

AEO / GEO for High Tech Products

Getting Ranked Fairly in AI Comparisons

White Paper



Michael Williams

October 2026



Table of Contents

1	Executive Summary	3
2	The Search Paradigm Shift: From Rankings to Citations	4
3	Test Methodology	6
4	Conclusions.....	10
5	Bibliography.....	12
6	About Michael Williams	13



1 EXECUTIVE SUMMARY

The rules of digital discovery have fundamentally changed. Traditional Search Engine Optimization (SEO) was built around keyword density, backlinks, and other factors. However, SEO is no longer the only way to guarantee brand visibility. Users are increasingly receiving their answers directly from artificial intelligence rather than clicking through a list of links.

This shift has given rise to two complementary disciplines: Generative Engine Optimization (GEO) and Answer Engine Optimization (AEO). Both describe the practice of structuring content, establishing authority, and building ecosystem presence. These practices optimize how AI-powered platforms (such as Perplexity, ChatGPT, and Google AI Overviews) cite, recommend, and surface a brand's products and expertise when generating responses for users.

This white paper outlines how to improve high-technology product rankings within an LLM. It records an experiment designed to explore how LLMs rank products, and how supplying accurate information can correct an unfairly low ranking. It provides insights into the best ways to do so, all within the scope of product marketing.

Results

A neutral buyer's guide carrying verified product figures did the most to improve how ChatGPT scored Securin Vulnerability Management against Cisco Kenna, a larger, more established vendor. Process:

1. **Map the evaluation criteria.** Asking the LLM for the market's key features, their categories and the metrics used to evaluate each produced a working hypothesis of what drives its scores. Most of the relevant Securin features were missing from public collateral, because they were too detailed for it.
2. **Content effectiveness testing.** The same figures were placed in three formats. The buyer's guide lifted Securin's Harvey balls most, the white paper less, and the data sheet least.
3. **Rerun the comparison.** With the buyer's guide as input, Securin's scores rose sharply, past what a formal comparison would support. That makes the rerun an upper bound, not a forecast.

These are field notes from about 10 tests run in ChatGPT during Q4 2025, inside chat sessions. They show how providing the right source information to an LLM can positively impact product ranking.

Key Insight: In these tests, a neutral buyer's guide was the most effective format. To affect rankings outside a test session, it must be published where LLMs can find and index it: the company's website, an industry site or a buyer's guide repository.

For a startup competing with a larger, more established firm, this can help correct an unfairly low ranking. Because the method can also overshoot, the content must stay accurate and credible.



2 THE SEARCH PARADIGM SHIFT: FROM RANKINGS TO CITATIONS

2.1 WHY TRADITIONAL SEO IS NO LONGER ENOUGH

Buyers researching technology products increasingly ask an AI assistant for a comparison instead of reading a page of search results. The answer often names a short list of vendors and summarizes their strengths, and that summary may be the first impression a buyer gets.

Generative Engine Optimization (GEO) and Answer Engine Optimization (AEO) build on SEO rather than replace it. What changes is the goal: a product needs to be described accurately inside the AI's answer, not just ranked on a results page.

So, how does product marketing need to adapt to these changes? Do the changes make it easier for startups / smaller teams to compete with larger, established companies or more difficult?

This paper explores those questions through a small experiment, run entirely within the scope of product marketing.

2.2 BACKGROUND FOR THE TEST

At Securin, I used an LLM to help build competitive battle cards for the product line. One comparison stood out: the LLM ranked Securin Vulnerability Management well below Cisco Kenna, a name-brand competitor. Given the two products' feature sets, that ranking seemed unjustifiably low.

LLM comparisons are a black box. They don't say which features matter most, or how each one is measured. So the challenge was to learn which product features the LLM valued most in vulnerability management, and which metrics it used to evaluate each one.

This paper describes how I approached it, what the results did and did not show, and what it means for product marketing.

Key assumption: This project had to improve AEO and GEO results using product marketing techniques alone. Technical changes, such as website markup or JSON-LD structured data, were out of scope.

