



TECHNICAL SEO FOR AI



Presented by
Alan Bleiweiss, Sustainable SEO Consultant

Alan Bleiweiss

[@alanbleiweiss.bsky.social](https://bsky.app/profile/alanbleiweiss.bsky.social)

<https://AlanBleiweiss.com>



ALAN BLEIWEISS



- Internet Marketing since 1995
- SEO since 2000
- Sustainable SEO Audits since 2006



Web Project Management:

- Managed World Class Web Development team
- 20 Developers, Designers & Support Staff
- Project Budgets Up to \$2,000,000



Search Engine Optimization:

Consultant, Trainer, Mentor, Speaker, Author





ALAN BLEIWEISS



Example Web Development Clients



Example SEO Clients

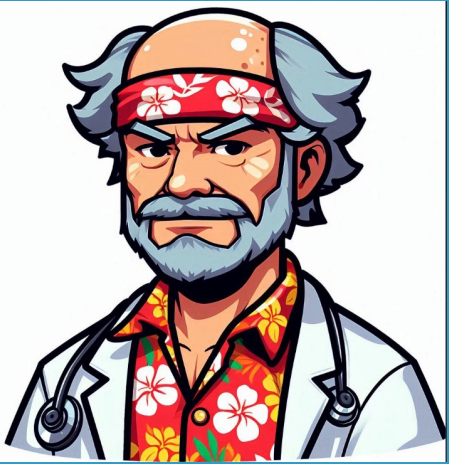


Alan Bleiweiss

[@alanbleiweiss.bsky.social](https://bsky.app/profile/alanbleiweiss.bsky.social)

<https://AlanBleiweiss.com>





TECHNICAL SEO FOR AI

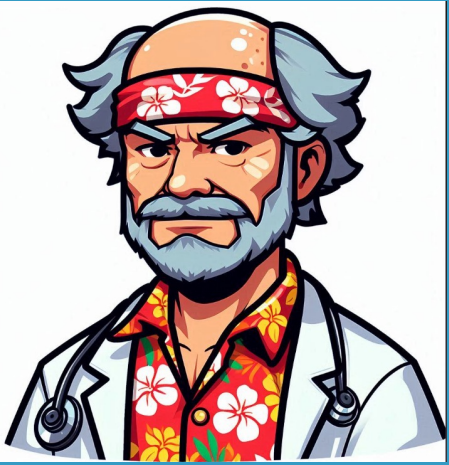


Generative AI services

- Rapidly changing
- Secretive in their approach
- Don't consistently following industry standards
- Operating on questionable legal premises

What's true today may not be true next week





TECHNICAL SEO FOR AI



Generative AI services

- Require Efficient Information Retrieval Methods
- Depend on Logic-Based Word Association Formulas
- Attempt to Understand The World “We” Created









TECHNICAL SEO FOR AI



Not Understanding Complexity of Challenges



<p>IA: INFORMATION ARCHITECTURE</p>  <p>The organizing and labeling of websites, intranets, online communities, software and more into content grouping.</p>	<p>TD: TOPICAL DISCOVERY</p>  <p>The process where search crawlers and algorithms work to find, identify, clarify and classify the primary topical focus of a web site as well as the intent of that presence.</p>	<p>TF: TOPICAL FOCUS</p>  <p>"What is this page/section/site about? What are the words associated with that?"</p>	<p>BR: BRAND REINFORCEMENT</p>  <p>Your brand is critical to your visibility, authority and trust. This leads to more interaction on the web site, and that means important value to SEO.</p>
---	--	---	---

From my "5 Super Signals of SEO" Infographic 2015

Alan Bleiweiss

[@alanbleiweiss.bsky.social](https://bsky.app/profile/alanbleiweiss.bsky.social)

