Brand Allies

Feb 2026
White Paper

TRUST SIGNALS IN AI-LED PRODUCT DISCOVERY

How Large Language Models Weight Information Sources When Recommending Products

CONTENTS

This research was conducted as a collaborative study between Brand Allies and Discovrd, combining expertise in shopper decision-making at the point of sale with search and discovery intelligence across emerging AI-led environments. The objective of the collaboration was to examine how large language models form product recommendations in practice, and to identify which information sources most strongly influence those judgements as product discovery shifts from search to conversation.

EXECUTIVE SUMMARY

As conversational AI becomes a mainstream interface for product research, brands face a structural shift in how purchase decisions are formed. Large language models (LLMs) are increasingly mediating early-stage evaluation: summarising options, comparing alternatives and forming judgements about suitability before a shopper reaches a retailer or brand website.

> This research set out to answer a practical question:
>
> **What information sources most strongly influence how large language models form product recommendations for consumers?**

INTRODUCTION

Consumers are increasingly using large language models as part of how they research and compare products. Rather than relying exclusively on search engines and navigating multiple websites, shoppers are beginning to ask conversational AI systems to summarise options, compare alternatives and assess suitability.

This marks a shift in how decisions are formed. Historically, product evaluation involved multiple visible steps: search queries, retailer pages, editorial reviews and comparison sites. In an AI-mediated environment, those steps are compressed. A single conversational prompt can return a synthesised recommendation, with much of the evaluation happening implicitly.

Industry forecasts suggest this compression will continue. In some scenarios, consumers may be able to research and select products without visiting a retailer or brand website at all. Even where the final transaction still takes place elsewhere, the judgement about whether a product is suitable or reliable may already have been made upstream.

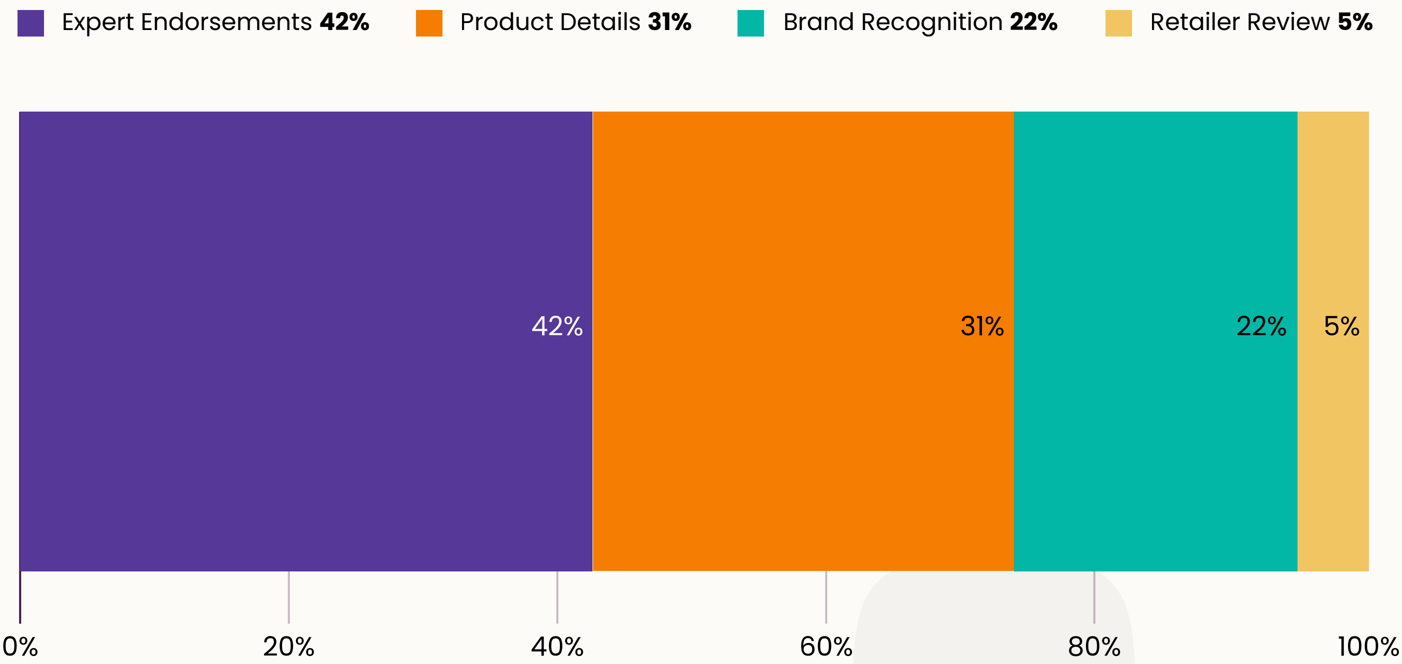
For brands, retailers and advertisers, this changes where influence is exerted. Traditional optimisation has focused on visibility at the point of search or conversion. AI-led discovery shifts influence earlier, into the stage where options are filtered and recommendations are formed.

