# Recommendation Capture:

# Adversarial Optimization, the Verification Vacuum, and the Hollowing of AI Product Recommendation

Travis Gilly
ORCID: 0009-0007-2954-6313
*Real Safety AI Foundation*
*Illinois, United States*

Working paper, June 2026 (v1)

**ABSTRACT**

Generative search interfaces are marketed on a single promise: the user no longer has to compare options, because the model has already done it. This paper argues that the same promise creates the incentive to corrupt the answer, and that a formalized optimization industry already does so at scale. Generative engine optimization, the practice of engineering content so that language models cite and recommend a product, derives much of its measured effectiveness from the manufacture of authority signals rather than from product merit. When every competitor in a category runs the same play, the model's reported best tracks the best optimizer, a quantity orthogonal to whether the product is good. I name this condition recommendation capture, by analogy to regulatory capture: an arbiter sold as neutral comes to serve whoever optimizes hardest, against the buyer it claims to serve. I argue that the resulting harm concentrates in a verification vacuum, the set of product categories where no independent ground truth exists, so that the model's recommendation is wholly a function of optimization, and that these categories tend to be the niche, the new, and the health adjacent, where vulnerable buyers are least able to check. Because models increasingly read the same open web that optimization floods, the corrupted output becomes the next input, and the interface erodes the trust it depends on. I sketch a tiered liability picture under existing false advertising and unfair practices law, identify which elements survive and which are untested, and argue that the deeper failure is one of market integrity and epistemics that current governance is structurally unable to detect.

**Keywords:** generative engine optimization, AI search, recommendation systems, consumer protection, false advertising, market integrity, AI governance, data integrity

## I. INTRODUCTION

Advertising is the thing almost everyone tolerates and almost no one likes. It funds the open web, and it is also the reason large parts of the open web are unpleasant to use. For two decades the labor that made advertising win, search engine optimization, was performed largely out of view. Optimizers shaped which pages rose to the top of a ranked list, and because the result was still a list of competing links that a reader could scroll, compare, and distrust, the manipulation stayed legible. The reader retained the final move. The discipline drew remarkably little public scrutiny

for an industry whose explicit purpose was to influence what billions of people see.

Generative search changes the terms. When a person asks a conversational model for the best product in a category, the model does not return a field of competing links. It returns a synthesized answer in a single authoritative voice, often naming a handful of options and sometimes only one, with a justification attached. The interface is sold precisely on the elimination of the reader's final move. The pitch, repeated across the launch materials of every major model and every agentic browsing product, is that the user no longer has to comparison shop, because the assistant has done the comparing. That pitch is the product. It is also, this paper argues, the vulnerability.

If the value of the interface lies in the user's trust that the answer is a neutral synthesis rather than a sales pitch, then the answer becomes the most valuable real estate on the internet to capture, and an industry has already formed to capture it. Generative engine optimization, sometimes marketed as AI search optimization, is the practice of engineering a brand's presence so that language models mention, cite, and recommend it. It is real, it is formalized, and it is growing (Aggarwal et al., 2024; Search Engine Land, 2026). The thesis of this paper is that this industry does not merely move products up a list. It manufactures the appearance of authority that the model consumes as evidence, and in doing so it converts an interface trusted as an arbiter into an advertising channel whose advertising is invisible as such to the person reading it.

The paper proceeds as follows. Section II describes the structural shift from ranked links to the single synthesized voice and what the reader loses in the transition. Section III examines what the optimization industry actually does and why its effectiveness comes substantially from authority signaling rather than from merit. Section IV defines recommendation capture and distinguishes it from ordinary marketing. Section V introduces the verification vacuum and argues that harm concentrates where independent verification is thinnest. Section VI describes the feedback loop by which the corrupted output becomes the next model's input. Section VII sketches a tiered liability picture under existing law and marks the elements that are untested. Section VIII argues that existing governance is structurally unable to detect the harm, and Section IX concludes.

