SearchPilot

Generative Engine Optimization (GEO): Stop Guessing and Start Testing
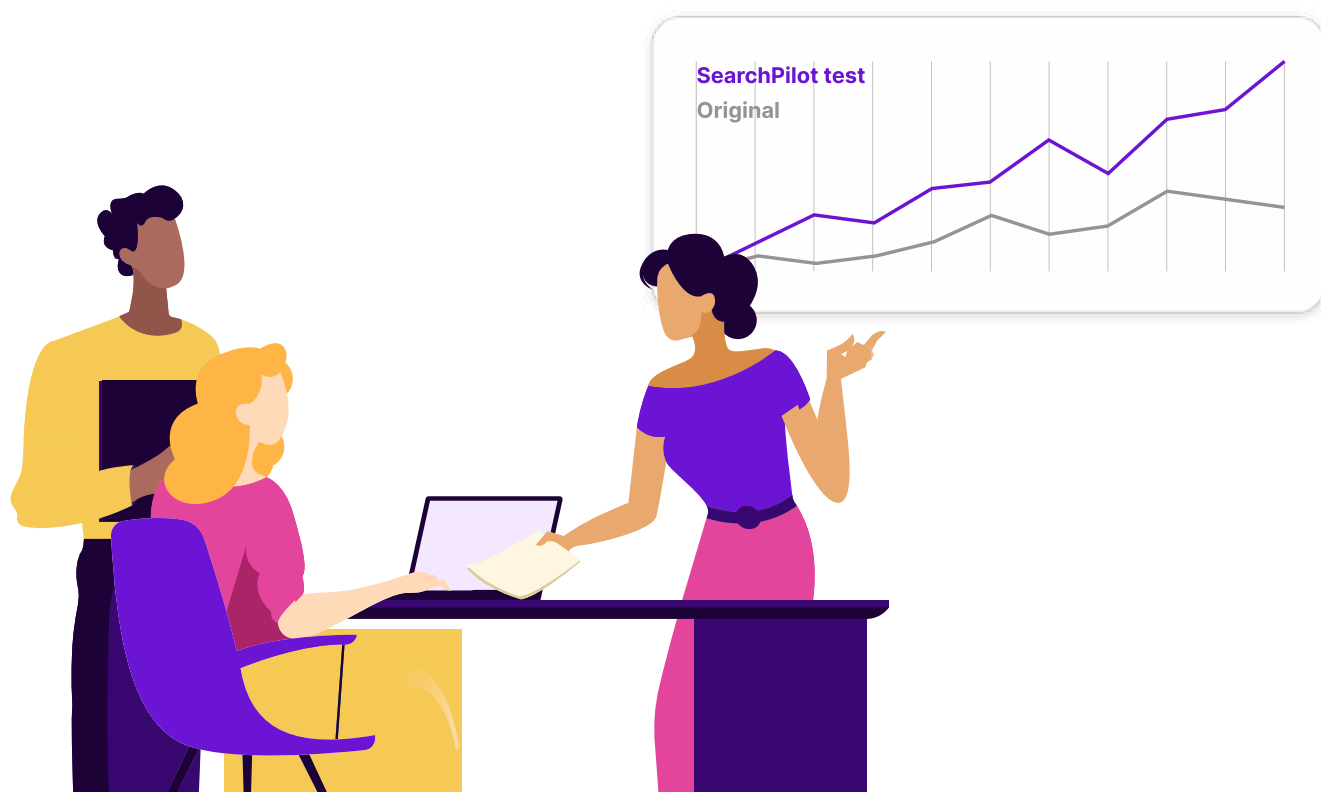
A Leader's Guide to GEO A/B Testing for Ecommerce

Your Biggest Channel is Changing. Are You Ready?

For years, senior leaders have known a simple truth: **search is your biggest channel, but it's the least understood**. You've built robust SEO strategies to compete on a familiar battlefield. But that battlefield is changing.

With the rise of AI-powered search, the customer journey is fundamentally shifting. More research, comparison, and consideration now happens within the AI model, long before a user ever clicks through to a website. This presents a new challenge: as the traditional search results page fades in importance, **search is going dark**. How do you influence a journey you can no longer see?