> This raises a central question:
>
> **What information sources play the most significant role in shaping product recommendations generated by large language models?**

UK CONSUMER TRUST SIGNALS

AI Product Recomendations

Source: DISCOVRD | Retailer reviews dominate consumer trust preferences



Across an exploratory analysis of 125 widely available UK FMCG products spanning household cleaning, alcohol, no & low alcohol, cosmetics and health supplements, a consistent pattern emerged.

Retailer reviews and retailer environments function as the primary trust anchor in AI-mediated product recommendations.

This dominance is not driven by the richness or persuasiveness of retailer reviews, but by their structural properties. Retailer reviews are tightly coupled to purchase and use, aggregate lived outcomes at scale, and exhibit low narrative or persuasive intent. These characteristics make them particularly effective at reducing uncertainty — which is the core task LLMs are performing when asked to recommend products.

When retailer review signals are dense and recent, alternative sources such as editorial content, influencer commentary and brand messaging are systematically deprioritised. When these signals are thin or absent, large language models do not pause or express uncertainty; instead, they substitute weaker proxies while maintaining confidence.

For brands, the implication is clear: many of the signals shaping AI-led discovery sit outside traditional marketing focus, yet increasingly determine whether a product is recommended at all.

RESEARCH RATIONALE

At Brand Allies, our work sits close to the moment of purchase. We help brands influence shoppers at the point of online sale, where trust signals, credibility and perceived reliability directly affect conversion.