## II. From Ranked Links to the Single Voice

The classic search results page presented a field. Ten links sat in ranked order, and the order was contestable, gamed, and frequently commercial, but the field itself remained visible. A reader who distrusted the top result could read the second, open three tabs, notice that the top entry was a sponsored placement, and form an independent judgment from a set of options laid side by side. Manipulation operated on the ordering, not on the existence of the field. The reader could always see that a field existed and that other entries were available to weigh.

Generative interfaces collapse the field. The model retrieves passages from across the web and composes them into a single answer, and the user receives a shortlist rather than a page (arXiv, 2025). In a ranked list, ten entries compete for attention. In a synthesized response, a few options are fused into one concise, justified recommendation, and the rest are simply absent. Absence here is not a low rank that a determined reader can scroll to reach. It is non appearance. The product that the model does not name does not exist for that user in that moment, and the user has no signal that anything was omitted, because the interface is designed to read as a complete answer rather than as the visible tip of a larger set.

Two things are lost in the transition, and both are load bearing for what follows. The first is the field of view. The reader can no longer see the set from which the recommendation was drawn, so the reader cannot notice what is missing. The second is the audit handle. A ranked list wears its commercial nature on its surface; sponsored entries are labeled, and a skeptical reader learns to discount the top of the page. A synthesized recommendation arrives with the model's own credibility attached and no visible seam between the parts that are neutral information and the parts that are the

residue of someone's marketing. The interface that was sold as a convenience, the removal of the reader's labor, is the same interface that removes the reader's ability to check.

## III. The Optimization Industry and the Manufacture of Authority

It would be a weaker argument if optimizing for model recommendation were impossible or merely speculative. It is neither. The practice has a founding academic literature and a maturing commercial apparatus. The work that formalized it provided the first systematic evidence that specific content interventions can change a model's visibility, and the term it introduced, generative engine optimization, is now the name of an industry (Aggarwal et al., 2024).

The uncomfortable finding in that founding work is not that optimization is possible but where its effectiveness comes from. Traditional keyword tactics had negligible or negative effect on model citation. What moved the needle was the appearance of authority. Inserting citations, statistics, and quotations, the surface features of a credible source, raised a lower ranked page's visibility in generated answers by a large margin in the reported experiments (Aggarwal et al., 2024). The lift attaches to the costume of evidence, not to the thing the costume signifies. A weaker product dressed in the formatting of a well sourced one can be promoted over a stronger product that lacks the dressing. This is the mechanism the rest of the paper depends on, so it is worth stating plainly: the model rewards the legible performance of authority, and the performance can be manufactured.

The commercial tactics follow from the finding. Practitioners advise securing placement in the comparison and best of articles that models copy, and the placement is obtained either by publishing the list oneself with one's own product included or by paying a review site for the slot (First Page Sage, 2026). Newer tooling is more direct still: vendors now offer to identify the specific forums and communities that models scan for consensus and then to mobilize a brand's own customers to post in exactly those places, so that the watering hole the model drinks from is seeded in advance (Yotpo, 2026). Each of these is a way of placing, into the corpus the model reads, content whose purpose is to be reflected back to a buyer as a neutral finding.

The act has a clean description. A commercial claim is deposited into the information environment; the model retrieves it, strips the commercial origin, and re emits it in its own voice as an apparent fact; and the buyer receives a recommendation that has been laundered of its provenance. The accountability half of this dynamic already has a name in the AI governance literature, where authority laundering denotes the move by which a decision is justified by the model having said so while responsibility for the decision flows away from every human in the chain (IRSA Institute, 2026). The commercial half is the same structure pointed at a purchase: the claim keeps the model's borrowed credibility and loses its return address.

## IV. Recommendation Capture

The individual case, one brand laundering one claim, is a consumer protection problem of familiar shape. The systemic case is different in kind, and it is the contribution this paper means to isolate.

*Recommendation capture* is the condition in which a recommendation layer that is presented to users as a neutral arbiter comes to be governed by whoever optimizes for it most effectively, rather than by the merits of the options it ranks. The analogy is to regulatory capture, in which an agency established to serve the public interest is gradually run in the interest of the industry it was meant to oversee. The captured agency still performs the outward motions of neutral judgment; the capture lives in whose inputs it actually responds to. The captured recommendation layer is the same. It still answers in the voice of a disinterested expert. What has changed is that the inputs it responds to are dominated by the parties with the resources and the incentive to shape them.