<https://AlanBleiweiss.com>



SEO & GAI



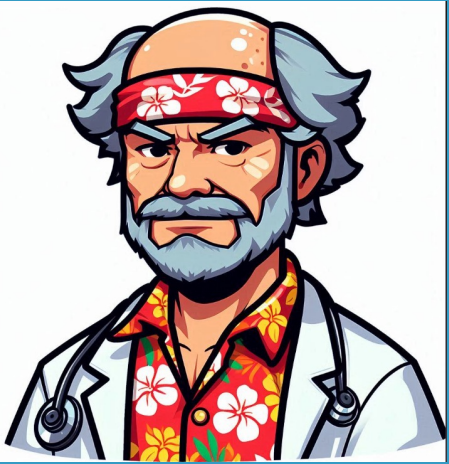
Expectation vs. Reality

AI Companies Claim Their Products Are Magic

Reality:

- Content Freshness Issue
- Hallucinations Not Going Away
- Legal Considerations of LLM Scrapers





SEO & GAI



Expectation vs. Reality

Content Freshness Issue

- GAI/LLM Service providers don't recrawl anywhere nearly as frequently as search engines
- Reliance on "RAG" is a band-aid





SEO & GAI



Retrieval-Augmented Generation (RAG)

“Retrieval-augmented generation (RAG) enables you to customize LLMs without fine-tuning, helping you save money and accelerate time to deployment.” *

*Intel Corporation

<https://www.intel.com/content/www/us/en/goal/how-to-implement-rag.html>





SEO & GAI



 Search Labs | AI Overview

Retrieval-Augmented Generation (RAG) is a framework, **not a proprietary technology owned by a specific entity**. It was developed and published in a 2020 paper by researchers at Meta (then Facebook). [🔗](#)

Here's a more detailed breakdown:

Origin:

The concept of RAG, which combines retrieval and generation, was introduced in a 2020 paper by [Patrick Lewis](#) and his colleagues at Meta AI. [🔗](#)

Purpose:

RAG allows [Large Language Models \(LLMs\)](#) to access and incorporate information beyond their training data, enabling them to answer questions and generate text with more context and accuracy. [🔗](#)

Not Proprietary:

RAG is a framework or technique, not a product or service owned by a single company. It's a general approach that can be implemented by various organizations and developers. [🔗](#)

Implementation:

Many companies and researchers are exploring and implementing RAG, including [NVIDIA](#) and [IBM](#). [🔗](#)



Alan Bleiweiss

[@alanbleiweiss.bsky.social](https://alanbleiweiss.bsky.social)

<https://AlanBleiweiss.com>



SEO & GAI



Expectation vs. Reality

Legal Considerations

- Who owns content LLM services scrape?
- Are you better off blocking LLM/GAI bots?





