

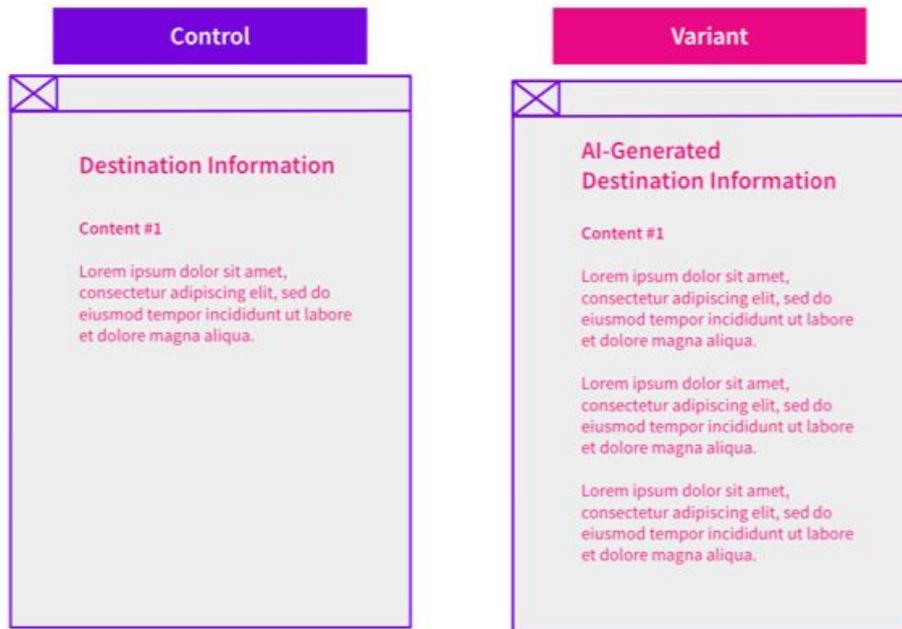
# AI Content Is Cheap. Lost Traffic Isn't.

A Roadmap for Testing What Works

@searchpilot

# Early AI content testing

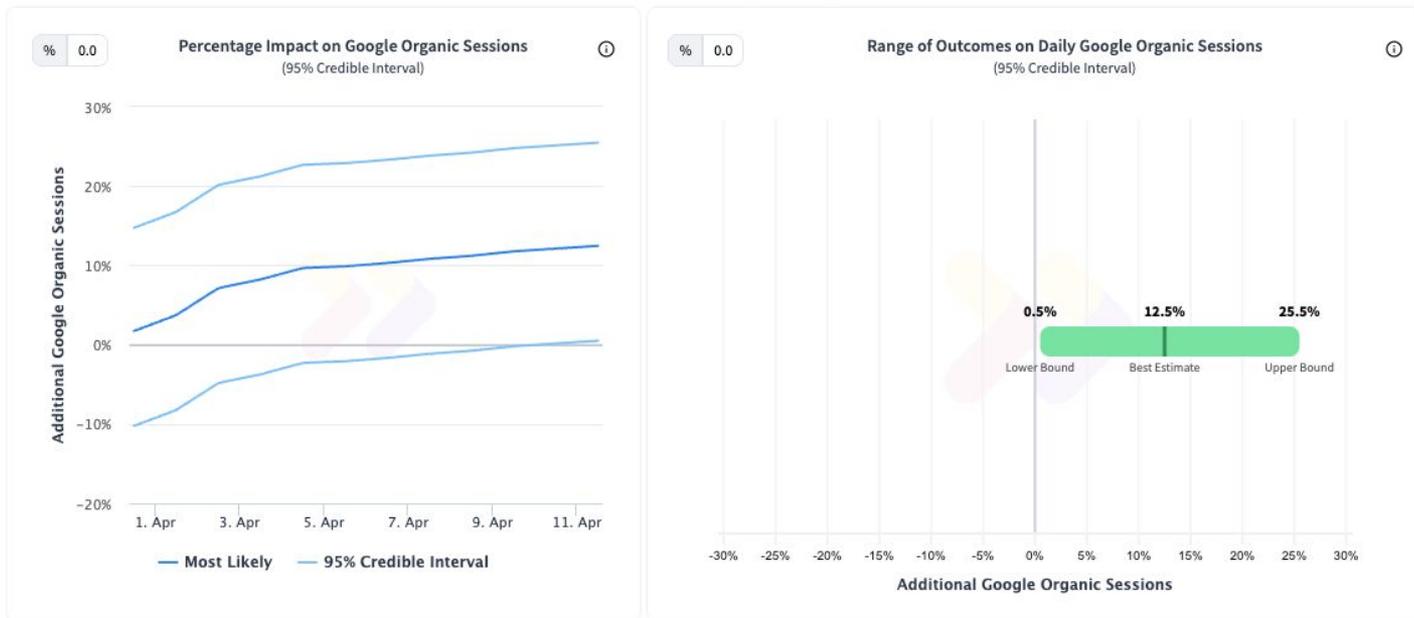
Can adding AI-generated content to pages improve organic traffic?