The existence of the optimization industry is itself the proof that the layer is capturable and contested. A profession cannot be paid to make a model choose a given client unless model choice is something that effort and expenditure can move. The whole sector is an existence proof of the battlefield. And once one competitor in a category optimizes, the rest must, because the cost of non appearance is total. When every serious competitor runs the same play, the equilibrium is not that the model surfaces the best product. The equilibrium is that the model surfaces the best optimized product, and optimization skill is orthogonal to whether the bag supports a spine or the supplement does what it claims. The reported research even cuts against the intuition that the largest firm simply wins; challenger brands can outperform incumbents on authority signaling and structural clarity (Aggarwal et al., 2024). That sounds meritocratic until one sees what the merit being rewarded is. It is merit at optimization, not merit at the underlying good.

The user, meanwhile, receives a single confident answer and is told nothing of the contest behind it. They believe they have been handed a verdict. They have been handed the scoreboard of a game played in a room they were not allowed to enter, and were not told the game existed.

## V. The Verification Vacuum

Recommendation capture does not do equal damage everywhere. Its severity is a function of how much independent ground truth exists about a category, and the relationship is inverse.

Consider a mature consumer electronics category, a flagship phone. The model that recommends one is anchored by an enormous body of independent evidence: professional review outlets, standardized benchmarks, teardown analyses, and a deep reservoir of user reports. Optimization in such a category is a thumb pressing on a scale that already has real weight on both pans. It can shift the result at the margin, but the abundance of independent verification limits how far the answer can be moved from the truth.

Now consider a category with little or no independent evidence. A niche posture support product, a new supplement, a small medical device, a service in a thin market. In categories like these, the only assertions that exist about the product are frequently the ones the seller wrote. There is no independent benchmark, no review outlet of record, no aggregated body of disinterested testing. Here optimization is not a thumb on a scale. It is the entire scale. When the only inputs about a product are self authored, any optimization that gets the model to repeat the seller's superiority claim is laundering self issued marketing into apparent fact, and the model has no means of distinguishing self issued authority from earned authority. It sees confident, well structured, citation dense claims and passes them on with its own credibility stamped on top.

I call this condition the verification vacuum: the set of categories in which no independent ground truth exists, so that the model's reported best is wholly a function of optimization. The argument of this section is that the categories most exposed to the vacuum are disproportionately the niche, the new, and the health adjacent. These are precisely the categories where a buyer has the least independent ability to check a claim, and frequently where the stakes for getting it wrong are highest. The structure of the harm is therefore regressive in the most literal sense. The contest decides the most where the buyer can verify the least. A person researching a flagship phone is well defended by the surrounding evidence even if they trust the model completely. A person researching an unfamiliar back brace for chronic pain, in a category with no independent record, is defended by nothing but the optimization budgets of the sellers, and is told none of this.

## VI. The Feedback Loop

The harm does not sit still. It compounds, through two loops that reinforce each other.

The first loop is adversarial and economic. The value of the recommendation layer to a buyer is its perceived neutrality. That same perceived neutrality is what makes the answer worth corrupting. The more users trust the answer, the more valuable it becomes to place a product inside it; the more it is optimized against, the less the answer deserves the trust that made it a target. An interface that runs on credibility creates, by running on credibility, the incentive that erodes it. Left unchecked, the layer consumes the trust it depends on.

The second loop is epistemic and concerns the supply of information itself. Models are trained on, and increasingly retrieve from, the open web. Optimization works by flooding that same web with brand favorable, authority costumed content. The corpus the model reads is therefore being actively shaped by parties whose goal is to be recommended, which means the poison and the food are the same substance. There is a terminological caution to register here. In machine learning security, data poisoning has a specific meaning, the injection of crafted examples into a training set to corrupt a model's behavior, and most of what optimization does is better described as adversarial manipulation of the retrieval and ranking layer rather than classical poisoning. The exception is genuine: when models train on the open web, flooding that web with brand favorable content does reach training data, and the poisoning frame is fair for that case. Either way, the degradation of a model's information environment by accumulated low quality, self interested content is a recognized concern, related to what the literature on training on generated data calls model collapse (Shumailov et al., 2024). The point for this paper is narrower and sturdier than any strong claim about collapse: the recommendation layer is being optimized against by many parties at once, in a manner structurally similar to coordinated automated flooding, and it reads the polluted record it helped create.

