it shifts depending on which platform the consumer uses and how they phrase their question.

An emerging pattern in the data, though based on small cluster samples, suggests that outdoor and adventure brands enjoy the highest combination of visibility and accuracy across both platforms, while fair trade and ethical labor brands underperform relative to their certification strength. The categories where sustainability claims are most complex appear to be the ones where AI struggles most to represent them.

METHODOLOGY

How the audit was conducted

What we tested

Seventeen sustainable fashion brands spanning three tiers: three established benchmarks, ten mid-market brands, and four emerging or niche brands. Each brand was tested on two AI platforms: OpenAI's GPT-4o and Google's Gemini. Web search and retrieval augmented generation were disabled on both platforms to test the models' embedded knowledge, not their ability to search the internet in real time. The audits were conducted using publicly available information on the brands' websites as well as third-party sources.

How we tested

Each brand was audited using a layered query methodology designed to simulate how consumers actually discover products through conversational AI. Queries progressed from broad ("What are the best sustainable fashion brands?") through subcategory-specific ("What brands make clothing from recycled materials?") to niche queries targeting each brand's specific differentiators without naming the brand directly. Verification queries then asked about each brand by name. A multi-turn purchase journey simulation tested whether brands persisted across a three-turn conversation that moved from discovery to evaluation to comparison.

Each query was run five times per platform to account for the stochastic nature of AI responses, producing a distribution of results rather than a single data point. Brands with dual positioning (for example, a fair trade brand with a secondary outdoor identity) received additional queries to test visibility across both contexts.

How responses were scored

Every brand in the sample was profiled with a ground truth fact sheet, compiled from the brand's own website, most recent sustainability reports, and independent third-party sources. Each fact sheet documented verified certifications, core sustainability claims, claims the brand does not make (for hallucination detection), and the brand's key differentiators (for completeness scoring).

AI responses were scored on three dimensions. Visibility measured whether the brand appeared at all and how prominently. Accuracy compared every factual claim in the AI's response against the verified fact sheet, with specific attention to fabricated certifications and false program claims. Completeness measured how many of the brand's verified differentiators the AI captured.

To prevent scoring bias, a cross-model judging protocol was used: OpenAI responses were scored by Gemini, and Gemini responses were scored by OpenAI. This ensures that neither platform grades its own work.

How we verified our results

The combination of layered queries, multiple iterations, ground truth fact sheets, and cross-model judging produces findings that are specific, reproducible, and grounded in verified data. The methodology was developed through a pilot audit and refined across all 17 brands before cross-brand aggregation. Every hallucination cited in this report is a specific, named claim that was checked against the brand's own published materials and independent third-party sources.

IMPLICATIONS

What this means for mission-driven brands

The data in this report describes a gap between what sustainable brands have built and what AI platforms communicate about them. That gap is not theoretical. It is quantified, brand by brand, across two platforms that collectively serve hundreds of millions of consumers.

For the benchmark and mid-market brands, this presents a specific strategic challenge. These brands have invested in certifications, supply chain transparency, and differentiated sustainability practices precisely because their customers demand proof. Their competitive advantage is specificity: not just “sustainable,” but B-Corp certified, Fair Trade across specific product lines, regenerative cotton sourced from named partners, carbon footprint data published at the individual product level. That specificity is what AI models struggle most to capture.

The accuracy gap is a trust gap

A 24-point accuracy difference between platforms is not an abstract measurement. It means that when a consumer using Gemini asks about a brand’s sustainability practices, there is a meaningful probability that the response will include fabricated certifications, invented programs, or outdated claims presented as current. For brands whose entire value proposition depends on consumer trust in their claims, this is a direct threat to the relationship that justifies their pricing.

Visibility is a prerequisite, not an outcome

The visibility cliff between established and mid-market brands suggests that most sustainable brands are not yet part of the AI-powered discovery conversation. If 52.9% of brands appear at the broadest query level on OpenAI, that means 47.1% do not. At the subcategory level, where purchase intent is higher, the numbers improve but remain uneven. For the brands that do not appear, the challenge is not whether their AI representation is accurate. The challenge is that they have no AI representation at all.

The completeness problem is structural

No brand in the sample, regardless of tier, size, or documentation quality, achieved a completeness score above 0.70. AI is telling a simplified version of every brand's story, even the brands it knows best. This is not a problem that individual brands have created, but it is a problem that individual brands bear the consequences of. The certifications, sourcing practices, and impact data that differentiate a brand from its competitors are the elements most likely to be omitted from an AI-generated recommendation.

The consumer shift to AI-powered discovery is no longer speculative. The research cited in this report, from BCG, Adobe, Gartner, McKinsey, and Salesforce, establishes that the channel is real, growing, and increasingly decisive. For brands built on trust, the question is whether their story is structured for the way consumers are now finding them.

REFERENCES

Endnotes

1. Adobe Digital Insights — *Generative AI-powered shopping rises with traffic to U.S. retail sites up 4,700%.* (August 2025)
2. Gartner — *Gartner Survey Finds Only One-Third of Consumers Say GenAI Rivals Search Engines; Marketers Must Optimize for Both AI-Driven and Traditional Search.* (January 2026)
3. Salesforce — *Holiday Season Rakes in Record $1.29T for Retailers, Salesforce Data Shows.* (January 2026)
4. Boston Consulting Group — *Consumers Trust AI to Buy Better. Brands Need to Move Quickly.* (January 2026)
5. McKinsey and Business of Fashion — *AI's Transformation of Online Shopping Is Just Getting Started.* (December 2025)

This report was produced by Rezonait (rezonait.com), an AI visibility consultancy for mission-driven brands. Methodology details and brand-level data are available upon request.