# Early AI content testing

Can adding AI-generated content to pages improve organic traffic?

Estimated Additional Google Organic Sessions



**We're ready to ask better  
questions**

# Your roadmap to quality content

Checking  
value

Choosing  
models

Choosing  
prompts

Content  
optimization

Translating/  
localizing



# It starts with the hypothesis

- Come up with the question you want to answer *first*
  - Is this content element valuable?
  - Am I using the best model to create my content?
  - Am I using prompts that output useful, relevant content?
  - Is my content the right length?
  - Is my content targeting the right keywords?
  - Should I add a disclaimer that content was created with AI?



# It doesn't end with the first test result

- Dig deeper into the insights you get from the first round of testing
  - How does this content compare to content created with a different model?
  - Is this content in the best location on the page?
  - Will this content perform well if it's translated into another language?
  - Will a more detailed or specific prompt give me content that performs better?
  - Are there semantic searches or other new forms of user behavior that may be impacted by this content?



**Checking the value of your content elements**

# Proving you can benefit from a content refresh

- Before we spend internal resources on generating AI or human-written content, we usually start with a series of tests proving that a content refresh is useful
- If we remove a content element, hide it on page load, or move it down the page and see an **inconclusive** or **positive** result, that tells us:
  - The content may be stale
  - The content may be unhelpful
  - The content may not be relevant to the purpose of the page
  - The content element may be pushing more useful or relevant content too far down the page

Test idea

**Test series proving that a page content element is ready for a refresh**

Test idea - Test 1 in series

**Remove the content from the page**

Test idea - Test 2 in series

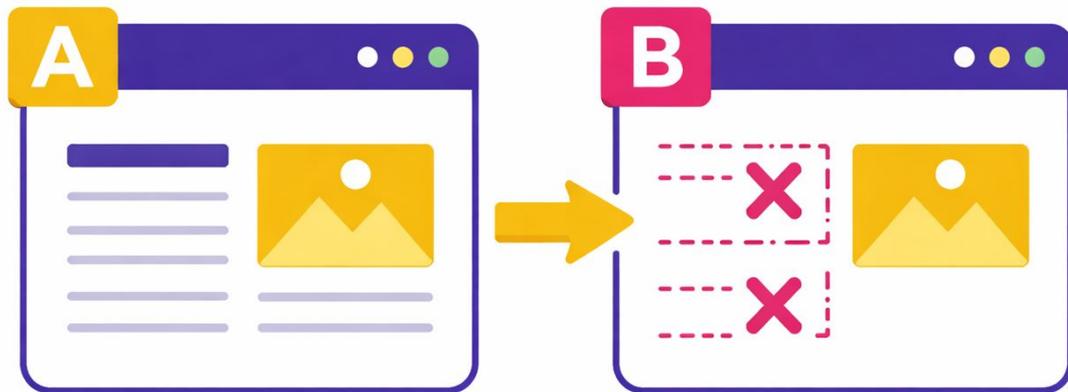
**Move the content to the bottom of  
the page**

Test idea - Test 3 in series

**Hide the content on page load**

# Consider your insights

- Have we shown that changes to this content element can impact traffic?
- Are we confident that this content element is sending the right signals with its visibility and placement?
- Is this the most high-impact content element we can test right now?
- Do we need any other data to create our testing roadmap?