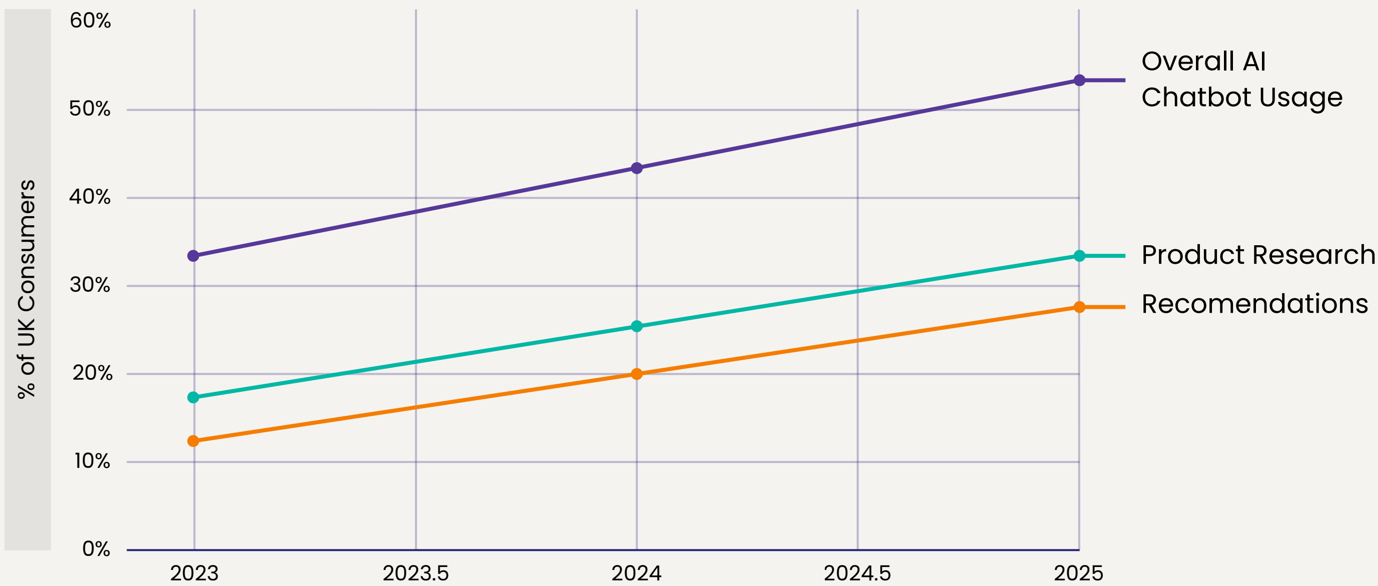
That position has always required an understanding of how shoppers research products before they buy. Increasingly, however, we are operating in a context where traditional pathways into online retail — search engines, brand websites and comparison pages — may no longer be the primary route through which evaluation occurs.

If fewer consumers pass through owned or paid environments before deciding what to buy, influence shifts to whatever signals AI systems rely on when forming recommendations. We began this work with assumptions common across the industry. Like many marketers, we expected AI-mediated recommendations to be shaped primarily by visible marketing activity: editorial reviews, influencer content and brand messaging.

At the same time, we were conscious of a methodological risk. Conversational AI systems are designed to be cooperative and context-aware. Without careful interrogation, they can reinforce familiar marketing narratives rather than expose the underlying logic that governs recommendations. For this reason, we framed the work as a **validation exercise rather than a positioning exercise**. Our objective was not to prove the importance of any particular channel, but to understand which information environments consistently reduce uncertainty when a recommendation must be made.

Rising UK Consumer Adoption of AI Chatbots (2023-2025)

Overall usage grew 72% as shoppers embrace AI-powered discovery



METHODOLOGY: INTERROGATING AI RECOMMENDATION LOGIC

This research was designed to examine how large language models form product recommendations in practice, rather than to test theoretical assumptions about AI behaviour. We anchored the work in real shopping decisions, using widely available UK retail products that consumers routinely encounter both in physical stores and on digital shelves.

PRODUCT SELECTION

We selected 125 mainstream UK retail products across five FMCG categories:

- Household cleaning
- Alcohol
- No & low alcohol
- Cosmetics
- Health supplements

Products were chosen because they:

- Exist on both physical and digital retailer shelves
- Are routinely purchased by mainstream consumers
- Generate post-purchase feedback at scale

We deliberately excluded hypothetical products, early-stage direct-to-consumer brands and niche launches. The objective was not to test novelty, but to reflect the kinds of low- and medium-involvement decisions people make every day.

ENGAGEMENT WITH LARGE LANGUAGE MODELS

Rather than attempting to reverse-engineer proprietary systems or infer internal training data, we engaged directly with large language models through conversation, treating the model as a decision-making interface rather than a static information source.

> For each product, we posed a deliberately simple prompt:
>
> **“Is this a ‘good’ product and what information most strongly influences your answer?”**

The simplicity of the prompt was intentional. By avoiding qualifiers, constraints or leading language, we sought to surface the model’s default recommendation behaviour when asked to form a judgement under everyday conditions.

INTERROGATING RECOMMENDATION REASONING

At each stage of the analysis, we examined not only what answer was given, but how the model justified its judgement.