For large ecommerce businesses managing thousands of products, where freshness of price, stock, and reviews is paramount, this uncertainty creates risk. The answer isn't to guess what the new algorithms want. Stop trying to predict the future. experiment to discover it.

Contents

From SEO to GEO: What is Generative Engine Optimisation?

Generative Engine Optimisation (GEO) is the evolution of SEO for the age of AI. It's the practice of optimising your website not just to rank in a list of links, but to be found, understood, and presented favourably within AI-generated answers and summaries.

While some influence can happen by being part of an AI's slow-moving training data, the real opportunity for dynamic ecommerce sites lies elsewhere. AI models need up-to-the-minute information, which they get by directly accessing URLs and conducting their own searches. This is where you can have an impact right now.

Your goal is to make your product and category pages the most reliable, authoritative, and compelling source of information for the AI as it gathers data to answer a user's query.

How GEO Testing Evolves From SEO Testing

If you're familiar with the **SearchPilot approach to SEO A/B testing**, the core methodology will feel familiar. This isn't about throwing out what works; it's about adapting a proven system for a new reality.

What Stays the Same

1. **The Core Mechanism:** The scientific approach of using control and variant page groups to measure impact remains the gold standard. You still need statistical rigour to know what truly works.
2. **The Testing Ground:** Your most important pages - Product Detail Pages (PDPs) and Product Listing Pages (PLPs) - are still the primary focus for optimisation.

The Critical Difference: A Focus on the Last Click

The biggest shift is in what you're optimising for. In traditional SEO, you might test for a higher click-through rate from a list of ten blue links. In GEO, the AI handles much of that initial research. By the time the user reaches your site, they have already read a range of summaries and comparisons.

The click you earn from an AI answer is therefore much further down the funnel. It's often the **final, decisive click** from a user who is ready to transact. This means your testing focus shifts from top-of-funnel visibility to bottom-of-funnel influence. You're no longer testing against a single search result, but against an entire AI-driven research journey.

How it Works: What Can We Influence?

LLMs present information from two conceptually-different places:

1. Some information is **"learned"** during the creation of the core model from the original training data and reinforcement learning. In the current state of the art, this is very static knowledge and there is a balance that the providers need to strike between increasing what we might call "intelligence" and "knowledge" in the training of their AI models.

2. Fresh, specific, time-sensitive or obscure information and deeper "knowledge" is **"piped in"** to the model at conversation time as augmented context. This process is called RAG from Retrieval Augmented Generation. This augmentation can come from sources the user specifies, from web searches carried out by the AI system, or from accessing specific web pages sourced from a variety of places.

The slow feedback loop, and batch update nature of changes to models' learned information makes it largely impossible to run statistical tests to isolate all the variables and tell if a particular website change has improved your presence in the training data in such a way that it is more likely to recommend your brand or website without accessing external search data. These step changes (e.g. from GPT-4 to GPT-5) are increases in intelligence more than increases in knowledge, however, and so are mainly independent of our efforts to optimise for AI discovery. This article is an excellent overview of the **trend towards more intelligence rather than more knowledge** and the reasons behind it.

Instead, GEO A/B testing focuses on the ways we can influence the RAG. This is beneficial because:

1. **It's real-time** - each conversation a user has with the LLM uses the latest search indexes and results behind the scenes.

2. **It's the primary source of ecommerce influence** - the history of Google has taught us all the importance of freshness in commercial queries - price, offers, reviews, ratings, stock levels and more are all real-time data that an AI wouldn't expect to "know" from its training and instead would take from its augmented data from Retrieval Augmented Generation. This augmentation can come from sources the user specifies, from web searches carried out by the AI system, or from accessing specific web pages sourced from a variety of places.

Crucially, the models don't do just a single search. They augment their training with a whole buyer's journey of searches based on a "fan out" set of queries - starting from the user's input prompt and making their own determination of what to research. GEO testing can therefore potentially affect **four levers** of influence:

1. Ranking better in the search queries performed by the AI.

2. Ranking for new keywords to increase coverage across the fan-out queries performed by the AI.

3. Appearing in a more compelling way in the search results returned to the AI [noting that this process is opaque, and we don't know whether individual products receive just a list of links, information that looks more like a search results page, or full page content for the ranking pages].