# Get the full picture with multi-metric testing

Your content doesn't just impact organic clicks. Find out what's happening beyond the 10 blue links:

[searchpilot.com/resources/blog/product-search-testing](https://searchpilot.com/resources/blog/product-search-testing)

## Range of Outcomes on Daily Google Organic Sessions

(95% Credible Interval)



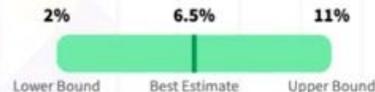
## Range of Outcomes on Daily Google Merchant Center Clicks

(95% Credible Interval)



## Range of Outcomes on Daily Google Merchant Center Impressions

(95% Credible Interval)



# Consider full-funnel testing

- Your content may be created for SEO purposes, but users can still see it



# Consider full-funnel testing

- Add CRO to your test cycle to be sure that your content isn't harming the user experience



# Learn directly from the experts

Sign up for our free SEO testing course to get a crash course in developing a testing program: [searchpilot.com/seo-testing-course](https://searchpilot.com/seo-testing-course)



## Your Instructor

**Craig Bradford, COO of SearchPilot**

At SearchPilot, I get to see thousands of SEO tests on all kinds of websites. I've learned what works and what doesn't. I want to help more marketers use data to show the value of their work. This course is my way of sharing what I've learned so you can make decisions that have a real impact.

# Comparing LLM models

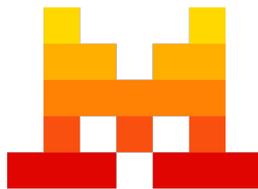
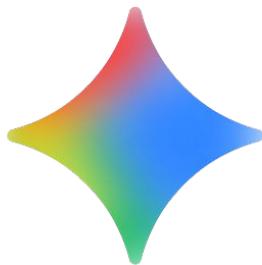
# Choosing the right text generation model

- If you're only sticking with ChatGPT or Gemini, **you may be missing out on the right model for your needs**
- If you're only using one model from one company, **you may be missing out on more specialized or lightweight options**



# Choosing the right text generation model

- A short list of companies and organizations working on text generation models right now:
  - OpenAI
  - Google
  - Anthropic
  - Meta
  - Mistral
  - X
  - DeepSeek
  - Microsoft



# Choosing the right text generation model

**Frontier models** OpenAI's most advanced models, recommended for most tasks.



## GPT-5.2

The best model for coding and agentic tasks across industries



## GPT-5 mini

A faster, cost-efficient version of GPT-5 for well-defined tasks



## GPT-5 nano

Fastest, most cost-efficient version of GPT-5



## GPT-5.2 pro

Version of GPT-5.2 that produces smarter and more precise responses.



## GPT-5

Previous intelligent reasoning model for coding and agentic tasks with configurable reasoning effort



## GPT-4.1

Smartest non-reasoning model

---

**Open-weight models** Open-weight models under a permissive Apache 2.0 license.



## gpt-oss-120b

Most powerful open-weight model, fits into an H100 GPU



## gpt-oss-20b

Medium-sized open-weight model for low latency

Test idea

**Using the same prompts, generate page content using models from 3 different companies**

Test idea

**Using the same prompts, generate page content with 3 different models from the same company**

# Refining LLM prompts

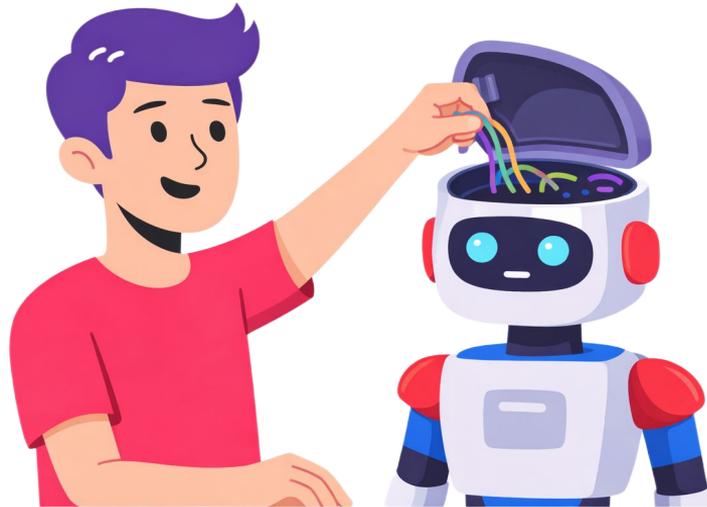
# Are your prompts working as intended?