2.3 CHALLENGES

I needed to work out how the LLM was evaluating products: which criteria it used, and how granular those criteria were for each feature. The obvious shortcut, extracting the model's judgments at scale by automation, was off the table.



2.3.1 Why Not Automate the Extraction?

Systematically harvesting a model's outputs to replicate its capabilities is called distillation. The technique itself is standard machine learning: Hinton, Vinyals and Dean described training a smaller model to mimic a larger one in 2015.¹ Doing it to someone else's model without permission is a different matter:

- **Provider terms prohibit it.** OpenAI's Terms of Use bar using output to develop competing models, and bar automatically or programmatically extracting data or output from ChatGPT.²
- **The US government treats it as theft.** In April 2025 the House Select Committee on the CCP concluded it was highly likely that DeepSeek used unauthorized distillation of US models, which it called unlawful.³
- **The allegations are ongoing.** In February 2026, OpenAI told the same committee that DeepSeek employees had built code to obtain outputs for distillation programmatically, using third-party routers to evade access restrictions.^{4, 5}

My goal was narrower than distillation. I did not want to copy the model, only to understand its evaluation criteria. But scraping ChatGPT at scale would still have broken its terms. As a result, the method had to work through ordinary, manual conversations rather than programmatically.

2.3.2 Automated Extraction Ages Quickly

As of this writing, each major LLM has its own priorities, architecture and results, and each is changing rapidly. Programmatic discovery across them is like taking one photograph of three express trains on separate tracks: by the time you have studied the picture, every train is somewhere else. Results tied to a specific model version are likely to be invalidated when the next version ships, often within months.

The evaluation criteria and metrics are a different matter. They come largely from the category's public content, such as analyst reports, buyer's guides, product documentation and reviews, and that content changes slowly. The bright side is that uncovering the criteria is not only more straightforward, it also stays valid far longer.



3 TEST METHODOLOGY

3.1 TEST OVERVIEW

Each test fed one piece of Securin collateral to a fresh chat, asked the LLM for the same competitive battle card, and compared its Harvey ball scores against a no-collateral baseline.

3.2 TEST DESIGN

The competitive study produced a battle card that scored each feature of Securin and Cisco Kenna with a Harvey ball. To compare runs, I treated each ball as a number.

Harvey ball	Score
Empty	0
Quarter	1
Half	2
Three-quarter	3
Full	4

The test loop. I needed fast iteration: give the LLM new Securin information, see the result, refine, repeat. Each run followed the same steps:

1. Open a new chat, so nothing from earlier runs carries over.
2. Provide one piece of collateral containing Securin's verified metrics.
3. Ask for the battle card by pasting the same prompt each time from a saved prompt document, so the wording never varied between runs.
4. Record each feature's Harvey ball for both products.
5. Delete the chat, then refine the collateral or switch formats and run again.

The baseline. The initial run used no collateral at all. The LLM scored both products from its existing knowledge of each company and product, presumably drawn from public sources, which in Q4 2025 may have included its own live web search. That run gave each format something to beat.

The formats. I tested three collateral types: a product specification, a white paper and a buyer's guide. The aim was a rough experimental feel for which format the LLM read most effectively, not a controlled study.

What this design measures, and what it doesn't. Supplying collateral in the chat does not train the model; it places the content in that conversation's context. So, each run showed how well a format communicated Securin's strengths to an LLM that had it in front of it. That makes it a fast proxy for content quality. It does not show how the LLM will rank the products for buyers who never see the collateral. That requires publishing the content and testing in clean sessions, which this report does not cover.

Test conditions. All runs used ChatGPT in Q4 2025, with whatever default model the app served at the time; the exact version was not recorded. I ran about 10 complete tests, iterating over several weeks. Because model versions can change without notice, results from different weeks may reflect different underlying models.



3.3 TACTIC 1: MAP THE EVALUATION CRITERIA

Working back from the LLM's scores to the criteria behind them showed which facts Securin needed to supply. The project began with a simple request: I asked ChatGPT to compare Securin and Cisco Kenna, summarizing each feature with Harvey balls. Securin's scores did not look accurate, so I set out to learn how the LLM arrived at them.