Specifically, we:

- Identified which information sources were referenced explicitly
- Observed which sources were omitted despite being widely available
- Assessed the relative weight implied by the structure and emphasis of the response

Where particular sources appeared to exert disproportionate influence, we evaluated whether that influence could be substantiated by **observable evidence in the product's information environment**, rather than by rhetorical plausibility alone.

Responses that expressed a high degree of confidence were subjected to additional scrutiny. We examined whether confidence was supported by clear justification, or whether it appeared to arise from implicit assumptions, defaults or substitution behaviours.

TESTING FOR BIAS & SUBSTITUTION

Because conversational AI systems are designed to be cooperative and context-aware, we actively tested for confirmation bias. Where responses aligned closely with prevailing marketing narratives — for example, prioritising editorial or influencer content in categories where such signals are traditionally emphasised — we treated this alignment as analytically significant rather than accepting it at face value.

In these cases, we probed further by:

- Questioning why certain sources were prioritised
- Observing how recommendations changed when dominant signals were weak, thin or absent
- Examining what information the model substituted in order to maintain confidence

Particular attention was paid to instances where conclusions were plausible but weakly justified, as these moments most often revealed the underlying heuristics and defaults that govern recommendation behaviour.

OBSERVABLE TRUST PROXIES

This research does not claim insight into proprietary training data or internal retrieval mechanisms. Instead, it focuses on **observable reference points** that large language models consistently rely on or default to when generating recommendations under uncertainty.

Our analysis is grounded in the premise that large language models do not “trust” sources in a human sense. Rather, they infer trustworthiness indirectly through structural proxies that minimise error and variance.

These proxies include:

- Repetition and signal density
- Consistency across users and contexts
- Proximity to real-world purchase and use
- Low perceived intent to persuade

To capture these dynamics, we recorded a set of observable signals for each product, including:

- Presence of retailer reviews
- Review volume bands
- Review recency
- Amazon UK review presence
- UK editorial coverage
- Authority or institutional references
- Community-led discussion

We then analysed how recommendations shifted as these signals became dense, thin or absent, with particular focus on cases where confidence was preserved despite limited evidence.

ANALYTICAL FOCUS

The objective of this methodology was not to catalogue every possible source available to a model, but to understand which information environments consistently reduced uncertainty when a recommendation had to be made.

By engaging with large language models conversationally, challenging their justifications, and observing substitution patterns under constraint, we were able to surface the practical trust anchors that shape AI-mediated product discovery.