- **Remember:** "AI" isn't true artificial intelligence
- Large language models use sophisticated algorithms trained on massive datasets to predict the likelihood of one word following another
- No AI model has a truly human-like understanding of your product, your intent, or your expectations for its output



# Are your prompts working as intended?

- Don't just test the models you're using. Test the prompts too.
- Understand what you're putting into the model to get what you want out of it
- You can even train your own models for internal use using larger amounts of data



Test idea

**Using 3 different prompts,  
generate title tag content with the  
same model from the same  
company**

Test idea - Prompt 1

**Using this list of existing title tag content, generate new title tag content**

Test idea - Prompt 2

**Using this list of existing title tag content, generate new title tag content, adding one new keyword that does not currently exist in the title tag**

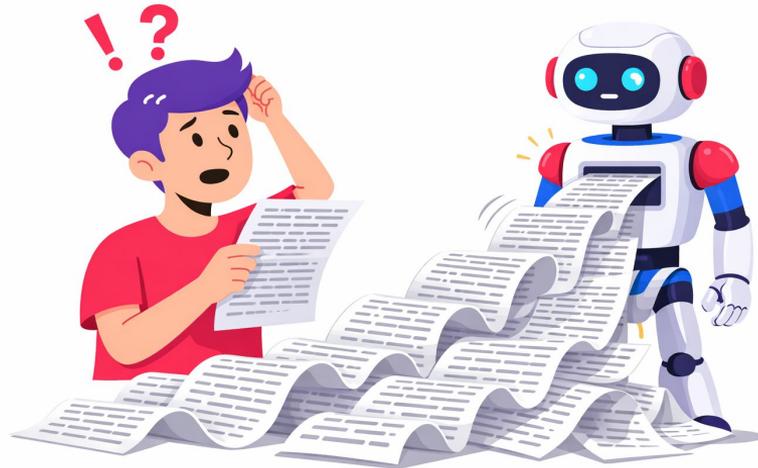
Test idea - Prompt 3

**Using this list of existing product descriptions, generate appropriate page title content for each product page**

**Optimizing your content**

# Asking the right questions

- LLMs can output far more content than you have the resources to write yourself
- As search engines expand their ability to parse and evaluate text, page content will remain a crucial ranking factor if it is useful and relevant
- Testing will give you a better understanding of whether your content is valuable



# Creating and expanding page content with AI

- LLMs can help you add more context to pages with thin content
- **Remember:** You know your own product better than an AI model ever can
  - Prompt with as many relevant details as possible
  - Add a human quality assurance step before publishing any new content to your site

Test idea

**Use AI to generate FAQ content for product pages**

Test idea

**Use AI to double the length of a content element**

Test idea

**Use AI to create a short, relevant piece of content above the fold based on a longer product description below the fold**

# Shortening content with AI

- More content isn't always better for users *and* search engines
- AI can help you:
  - Streamline content elements that are too long
  - Summarize user-created content
  - Trim irrelevant sections from page content

Test idea

**Use AI to reduce the length of a piece of page content while preserving the most relevant keywords**

Test idea

**Use AI to turn a lengthy product description into a bulleted list of key features**

Test idea

**Use AI to summarize user reviews  
on a product page**

# Translating and localizing your content

# Using AI to create content for a global portfolio

- AI is a blessing *and* a curse for sites that have to maintain content in dozens of languages
- **The upside:** You can use AI to create content in all those languages, or to translate content created in English for those languages
- **The downside:** Not all AI models are set up to handle non-English languages well, and you may get yourself in trouble if you're creating content that your team cannot read and assess for accuracy
- If you're going to use content created or translated by large language models in a language you don't speak, you need additional testing and quality assurance

# Using AI to create content for a global portfolio

- Even if you've tested enough to be confident that you've chosen the right model for English, test again in other languages
- Consider whether the model you're using was developed primarily for the language and regional market you need to use it for
- Models that perform well in US English may use archaic or incorrect language for UK, Australian, or Indian English
- Models that perform well in English may not be the right choice for French, German, Japanese, Chinese, and so forth

Test idea

**Using the same prompts in the same languages, generate page content using LLM translation models from Mistral, Gemma2, DeekSeek, and DeepL**

# Not all machine translation works the same way

## Statistical Machine Translation (SMT)

- Uses human-translated texts compared to the original language in its dataset to develop statistical models of the relationships between phrases
- Big jump in translation quality from earlier rule- or example-based methods
- Struggles with words and phrases that are rare or not present in its training data
- May not be able to output the most natural human-like language

# Not all machine translation works the same way

## Neural Machine Translation (NMT)

- Uses a neural network to predict the likelihood of a series of words in a text it is translating
- Generally considered the gold standard for *accurate* translations, with caveats:
  - Only works well in language it has been trained on with extremely large datasets
  - Translation can be too literal, missing nuance or naturalistic localization of idioms and writing styles

# Not all machine translation works the same way

## Large Language Model Translation (LLM)

- Uses Transformer decoder architecture to improve understanding of context and nuance across languages
- Generally the best for *humanlike* translations with a nuanced understanding of idioms and context, with caveats:
  - Not all models have appropriate training sets for all languages
  - Translation is *not* literal, and the model may output inaccurate or hallucinated results without being able to correct itself