4. Improving the pages themselves so that when the models ingest it, its information is more likely to be included prominently in the output for the user.

   a. Different products likely carry out this step differently - with Google's Gemini likely getting a more direct feed of page content from Google's index, compared to other tools that are less connected to a search engine provider that may visit the page directly. In the latter case, it's worth remembering that many **do not execute JavaScript**.

Note that the outputs of LLMs don't constitute "search results" in the same way that we are used to, and there isn't the same concept of "ranking" within a conversation. Hypotheses and tests are aimed at a combination of being featured in more conversations, being featured more prominently, or in a more compelling way.
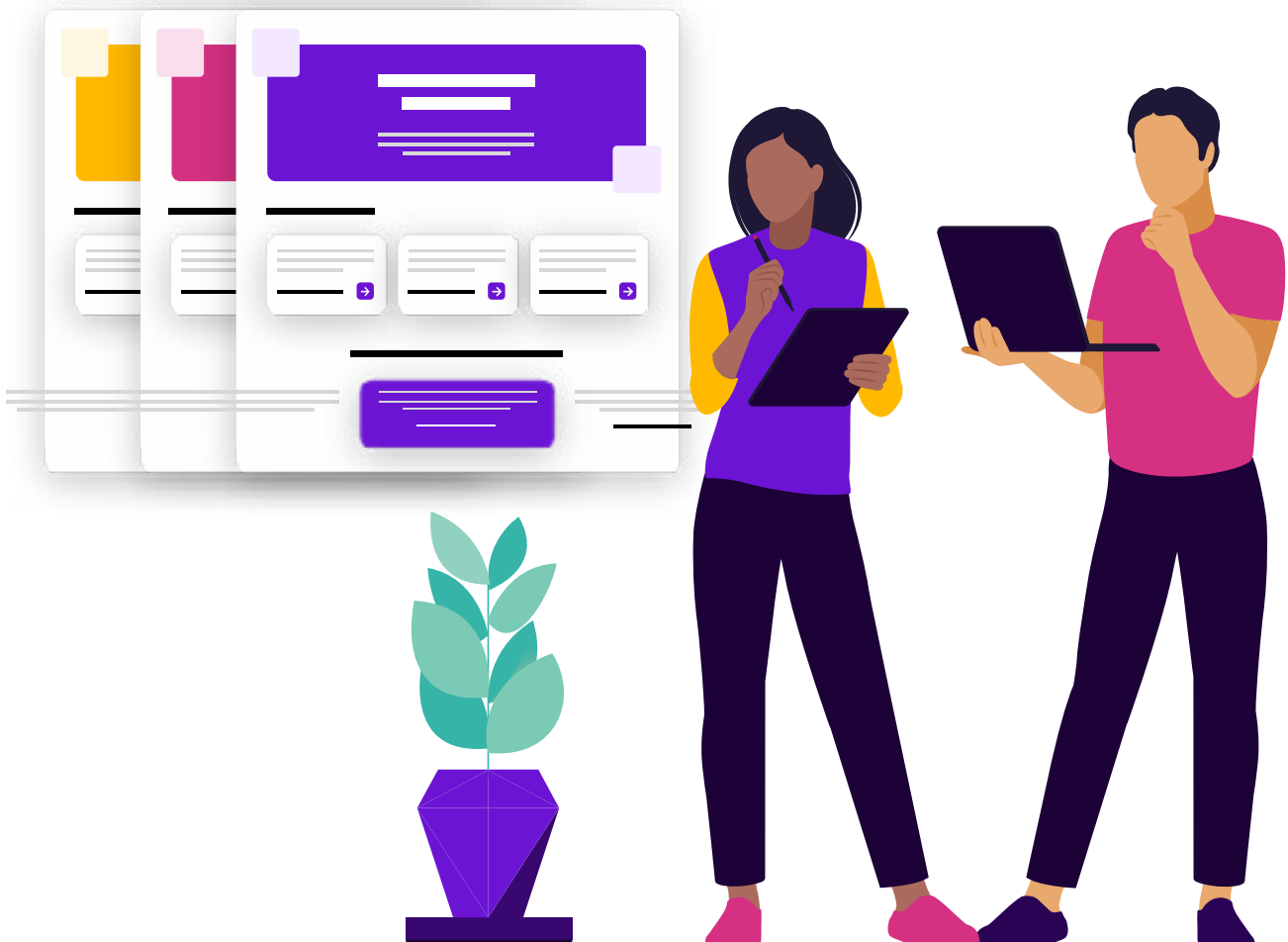
Putting GEO Testing Into Practice: Moving from React Mode to Control Mode