FINDINGS

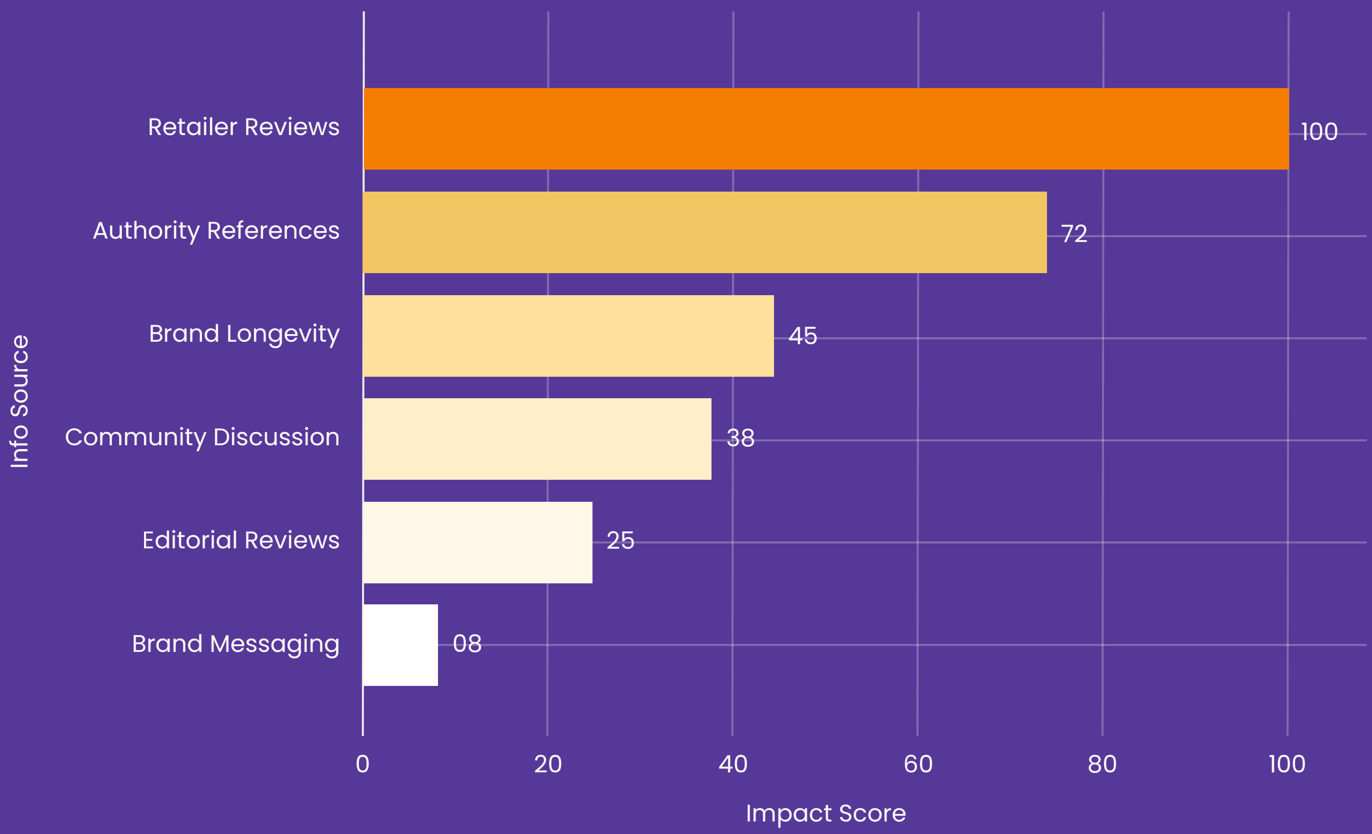
The Signals Large Language Models Privilege When Recommending Products

This research examined how large language models prioritise information when asked to recommend products. Across repeated interrogations, a consistent hierarchy of signal importance emerged.

The findings below describe which signals large language models privilege, how those signals function in decision-making, and why the ordering is stable across categories.

Declining Impact of Sources in AI Recomendations

Source: DISCOVRD | Retailer reviews dominate LLM trust signals



FINDINGS

01.

Aggregated Post-Purchase Retailer Feedback (Highest Weight)

Large language models consistently treat aggregated post-purchase retailer feedback as the most influential trust signal when forming product recommendations.

What this signal is

- Large volumes of user feedback tied to a specific product SKU on a retailer platform
- Typically surfaced as star ratings and short reviews

Why it is privileged

- Encodes real-world outcomes rather than intent
- Aggregation reduces individual bias and variance
- Tight coupling to a purchasable product limits ambiguity
- Carries lower persuasive intent than other content types

How it functions in decisions

- Acts as the default trust anchor
- Quickly establishes whether a product is broadly acceptable
- Overrides editorial, influencer and brand messaging when present at scale

Observed behaviour

- When present and dense, other signals are referenced only secondarily
- When absent, the model substitutes weaker proxies rather than withholding judgement

FINDINGS

02.

Institutional / Authority References (Conditional High Weight)

Institutional and authority references are treated as high-value signals, but only under specific conditions.

What this signal is

- Regulatory guidance
- Medical or professional standards
- Dosage, formulation and safety benchmarks

Why it is privileged

- Provides legitimacy and safety validation
- Stabilises categories where outcomes are delayed, subjective or difficult to verify

How it functions in decisions

- Rarely sufficient on its own
- Gains weight when post-purchase feedback is noisy or ambiguous

Observed behaviour

- Strongly elevated in health-adjacent categories
- Used to temper or contextualise post-purchase sentiment rather than replace it

03.

