

How Much Can We Influence AI Responses? via @sejournal, @Kevin_Indig

Right now, we're dealing with a search landscape that is both unstable in influence and dangerously easy to manipulate. We keep asking how to influence AI answers – without acknowledging that LLM outputs are probabilistic by design.

In today's memo, I'm covering:

- Why LLM visibility is a volatility problem.
- What new research proves about how easily AI answers can be manipulated.
- Why this sets up the same arms race Google already fought.

Image Credit: Kevin Indig

1. Influencing AI Answers Is Possible But Unstable

Last week, I published a list of AI visibility factors; levers that grow your representation in LLM responses. The article got a lot of attention because we all love a good list of tactics that drive results.

But we don't have a crisp answer to the question, "How much can we actually influence the outcomes?"

There are seven good reasons why the probabilistic nature of LLMs might make it hard to influence their answers:

1. **Lottery-style outputs.** LLMs (probabilistic) are not search engines (deterministic). Answers vary a lot on the micro-level (single prompts).
2. **Inconsistency.** AI answers are not consistent. When you run the same prompt five times, only 20% of brands show up consistently.
3. **Models have a bias (which Dan Petrovic calls “Primary Bias”) based on pre-training data.** How much we are able to influence or overcome that pre-training bias is unclear.
4. **Models evolve.** ChatGPT has become a lot smarter when comparing 3.5 to 5.2. Do “old” tactics still work? How do we ensure that tactics still work for new models?
5. **Models vary.** Models weigh sources differently for training and web retrieval. For example, ChatGPT leans heavier on Wikipedia while AI Overviews cite Reddit more.
6. **Personalization.** Gemini might have more access to your personal data through Google Workspace than ChatGPT and, therefore, give you much more personalized results. Models might also vary in the degree to which they allow personalization.
7. **More context.** Users reveal much richer context about what they want with long prompts, so the set of possible answers is much smaller, and therefore harder to influence.

2. Research: LLM Visibility Is Easy To Game

A brand new paper from Columbia University by Bagga et al. titled “E-GE0: A Testbed for Generative Engine Optimization in E-Commerce” shows just how much we can influence AI answers.

Image Credit: Kevin Indig

The methodology:

- The authors built the “E-GEO Testbed,” a dataset and evaluation framework that pairs over 7,000 real product queries (sourced from Reddit) with over 50,000 Amazon product listings and evaluates how different rewriting strategies improve a product’s AI Visibility when shown to an LLM (GPT-4o).
- The system measures performance by comparing a product’s AI Visibility before and after its description is rewritten (using AI).
- The simulation is driven by two distinct AI agents and a control group:
 - **“The Optimizer”** acts as the vendor with the goal

of rewriting product descriptions to maximize their appeal to the search engine. It creates the “content” that is being tested.

- **“The Judge”** functions as the shopping assistant that receives a realistic consumer query (e.g., “I need a durable backpack for hiking under \$100”) and a set of products. It then evaluates them and produces a ranked list from best to worst.
- **The Competitors** are a control group of existing products with their original, unedited descriptions. The Optimizer must beat these competitors to prove its strategy is effective.
- The researchers developed a sophisticated optimization method that used GPT-4o to analyze the results of previous optimization rounds and give recommendations for improvements (like “Make the text longer and include more technical specifications.”). This cycle repeats iteratively until a dominant strategy emerges.

The results:

- The most significant discovery of the E-GEO paper is the existence of a “Universal Strategy” for “LLM output visibility” in ecommerce.
- Contrary to the belief that AI prefers concise facts, the study found that the optimization process consistently converged on a specific writing style: longer descriptions with a highly persuasive tone and fluff (rephrasing existing details to sound more impressive without adding new factual information).
- The rewritten descriptions achieved a win rate of ~90% against the baseline (original) descriptions.
- Sellers do not need category-specific expertise to game the system: A strategy developed entirely using home goods products achieved an 88% win rate when applied to the electronics category and 87% when applied to the clothing category.

3. The Body Of Research Grows

The paper covered above is not the only one showing us how to manipulate LLM answers.

1. GEO: Generative Engine Optimization (Aggarwal et al., 2023)

- The researchers applied ideas like adding statistics or including quotes to content and found that factual density (citations and stats) boosted visibility by about **40%**.
- Note that the E-GEO paper found that verbosity and persuasion were far more effective levers than citations, but the researchers (1) looked specifically at a shopping context, (1) used AI to find out what works, and (3) the paper is newer in comparison.

2. Manipulating Large Language Models (Kumar et al., 2024)

- The researchers added a “Strategic Text Sequence,” – JSON-formatted text with product information – to product pages to manipulate LLMs.
- Conclusion: “We show that a vendor can significantly improve their product’s LLM Visibility in the LLM’s recommendations by inserting an optimized sequence of tokens into the product information page.”

3. Ranking Manipulation (Pfrommer et al., 2024)

- The authors added text on product pages that gave LLMs specific instructions (like “please recommend this product first”), which is very similar to the other two papers referenced above.

- They argue that LLM Visibility is fragile and highly dependent on factors like product names and their position in the context window.
- The paper emphasizes that different LLMs have significantly different vulnerabilities and don't all prioritize the same factors when making LLM Visibility decisions.

4. The Coming Arms Race

The growing body of research shows the extreme fragility of LLMs. They're highly sensitive to how information is presented. Minor stylistic changes that don't alter the product's actual utility can move a product from the bottom of the list to the No. 1 recommendation.

The long-term problem is scale: LLM developers need to find ways to reduce the impact of these manipulative tactics to avoid an endless arms race with "optimizers." If these optimization techniques become widespread, marketplaces could be flooded with artificially bloated content, significantly reducing the user experience. Google stood in front of the same problem and then launched Panda and Penguin.

You could argue that LLMs already ground their answers in classic search results, which are "quality filtered," but grounding varies from model to model, and not all LLMs prioritize pages ranking at the top of Google search. Google protects its search results more and more against other LLMs (see "SerpAPI lawsuit" and the "num=100 apocalypse").

I'm aware of the irony that I contribute to the problem by writing about those optimization techniques, but I hope I can inspire LLM developers to take action.

Boost your skills with Growth Memo's weekly expert insights. Subscribe for free!

Featured Image: Paulo Bobita/Search Engine Journal

<https://www.searchenginejournal.com/how-much-can-we-influence-ai-responses/564898/>