SEO & GAI



Get Found in GAI Output

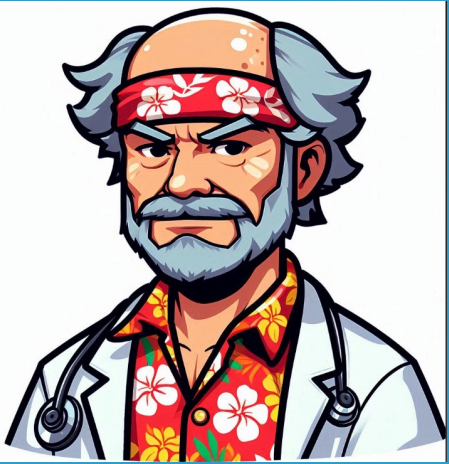


The screenshot shows a search interface for Perplexity AI. The search query is "Is Alan Bleiweiss an author at Search Engine Journal? Wh...". The main result is titled "Who is Alan Bleiweiss?". It includes a "Sources" section with three items: "Alan Bleiweiss Consulting - ..." (linked to linkedin), "Alan Bleiweiss - SEO Hero" (linked to seoheronews), and "Interview with Alan Bleiweiss: Tips fro..." (linked to sitechecker). Below this is a "Perplexity" section with a paragraph of text: "Alan Bleiweiss is a seasoned SEO consultant, speaker, trainer, and author with over 20 years of experience in internet marketing 3 1. He specializes in forensic SEO audits and provides consulting services to a wide range of clients, from mid-size to large-scale, including Fortune 500 companies 1. Bleiweiss focuses on helping clients improve their organic search visibility and address issues impacting website performance 4. He is known for his expertise in identifying and resolving complex SEO challenges, with a focus on sustainable SEO practices 1 4." On the right side, there is a "Shareable" button, a "Share" button, and a "Search Videos" button with a plus sign. Below that is a "Generate Image" button with a plus sign and a "PRO" label.

Alan Bleiweiss

[@alanbleiweiss.bsky.social](https://alanbleiweiss.bsky.social)

<https://AlanBleiweiss.com>



SEO & GAI

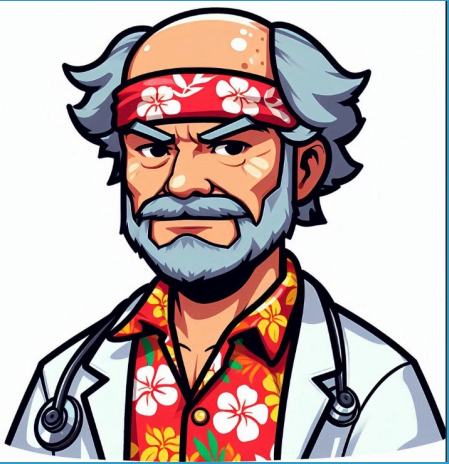


Get Found in GAI Output

Many SEO Best Practices are essential

- Crawl efficiency
- Overall Information Architecture
- Topical Organization
- Topical Focus
- AVOID Client-Side JavaScript (CSR-JS)





SEO & GAI



Get Found in GAI Output

LLMs depend on technical aspects of Information Retrieval

- Crawl Efficiency
- Frictionless Internal Linking
- Topical Relationship Mapping
- Context Reinforcement





SEO & GAI



Get Found in GAI Output

LLMs don't currently process Schema markup the way search engines do.

Use Schema Anyway

- Should already do it for SEO
- It's already established IR refinement
- LLMs adapting Schema more (especially JSON)





SEO & GAI



Get Found in GAI Output

LLMs.txt

The /llms.txt file

A proposal to standardise on using an `/llms.txt` file to provide information to help LLMs use a website at inference time.

AUTHOR

Jeremy Howard

PUBLISHED

September 3, 2024

<https://llmstxt.org/>

Alan Bleiweiss

[@alanbleiweiss.bsky.social](https://alanbleiweiss.bsky.social)

<https://AlanBleiweiss.com>





SEO & GAI



Get Found in GAI Output

How to Create an llms.txt File for Any Website

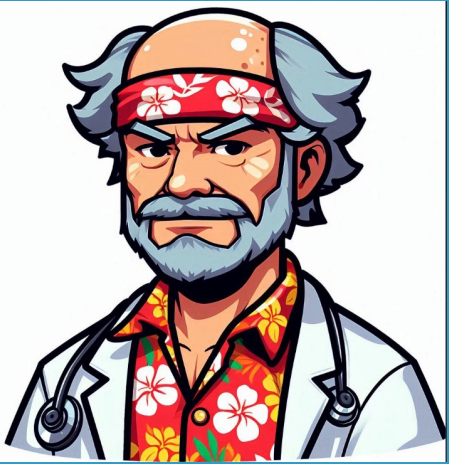
<https://www.firecrawl.dev/blog/How-to-Create-an-llms-txt-File-for-Any-Website>

Alan Bleiweiss

[@alanbleiweiss.bsky.social](https://bsky.app/profile/alanbleiweiss.bsky.social)

<https://AlanBleiweiss.com>





SEO & GAI

The Role of GAI in Organic Search



Get Found in GAI Output

A guide to Semantics or how to be visible both in Search and LLMs.