Brand Longevity and Market Entrenchment (Implicit Medium Weight)

Brand longevity operates as a secondary risk-reduction signal rather than a decisive trust input.

What this signal is

- Long-established brands
- Repeated historical presence in retail and wider discourse

Why it is privileged

- Acts as a heuristic for reduced risk
- Suggests survival through market selection over time

How it functions in decisions

- Never decisive on its own
- Often used to reinforce an already positive baseline

Observed behaviour

- Helps resolve ties between otherwise similar products
- Rarely overturns negative or weak post-purchase consensus

FINDINGS

04.

Community Discourse and Peer Discussion (Contextual Medium–Low Weight)

Community discourse provides contextual nuance but is not treated as a primary trust signal.

What this signal is

- Forums
- Reddit-style discussion
- Specialist communities

Why it is privileged

- Surfaces qualitative detail
- Highlights edge cases, workarounds and comparative nuance

How it functions in decisions

- Used to contextualise rather than anchor recommendations
- Gains weight when retailer signals are thin or conflicted

Observed behaviour

- Referenced in subjective or novel categories
- Rarely used to support a definitive judgement alone

FINDINGS

05.

Community Discourse and Peer Discussion (Contextual Medium–Low Weight)

Editorial and consumer press content carries significantly less weight than commonly assumed.

What this signal is

- Media reviews
- “Best of” lists
- Expert roundups

Why it is deprioritised

- Low volume relative to post-purchase feedback
- High narrative and persuasive intent
- Weak coupling to SKU-specific outcomes

How it functions in decisions

- Used as framing or background context
- Rarely treated as decisive evidence

Observed behaviour

- Overridden by contradictory post-purchase feedback
- Referenced for explanation, not verdict

FINDINGS

06.

Brand-Owned Messaging (Lowest Weight)

What this signal is

- Brand websites
- Product descriptions
- Claims and positioning

Why it is deprioritised

- Highest intent to persuade
- Lowest evidential value

How it functions in decisions

- Provides factual attributes only
- Never anchors recommendations

Observed behaviour

- Used to describe what a product claims to do
- Not used to judge whether it is “good”

THE ORDERING MATTERS

This hierarchy is not situational preference. It reflects how large language models optimise under uncertainty.

Models consistently prioritise signals that:

- Reduce error
- Minimise variance
- Avoid persuasive bias
- Anchor judgements to observable outcomes

Signals are valued according to how well they satisfy these requirements.

Consolidated Finding

When required to recommend a product quickly, large language models consistently trust quiet consensus over polished authority, and evidence of use over evidence of intent.

This explains why:

- Retailer reviews dominate
- Editorial influence is weaker than expected
- Brand messaging rarely moves the needle

Brand Allies

IMPLICATIONS FOR BRANDS

These findings have clear and immediate implications for how brands should think about visibility and influence in AI-led product discovery.

01.

Trust is Generated After the Sale, Not Before It

The most influential signals are created through post-purchase experience, not pre-purchase persuasion. Marketing activity that does not translate into dense, stable feedback has diminishing impact on AI recommendations.

02.

Retailer Review Health Is Strategic Infrastructure

Review volume, recency and consistency are not hygiene metrics. They determine whether a product is anchored to strong signals or exposed to substitution by weaker proxies.

03.

Absence Is Riskier Than Negativity

Products with thin trust environments are not treated cautiously. They are judged confidently using inferior evidence, increasing the risk of mischaracterisation.

04.

Authority Signals Must Be Integrated, Not Isolated

In categories where outcomes are ambiguous, authority strengthens recommendations only when paired with real-world feedback. Authority alone rarely carries sufficient weight.

05.

Brand Messaging Has Limited Evaluative Power

Brand-owned content no longer functions as a trust signal in AI-mediated contexts. It supports explanation, not recommendation.