Test idea

**Using the same prompts in the same languages, translate page content using Google Translate (NMT) vs. Google Gemini (LLM)**

Test idea

**Translate English-language  
human-written content to another  
language using a human translator  
vs. DeepL's NMT model**

Test idea

**Translate English-language human-written content to another language using an LLM model, then use an NMT model to compare the translation to the original text to flag and fix accuracy issues**

# Test cycle: Your roadmap to content that performs

Checking value

Choosing models

Choosing prompts

Content optimization

Translating/  
Localizing



```
graph LR; A[Checking value] --> B[Choosing models]; B --> C[Choosing prompts]; C --> D[Content optimization]; D --> E[Translating/Localizing]
```

# Test cycle: Your roadmap to content that performs

Checking value

Choosing models

Choosing prompts

Content optimization

Translating/  
Localizing

- Remove content element
- Hide content element on page load
- Move content element to a different place on the page

# Test cycle: Your roadmap to content that performs

Checking value

Choosing models

Choosing prompts

Content optimization

Translating/  
Localizing

- Test models from different companies
- Test models from the same company

# Test cycle: Your roadmap to content that performs

Checking value

Choosing models

Choosing prompts

Content optimization

Translating/  
Localizing

- Test more detailed prompts
- Test prompts that use existing content as basis for new content generation
- Prompt for specific keywords or user types

# Test cycle: Your roadmap to content that performs

Checking value

Choosing models

Choosing prompts

Content optimization

Translating/  
Localizing

- Expand content
- Reduce content length
- Compare against human-written content
- Target specific keywords
- Target semantic queries and related searches

# Test cycle: Your roadmap to content that performs

Checking value

Choosing models

Choosing prompts

Content optimization

Translating/  
Localizing

- Add NMT-translated content
- Add LLM-translated content
- Vary LLM models used for translated content
- Create custom model using NMT translation to verify LLM translation accuracy

# Test cycle: Your roadmap to content that performs

## Checking value

- Remove content element
- Hide content element on page load
- Move content element to a different place on the page

## Choosing models

- Test models from different companies
- Test models from the same company

## Choosing prompts

- Test more detailed prompts
- Test prompts that use existing content as basis for new content generation
- Prompt for specific keywords or user types

## Content optimization

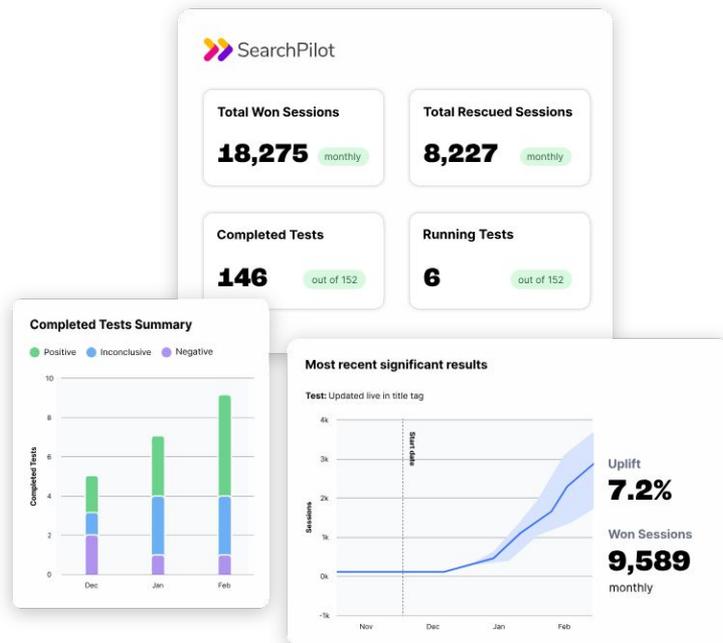
- Expand content
- Reduce content length
- Compare against human-written content
- Target specific keywords
- Target semantic queries and related searches

## Translating/Localizing

- Add NMT-translated content
- Add LLM-translated content
- Vary LLM models used for translated content
- Create custom model using NMT translation to verify LLM translation accuracy

# Ready to put your content to the test?

Book a meeting with our team at [searchpilot.com/request-a-demo](https://searchpilot.com/request-a-demo)



# Thanks!

@searchpilot

Ask me any questions you have about SEO and CRO testing in the chat!

**Book a meeting** with our team at  
[searchpilot.com/request-a-demo](https://searchpilot.com/request-a-demo)