1. **Establish an industry-wide feature baseline.** Battle cards were needed for every major competitor, so the baseline had to cover the whole market, not one rival. I asked: "Which specific features are most important for vulnerability management products across the entire market?"
2. **Group the features into categories.** I asked the LLM to organize the features into categories, which made the analysis more intuitive to read and compare.
3. **Get the evaluation metrics for each feature.** I asked the LLM which metrics are used to evaluate that feature across products. I treated its answer as a hypothesis about what drives the Harvey ball scores, to be tested in the tactics that follow.
4. **Compare the features with Securin's public footprint.** I checked each feature against Securin's data sheets, website, press releases and other public sources. Most features were not listed at all, because they were too detailed for typical marketing collateral. That gap likely explains the low scores: the LLM had no public Securin figures to draw on.
5. **Quantify Securin against each criterion.** I took the criteria to Securin's technical teams, who provided Securin's value for each feature.

Avoiding bias: Cisco Kenna's Harvey balls came only from the LLM's own comparison. I did not ask the LLM for Kenna's scores or underlying values on each feature, so the criteria and Securin's figures were not framed around the competitor's numbers.

With the criteria and Securin's values in hand, I was ready to test which reference formats communicated them best (Tactic 2).



3.4 TACTIC 2: CONTENT EFFECTIVENESS TESTING

The same Securin figures moved the LLM's Harvey ball scores most when presented in a neutral buyer's guide, and least when appended to a data sheet.

Hypotheses. This stage tested one question: which document format makes Securin's verified figures count most with the LLM?

Formats tested. I put the same feature tables into three formats, then ran each through the test loop described under Test methodology:

Format	What it contained	Effect on Harvey ball scores
Buyer's guide	Feature tables with a neutral discussion of each feature	High
White paper	Feature tables with a brief marketing-oriented discussion	Medium to low
Product specification	Feature tables added to the bottom of the standard data sheet	Low

I judged each format by how much it changed the competitive matrix: how many Securin features moved up, and by how many Harvey ball steps.

What changed: Every format improved Securin's scores. Specifying more features and values raised Harvey balls across the board, in some cases from empty all the way to full. That fits the Princeton GEO study, which found that adding statistics and cited sources raised a source's visibility in generative answers by up to roughly 40% in its benchmark.⁶ The buyer's guide, however, lifted more features to higher Harvey balls than the other two formats. The difference was clear from a side-by-side look at each format's feature chart, so I did not calculate it precisely.

Why the buyer's guide likely won. Its neutral tone and explanation of each feature made it read as reference material rather than promotion. The data sheet held the same numbers, but appended tables with little context gave the LLM less to connect each figure to its evaluation criteria.

These results show how well each format communicated Securin's figures when the LLM had the document in front of it. Whether the buyer's guide also moves rankings once published is a separate test, outside the scope of this report.



3.5 TACTIC 3: RERUNNING THE COMPETITIVE COMPARISON

With the buyer's guide as input, the rerun scored Securin far higher than the original comparison, higher than I would publish in a formal comparison.

Method. I closed the previous chat and started a new one. I supplied the buyer's guide and asked the LLM to rerun the original competitive analysis of Securin versus Cisco Kenna.

Result. Securin's evaluation improved sharply, with higher Harvey balls across its features. In fact, the competitive scores were high enough that I would have toned them down before using them in any formal comparison.

What the overshoot shows. With the buyer's guide as the main source in the conversation, the LLM leaned heavily on it, while Cisco Kenna's scores rested on what the model already knew or found through its own search. The rerun confirms the guide communicated Securin's strengths. The guide can correct an unfairly low ranking, though the method can overshoot, so the content must stay accurate.

What would confirm it in future tests. Publish the buyer's guide, allow time for it to be indexed, then rerun the comparison in clean sessions with nothing attached and compare against the original baseline.



4 CONCLUSIONS

4.1 SUMMARY OF RESULTS

A neutral buyer's guide carrying verified product figures did the most to improve how ChatGPT scored Securin against Cisco Kenna. Each format improved the scores, and the rerun with the buyer's guide overshoot what a formal comparison would support, so it marks an upper bound rather than a forecast.