Without a testing methodology, adapting to AI search is just guesswork - you're in **React Mod**e making changes and hoping for the best without a view of the trade-offs between SEO and GEO. GEO testing moves you to **Control Mode**, allowing you to build a playbook for what actually influences AI-driven traffic and revenue.

The hypotheses are similar in structure, but different in intent. Measurement, tracked through your existing web analytics, is focused on observing uplifts in qualified, high-intent traffic.

Here are two examples of potential GEO tests:

1. **Hypothesis for a Product Detail Page (PDP)**

   - **Goal:** To be the product chosen and recommended in an AI-generated comparison (e.g., "compare the best running shoes for beginners").

   - **Potential Hypothesis:** "By reformatting our product specifications into a concise, scannable 'Key Features' summary at the top of the PDP, we hypothesize that generative engines will more frequently cite our product in comparison answers, leading to an increase in qualified last-click traffic."

2. **Hypothesis for a Product Listing Page (PLP)**

   - **Goal:** To become the AI's preferred source for a category-level query (e.g., "where's the best place to buy 4K TVs under £1000?").

   - **Potential Hypothesis:** "By adding a clear, authoritative introductory paragraph above the product grid on our category pages, we hypothesize that AI models will use our page as a primary source for 'best of' queries, resulting in more citations and an uplift in high-intent traffic to the entire category."

Stop Trying to Predict the Future. Experiment to Discover it

The rise of generative AI is the most significant shift in search in two decades. It brings uncertainty, but it also brings opportunity for those prepared to adapt. While your competitors are guessing, you can be learning.

By evolving your SEO testing program to include a GEO testing program as your LLM-powered traffic grows, you can systematically reduce risk and find a competitive advantage.
The world of GEO is new and evolving. For a deeper discussion on how your SEO team can prepare for this new challenge, **contact us**.

SearchPilot supports GEO A/B testing

We are excited to announce that SearchPilot now supports GEO A/B testing, empowering you to move from guesswork to a data-driven strategy for AI search. While competitors are reacting and hoping for the best, you can be in control, building a playbook for what actually drives high-intent traffic and revenue in this new landscape. The battlefield is changing, and with GEO testing, you can stop trying to predict the future and start experimenting to discover it.
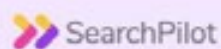
The rise of AI-powered search is fundamentally changing your most important revenue channel. As more customer research happens within AI models before anyone clicks to your site, the traditional search journey is becoming less visible - in essence, search is going dark. For large e-commerce businesses managing thousands of products and many categories, this growing uncertainty creates significant risk and makes it harder to influence a journey you can no longer see.

The answer to the uncertainty, speculation and misinformation is to adapt a proven system for this new reality. Generative Engine Optimisation (GEO) A/B testing applies the same statistical rigor of SEO testing to this new AI-driven world. The focus shifts from optimizing for initial visibility towards ensuring your product and category pages are the most compelling source of information for an AI model as it guides a user toward a purchase.

If you'd like to learn more about **how GEO A/B testing works, check out our detailed guide**.

If you'd like to speak to the SearchPilot team about GEO testing, **get in touch**.

(You might notice that along with the debate about what works and how optimizing for AI recommendations differs from regular SEO, there has been significant discussion in the industry about what to call this emerging practice. I've heard GEO (for Generative Engine Optimization), AIO (for AI Optimization), LLM Optimization (because LLMO is too silly even for those advocating this name), and even AISEO which, honestly, sounds like a verse of Old MacDonald. We're going with GEO for now, because it captures the essence of the difference which is that these "generative engines" are a fundamentally different user experience more because of the mode of interaction they enable, and less because there is an AI under the hood.)

It's early days, and you need to be getting enough clicks from these new sources in order to be able to run experiments against them. We are just starting this process with our customers, but you can be sure that we will be publishing interesting case studies as soon as we can. Sign up below to be the first to hear!