The honest formulation of the resulting claim is not that every AI recommendation is a lie. It is that the recommendation layer is contestable, it is actively being contested, and the consumer can see neither the contest nor who is winning it. That is a more durable claim than the overstatement, and it is sufficient.

## VII. A Tiered Liability Picture

The legal analysis that follows is secondary to the diagnosis above, and it is offered as a sketch of available theories rather than as the paper's contribution. It is included because the conduct described is not merely distasteful; in a range of cases it is already actionable, and naming the available doctrines clarifies where enforcement could attach. The conduct supports a tiered picture with three potential defendants for a single laundered sentence.[1]

*The brand as false advertiser.* A claim under Section 43(a)(1)(B) of the Lanham Act, 15 U.S.C. 1125(a)(1)(B), requires a false or misleading statement of fact in commercial advertising or promotion of the defendant's own goods. Where the false statement appears in advertising for a different party's product, as when the misrepresentation is delivered by a platform's generative model about a third party's goods, that element is not satisfied as to the platform, which is selling the model rather than the goods. Pointed at the brand and its optimizer, however, the element stands back up. A superiority claim about the brand's own product, authored and seeded for the purpose of being repeated by the model in order to sell that product, is a statement about the defendant's own goods made to promote them. The standing problem that excludes the deceived consumer resolves in the same move: a competing seller losing sales to the laundered claim is the commercial party the zone of interests under Lexmark Int'l v. Static Control Components, 572 U.S. 118 (2014), was drawn to protect.

*The substantiation seam.* A bare assertion that a product is best can be non actionable puffery. It becomes actionable when it reads as an establishment claim, an assertion of test proven superiority, which then requires substantiation. The

[1]The doctrinal claims in this section are stated at the level of available theory. Specific statutory and case citations require verification against current authority before any reliance, and the application of advertising law to optimization content disseminated through a model is, as noted below, untested.

vulnerable target is therefore not a soft claim that a product helps, which a seller with even modest evidence can defend, but the universal superiority claim, best among any in its class, which the evidence in a verification vacuum category will not support. Enforcement should aim at the unsubstantiated universal, not at the defensible specific.

*The optimizer as participant.* The agency that crafts and seeds the claim is reachable in circuits that recognize contributory or participant theories of false advertising liability, though this is messier than the claim against the brand and should be kept as a separate tier rather than blurred into it.

*The unfair practices lane.* Independent of the Lanham theories and their standing puzzles, Section 5 of the Federal Trade Commission Act reaches deceptive practices without a competitor standing requirement. A health or safety superiority claim requires competent and reliable scientific evidence, and the Commission has long held that a marketer cannot make through an intermediary a claim it could not make directly, which describes seeding a claim for a model to relay. Where the optimization touches fabricated or undisclosed reviews, the Commission's 2024 rule addressing fake and AI generated reviews moves the conduct close to a per se violation. This lane reaches the conduct most cleanly.

*The untested novelty.* The load bearing open question is whether optimization content qualifies as commercial advertising or promotion disseminated to the relevant purchasing public when the dissemination runs through a model rather than directly to the buyer. The content is public, the model's crawler ingests it, and the public hears it back from the model, so the argument that it is promotion to the public is strong. It is also, at the time of writing, untested, and it should be named as the open question rather than assumed. The broader uncertainty about whether model output is protected expression remains unresolved as well; courts have so far declined to decide it (Garcia v. Character Technologies, 2025).

## VIII. Why Existing Governance Misses It

The harm described here is difficult for current oversight to see, and the difficulty is structural rather than accidental. Three properties of recommendation capture place it outside the reach of the instruments built to catch advertising harm.