4.2 COMPARING RESULTS TO PUBLISHED RESEARCH

The same verified product figures changed ChatGPT's competitive scores most when presented in a neutral buyer's guide, and least when appended to a data sheet. Published research puts these results in context, as shown below.

Finding	Figure	Source	Caveat
Content changes can raise visibility in generative answers	Up to about 40%	Aggarwal et al., KDD 2024 ⁶	Measured on a research benchmark; results vary by domain
AI Overviews show on a minority of Google queries	6.5% (Jan 2025) to 24.6% (Jul 2025) to 15.7% (Nov 2025)	Semrush ⁷	Share of keywords, not of searches
Queries that trigger AI Overviews skew informational	91.3% informational (Jan 2025), falling to 57.1% by Oct 2025	Semrush ⁷	This is the share of triggered queries, not the share of informational queries that trigger one

Google states there are no extra requirements, special schema or AI text files needed to appear in AI Overviews or AI Mode.⁸ Standard SEO fundamentals apply. That is why this report focuses on content, not markup.

4.3 PRODUCT MARKETING INSIGHTS

Now more than ever, core product marketing discipline matters. Each feature and its corresponding value needs to be understood, prioritized in a release, and delivered to the market.

Once the features are delivered, the next challenge is making sure LLMs pick up the updates. For startups without the resources of large companies, that process needs to be something product marketing can run on its own.

Fortunately, a buyer's guide approach seems effective as of this writing, based on these tests. But the LLM must be able to access and index the content. So the buyer's guide must be publicly available on a website, whether the company's own, an industry site or a buyer's guide repository, for this technique to work.



AEO / GEO for High Tech Products

As a final product marketing note, this approach can help a startup competing with a larger, more established firm correct an unfairly low ranking, rather than accept one that reflects a rival's larger budget. Because the method can also overshoot, the content must stay accurate and credible.

4.4 AI APPLICATION THOUGHTS

This methodology was performed largely by hand, as an experiment. Codified into an application or a set of agents, each step above could run systematically, through official APIs and within each provider's terms. That project may be the inspiration for a future white paper.



5 BIBLIOGRAPHY

1. Hinton, G., Vinyals, O. and Dean, J. [Distilling the Knowledge in a Neural Network](#). arXiv:1503.02531, 2015.
2. OpenAI. [Terms of Use](#). Effective Jan 1, 2026.
3. Mintz. [House Select Committee Publishes Report on DeepSeek](#). Apr 24, 2025.
4. Reuters, via Yahoo Finance. [OpenAI says China's DeepSeek trained its AI by distilling US models, memo shows](#). Feb 12, 2026.
5. Burnham, J. [OpenAI Alleges China's DeepSeek Stole its Intellectual Property to Train its Own Models](#). FDD, Feb 13, 2026.
6. Aggarwal, P. et al. [GEO: Generative Engine Optimization](#). KDD 2024. arXiv:2311.09735.
7. Garanko, J. et al., Semrush. [Semrush AI Overviews Study](#). Updated Dec 15, 2025; data Jan to Nov 2025.
8. Google Search Central. [AI features and your website](#). Updated Dec 10, 2025.

AI-use note: I used AI assistants to research and edit this report. The experiments, results and conclusions are mine, and I checked every cited figure against the source linked above.



6 ABOUT MICHAEL WILLIAMS

Michael Williams is an accomplished product marketing leader with a passion for driving exceptional business outcomes through impactful content. A diverse background as security sys admin, developer, sales, BD, product management, and product marketing enables him to collaborate effectively across the company.

His core product marketing differentiator is establishing the path to and removing barriers to revenue growth. He has established four new market and product categories, taken over 20 products from entry to #1 market and/or thought leadership, helped grow horizontal markets to over \$1B, provided content for campaigns that exceeded goals by 300%, and driven revenue gains of 30% QoQ over consecutive years.

Key Links:

- LinkedIn: <https://www.linkedin.com/in/williamsma/>
- Portfolio: <https://revgrow.pro/portfolio>
- GitHub: <https://github.com/michaelrobotai>