by Gianluca Fiorelli

<https://bsky.app/profile/gfiorelli1.bsky.social>

<https://www.iloveseo.net/a-guide-to-semantics-or-how-to-be-visible-both-in-search-and-llms/>

Alan Bleiweiss

[@alanbleiweiss.bsky.social](https://alanbleiweiss.bsky.social)

<https://AlanBleiweiss.com>





SEO & GAI



Stay Out of AI Output

Robots.txt – Prevent LLM access



User-agent token in robots.txt

Google-Extended

Example robots.txt group

```
user-agent: Google-Extended
allow: /archive/1Q84
disallow: /archive/
```

Google-Extended is a standalone product token that web publishers can use to manage whether their sites help improve [Gemini Apps](#) and Vertex AI generative APIs, including future generations of models that power those products. Grounding with Google Search on [Vertex AI](#) does not use web pages for grounding that have disallowed Google-Extended. Google-Extended does not impact a site's inclusion or ranking in Google Search.

<https://developers.google.com/search/docs/crawling-indexing/google-common-crawlers>

Alan Bleiweiss

[@alanbleiweiss.bsky.social](https://bsky.app/profile/alanbleiweiss.bsky.social)

<https://AlanBleiweiss.com>



SEO & GAI



Stay Out of AI Output

Robots.txt – Prevent LLM access

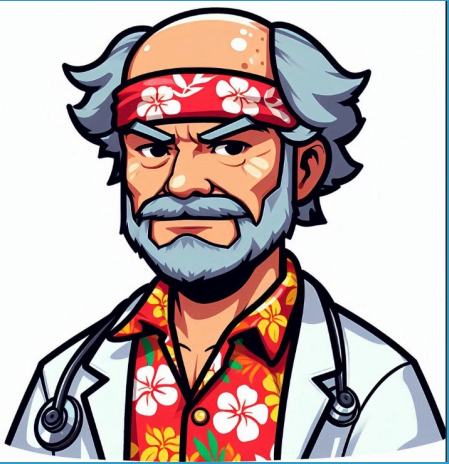
Identifying Applebot

Traffic coming from Applebot is generally identified by using reverse DNS in the [*.applebot.apple.com](https://applebot.apple.com) domain.

Another way is to match the IP address with a CIDR prefix contained in the following JSON file: [Applebot IP CIDRs ↗](#) .

<https://support.apple.com/en-us/119829>





SEO & GAI



Get Found in GAI Output

BONUS: Track GAI Traffic

By Dana DiTomaso

<https://bsky.app/profile/danaditomaso.com>

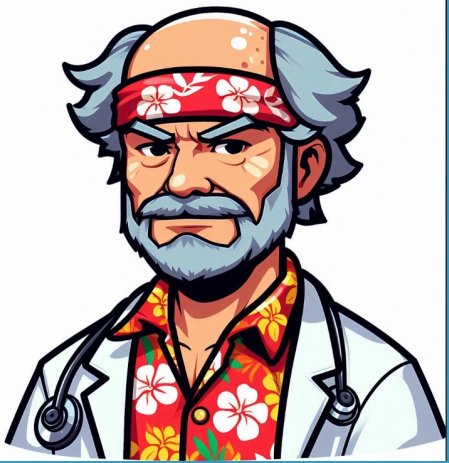
<https://kpplaybook.com/resources/how-to-track-traffic-from-aio-featured-snippets-paa-results-ga4/>

Alan Bleiweiss

[@alanbleiweiss.bsky.social](https://alanbleiweiss.bsky.social)

<https://AlanBleiweiss.com>





SEO & GAI

Technical SEO for AI



Download This Deck in a PDF

<https://alanbleiweiss.com/seo-for-gai/>



Alan Bleiweiss

[@alanbleiweiss.bsky.social](https://alanbleiweiss.bsky.social)

<https://AlanBleiweiss.com>