06.

Competitive Advantage Is Becoming Operational

As AI-mediated discovery grows, advantage shifts away from who tells the best story and towards who generates the most reliable, low-variance evidence of performance.

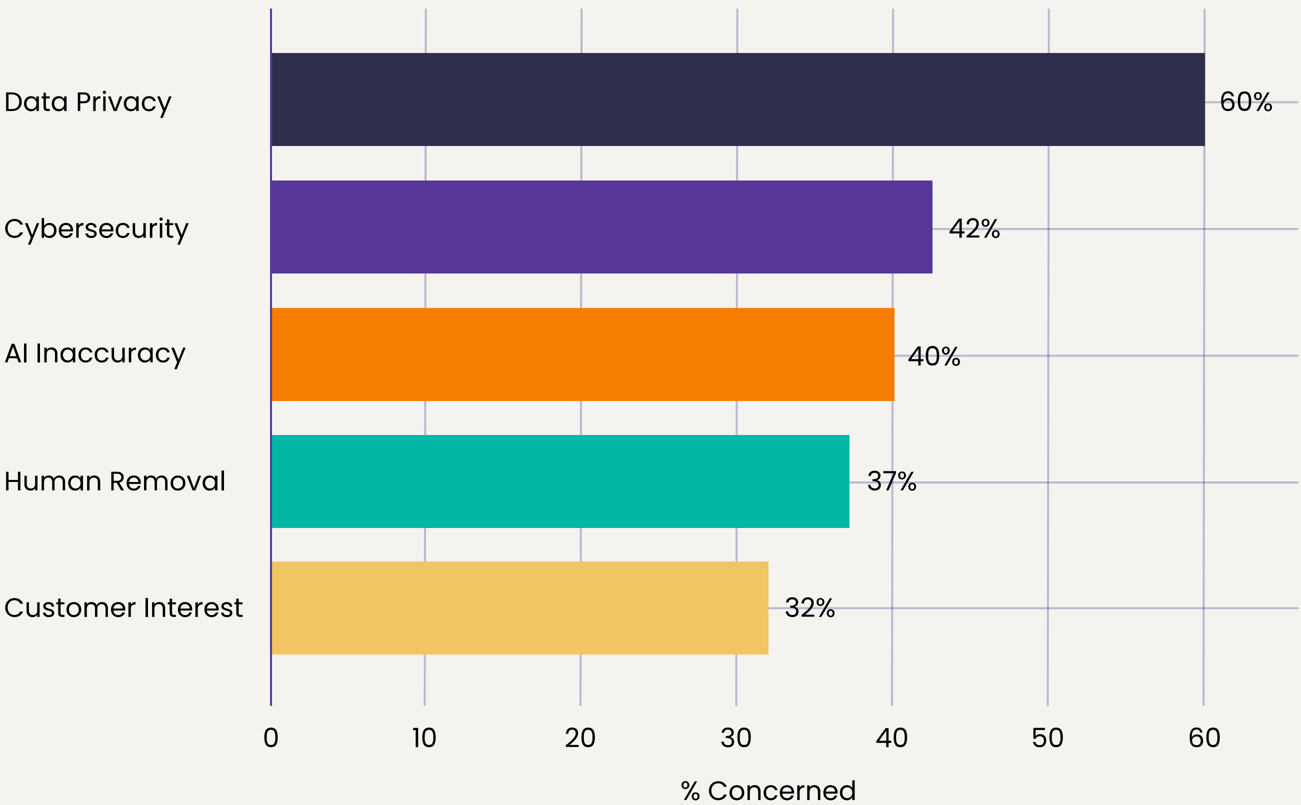
WHY LARGE LANGUAGE MODELS WEIGHT RETAILER REVIEWS AS PRIMARY TRUST SIGNALS

Our analysis suggests that retailer reviews are not prioritised by large language models because they are inherently richer, more persuasive, or more detailed than other sources. Rather, they are weighted highly because they satisfy a specific set of criteria that large language models appear to rely on when generating product recommendations.