The first is a power asymmetry that is invisible at the point of consumption. The party that decides what the buyer sees is whoever spends and optimizes most, but the spending and optimizing occur upstream, in the corpus and in the ranking, not in a labeled advertisement the buyer can identify. The consumer protection apparatus was built around the disclosed advertisement, the thing marked as a sales message. A recommendation laundered of its provenance presents no such mark.

The second is the invisibility of the contest to the affected party. A harm that a person cannot perceive is a harm they cannot report, and a regime that depends on complaints will not receive them. The buyer who is steered toward an inferior product by a captured recommendation does not experience being steered. They experience receiving an answer. There is no moment of friction at which a grievance forms.

The third is the feedback loop described above, in which the corrupted output becomes the next system's input. Oversight regimes assume a stable target. A harm whose outputs are continuously recycled into the inputs of the next model generation is a moving target that degrades the evidentiary baseline against which any later assessment would be made. By the time a category's record is examined, the record itself has been shaped by the conduct under examination.

These three properties, asymmetric and upstream power, invisibility to the affected party, and a self reinforcing loop, are the signature of a harm that distributes itself below the resolution of the systems meant to detect it. That is why the appropriate register for this problem is market integrity and information integrity rather than individual deception, and why a remedy aimed only at the individual false sentence will not reach the condition that produces the sentences.

## IX. Conclusion

The promise that sold the generative interface, that the user no longer has to compare because the assistant has compared for them, is the same promise that makes the answer worth corrupting and an industry has formed to corrupt it. The optimization of that answer is not a marginal nuisance layered on top of an otherwise sound system. In categories without independent verification it is the whole of the system, and those categories are disproportionately the ones where buyers can check the least and the stakes run highest. The layer eats the trust it runs on, and because it reads the web it pollutes, the damage compounds.

Several minimal interventions follow from the diagnosis without requiring its full acceptance. Provenance and attribution at the point of recommendation would restore part of the audit handle that the single voice removed, letting a buyer see that a claim originated with a seller. Substantiation enforcement that explicitly reaches claims laundered through an intermediary would close the gap that lets a marketer say through a model what it could not say directly. Preserving a visible field, some signal that alternatives exist and were considered, would restore part of the reader's lost final move. And treating the recommendation layer as the advertising surface it has functionally become, for the limited purpose of disclosure, would align the regime with the conduct. None of these requires resolving the harder questions about model expression or platform liability. They require only recognizing that an interface trusted as an arbiter has been quietly turned into a channel, and that the people least able to tell are the ones it will cost the most.

## ACKNOWLEDGMENTS

The author discloses the use of speech-to-text software (Wispr Flow) and Anthropic's Claude as assistive technology for a diagnosed neurodevelopmental condition. All intellectual contributions, research questions, theoretical arguments, methodological decisions, and conclusions are solely the author's own.

Correspondence concerning this article should be addressed to Travis Gilly, Real Safety AI Foundation, Illinois, United States. Email: t.gilly@ai-literacy-labs.org.

## REFERENCES

Aggarwal, P., et al. (2024). GEO: Generative engine optimization. *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. [Citation requires verification of authors, venue, and reported effect sizes.]

arXiv. (2025). Consumer behavior in AI assisted shopping [Preprint, identifier 2509.08919]. https://arxiv.org/abs/2509.08919 [Author and title require verification.]

First Page Sage. (2026). *Generative engine optimization (GEO) strategy guide*. https://firstpagesage.com/seo-blog/generative-engine-optimization-geo-strategy-guide/

Garcia v. Character Technologies, Inc. (2025). [Citation and holding require verification.]

IRSA Institute. (2026). *Authority laundering* [Glossary entry]. https://www.irsa.institute/research/glossary/authority-laundering

Lanham Act, 15 U.S.C. § 1125(a)(1)(B).

Lexmark International, Inc. v. Static Control Components, Inc., 572 U.S. 118 (2014).

Search Engine Land. (2026). *What is generative engine optimization (GEO)*. https://searchengineland.com/what-is-generative-engine-optimization-geo-444418

Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N., & Anderson, R. (2024). The curse of recursion: Training on generated data makes models forget. *Nature*. [Final citation details require verification.]

Yotpo. (2026). *Best AI search engines and strategies*. https://www.yotpo.com/blog/best-ai-search-engines-strat