

BRAND ACCURACY · BRAND MESSAGING · AI MISINFORMATION · MONITORING · HALLUCINATION

When AI gets your brand wrong, and why it's hard to fix

AI systems describe your brand based on stale training data and third-party sources — not your website. Most brands don't know what's wrong, and there's no official correction portal. Here's what to do.

AIVIARA RESEARCH · MAY 2026

B2B buyers increasingly research vendors in ChatGPT, Perplexity, Google AI Overviews, and Gemini before talking to anyone on a sales team. Bain & Company found that ChatGPT usage for shopping-related queries grew 25% between January and June 2025, on top of roughly 70% growth in overall usage. For many buyers, the first detailed description of your company now comes from an AI system rather than your website. What it says about you is often inaccurate.

Metricus, an AI brand monitoring company, reports from its internal audit data that 72% of brands have at least one factual error in AI-generated responses, averaging 3.4 errors per brand. Those figures come from a vendor selling correction services; treat them as directional rather than definitive. The independent evidence is harder to set aside. Columbia University's Tow Center for Digital Journalism tested eight AI search engines across 1,600 queries and found they "failed to retrieve the correct information more than 60% of the time." In that study, ChatGPT used qualifying language on only 15 of its 134 incorrect responses.

72% of brands have at least one factual error in AI-generated responses, averaging 3.4 errors per brand.

METRICUS INTERNAL AUDIT DATA

Most brands track whether they appear in AI responses. Fewer are checking whether what AI says about them is accurate.

Three mechanisms behind most AI misinformation about brands

AI models train on data compiled months before deployment. Dana Davis at RankScience estimates typical lag at 6 to 12 months, with search-heavy platforms refreshing more frequently and some general models carrying brand information 18 months or more out of date. A 2022 study in Nature Scientific Reports (Vela et al.) found 91% of deployed machine learning models showed temporal degradation as time passed from their training date.

RankScience documented what that gap looks like. After discontinuing several services in 2024, including its A/B testing software, ContentEdge AI copywriting tool, and PPC offerings, the company found ChatGPT recommending it for "A/B testing" and "SEO-PPC Channel Alignment," Gemini listing "Data-Driven SEO Experimentation" as a top service driver, and Perplexity ranking "Integrated SEO and PPC Engine" as their #2 service. Dana Davis, reporting in January 2026, noted all of those services had ceased to exist.

A brand's own website is one signal in the training corpus, not a privileged one. Sight AI, an AI brand monitoring service, observes that "AI systems often give more weight to content from established domains." Errors on high-authority review platforms and comparison sites can dominate over accurate content on a brand's own site. ZipTie.dev describes the practical version of this: "If twelve outdated comparison articles mention wrong pricing and only your website shows the correct number, AI sees the wrong information more often."

Beyond stale data and conflicting signals, AI systems can confabulate, generating wrong answers that don't trace to any specific wrong source. Anqi Shao at the University of Wisconsin-Madison, writing in the Harvard Kennedy School Misinformation Review in August 2025, identifies this as a distinct production logic. Hallucinations, Shao writes, "emerge from human-machine interactions without deliberate deception" because training data "often contain biases, omissions, or inconsistencies." The result is a response that sounds authoritative and is occasionally entirely fabricated.

In practice, most brand errors are not dramatic hallucinations. They're usually older pricing or discontinued services, surviving in training data long after the business moved on.

Hallucination is not deception

Framing AI brand errors as deception misses how these systems work.

What looks like deception is usually pattern completion. These systems are not attempting to mislead anyone; they're extending statistical patterns from incomplete or stale information. Anqi Shao's August 2025 paper in the Harvard Kennedy School Misinformation Review treats hallucination as a structurally different phenomenon from human misinformation for exactly this reason.

Legal consequences don't require intent. Air Canada's own chatbot told customer Jake Moffatt he could apply for a discounted bereavement fare within 90 days of purchasing a ticket. No such retroactive policy existed. The British Columbia Civil Resolution Tribunal ruled in February 2024 (Moffatt v. Air Canada, 2024 BCCRT 149) that Air Canada was liable for negligent misrepresentation. Tribunal Member Christopher Rivers wrote that the company "did not take reasonable care to ensure its chatbot was accurate" and "still bore responsibility for all the information on its website, whether it came from a static page or a chatbot." Air Canada was ordered to pay \$812.02. Intent played no part in the ruling.

Monitoring tools track visibility, not accuracy

A category of AI monitoring tools has emerged to handle the tracking work. These tools track where brands appear in AI responses across platforms and are useful for visibility measurement: does the brand appear, how often, in what position. Factual accuracy is a different question, and largely an absent one.

Errors can propagate further through AI-generated content. GPTZero's investigation in June 2024 found that "the average user only needs three calls to Perplexity before encountering an AI-generated source." Wrong descriptions spread through AI-to-AI retrieval without tracing back to any original authored source.

The average user needs just **three calls** to Perplexity before encountering an AI-generated source.

GPTZERO INVESTIGATION, JUNE 2024

The most reliable starting point is a manual baseline. Run the questions your buyers actually ask about your category, covering each across ChatGPT, Perplexity, Google AI Overviews, and Gemini. PageCrawl.io recommends starting with your top 10 brand-related queries. Record the responses verbatim. Score accuracy on a 1-5 scale. The process requires no tooling and surfaces the specific error types affecting your brand. Monitoring tools become useful once you have something to measure against.

Correction is indirect and slow

There is no official correction portal at OpenAI, Google, or Anthropic for brand misinformation. Metricus states this explicitly, and ZipTie.dev confirms it. The structural reason is that AI models generate responses from training data patterns rather than retrieving facts from an editable database. There is no entry to update.

Fixing third-party sources is the most effective indirect lever. Outdated comparison articles, review listings on G2 and Capterra, and Wikipedia pages are the sources AI encounters most often. Correcting those changes the signal distribution AI sees over time. ZipTie.dev describes source fixes as "faster, more durable, and more effective than platform correction requests in most cases."

Publishing structured factual statements on owned properties helps for platforms that retrieve from the live web. Schema markup using Organisation, Product, and FAQPage types makes it easier for AI systems to parse your brand's specific claims. This matters more for platforms retrieving from the live web than for queries served from training data, where your website changes won't appear until the next training cycle.

llms.txt is worth implementing as a hygiene step. Over 844,000 sites had adopted the standard as of October 2025. SE Ranking's Yulia Deda ran a controlled test and found that removing the llms.txt variable actually improved the model's citation predictions. SE Ranking concluded that "LLMs.txt doesn't impact how AI systems see or cite your content today." Implement it for structured discoverability; don't count on it to correct inaccurate descriptions.

Corrections surface gradually and on different schedules across platforms. Sight AI notes that improvements appear as different platforms update their knowledge bases. For queries served from training data rather than live retrieval, corrections may not appear until the next training cycle.

Visibility and accuracy are different problems

Most brands don't know what AI is saying about them. Standard analytics don't surface the buyer who encountered wrong pricing in an AI response and never made contact. The gap between what AI says about your brand and what is actually true is mostly invisible through standard measurement.

A brand that appears frequently in AI responses with wrong pricing is not well-served by improving its appearance rate. Finding out whether the description AI produces is correct

requires running the queries and checking responses against known facts. Most teams are not structured to do that systematically across multiple platforms.

Aiviara is building infrastructure for monitoring AI brand citations and factual accuracy across LLM platforms. Early access information is available at aiviara.com.