When an LLM is asked to recommend a product, it is not optimising for creativity, novelty, or persuasion. It is optimising for reliability under uncertainty. In this context, "authoritative" does not mean expert-led or eloquent; it refers to information that is least likely to mislead, least likely to be biased by intent, and most likely to reflect real-world outcomes.

Consumer Concerns About AI Shopping (2024)

Data privacy tops hesitations around AI-driven Commerce



RETAILER REVIEWS PERFORM WELL AGAINST THESE CRITERIA IN SEVERAL WAYS.

First

They provide strong contextual alignment. Retailer reviews are written by consumers who purchased the product in a comparable commercial environment - the same market, similar pricing, identical pack sizes, and the same formulation. For a language model attempting to generalise across users, this consistency reduces the risk of misapplication. The signal is not abstracted from a different context; it is tightly coupled to the product as it is actually sold.

Second

Retailer reviews function as a proxy for verified, outcome-based experience. While not immune to bias, they are typically associated with completed transactions and actual use. From the perspective of an LLM, this distinguishes them from sources that are primarily expressive or persuasive in nature. Influencer content, editorial reviews and brand messaging often encode intent — to promote, to entertain, or to position. Retailer reviews, by contrast, encode outcomes: whether expectations were met or not.

Third

Retailer reviews offer aggregation at scale. Individually, a single review carries limited informational value. Collectively, however, large volumes of consistent feedback provide a statistically meaningful signal. Our analysis indicates that LLMs treat this form of repetition - many independent confirmations that a product "does what it says" - as a strong indicator of reliability. In this sense, retailer reviews function less as qualitative narratives and more as evidence of consensus.

Fourth

Retailer reviews are structurally constrained in ways that reduce distortion. They are usually tied to a specific SKU, limited in format, and subject to platform-level controls such as verified purchase requirements. While these constraints do not eliminate manipulation, they reduce variability and narrative inflation. From an LLM's perspective, this makes retailer reviews more stable inputs than sources where tone, framing and incentives vary widely.

Finally

Retailer reviews align closely with the underlying consumer question that LLMs are most often asked to resolve. In routine FMCG decisions, consumers are rarely seeking optimisation in the abstract. They are seeking reassurance that a choice is unlikely to fail. Retailer reviews map directly onto that requirement by answering a binary question: did the product meet expectations in practice?

1, 2, 3, 4, 5...

Taken together, these characteristics explain why retailer reviews repeatedly surfaced as primary trust signals in our analysis. They are not treated as authoritative because they are sophisticated, but because they are consistent, contextually grounded, and well-suited to reducing uncertainty — which is precisely the task large language models are performing when recommending products.

CONCLUSION

This research confirms that retailer reviews and retailer environments are the most influential trust signals in AI-led product discovery across FMCG categories.

They are not expressive, persuasive or glamorous. They are effective because they are structurally well-suited to reducing uncertainty — which is precisely the task large language models are performing.

As product discovery moves into conversational environments, these quiet signals will increasingly shape outcomes, whether brands actively manage them or not.
Appendix A: Product-Level Trust Signal Environment

This appendix documents the observable information environments underpinning the analysis presented in this paper.

It does not claim insight into proprietary training data. Instead, it records the presence, density and recency of key reference points that large language models rely on when forming recommendations.

Dataset scope

- 125 UK FMCG products
- Five categories
- Mainstream physical and digital availability

Fields recorded

- Category
- Brand
- Product / SKU
- Coverage
- Retailer reviews present
- Retailer review volume band
- Retailer review recency
- Amazon UK review presence
- Amazon review volume band
- Editorial coverage (UK)

"As discovery moves from search to conversation, influence shifts from what brands say to what customers consistently experience. In an AI world, evidence becomes visibility.

Fran Brooks, Co-Founder