Rank Higher in Al Search Queries and LLMs

A Complete Guide by Nightwatch

Here's a trend that might be worth paying attention to: Al search is increasingly eating away the traditional search queries from Google.

And here's the absence of a trend... we should be seeing by now: talking about how to rank higher in both AI search and LLM models.

We analyzed the top 1000 websites by traffic and discovered something weird.

In this moment 25% of them are already blocking OpenAI.

Although this may seem as the right approach, it is in fact, counterproductive. Instead, why not focus on ranking *higher* in AI and LLM Search Queries?

This question pushed us to analyze how to appear as top results when LLMs generate answers.

After dedicating tens of hours, analyzing the top-ranking services and pages in LLM outputs, we put it all together in this article.

Stick along (and thank us later), to enjoy the results of our deep dive:

- ✓ How to structure the content to attract LLM indexing and high ranking
- Seed websites to maximize brand attribution
- Key technical optimizations to signal LLM authority
- Best tools and methods for tracking LLM rankings
- ✓ Content partnerships with Al-friendly platforms
- Semantic tags and linking that attract AI indexing

Think of this guide as your shortcut to all the strategies and tactics that make your website visible, when people search for specific services you offer through LLMs.

To set the stage: Understanding the AI Landscape

A Snapshot of how LLMs are Trained

Data Sources for Training:
 Large Language Models (LLMs) are trained on a diverse dataset that includes publicly

available web content, books, research papers, forums, and more. But it's good to know that key sources include:

- Wikipedia: For general knowledge and facts.
- Stack Overflow, GitHub: For technical topics and code-related queries.
- News sites and blogs: For current events and general insights.
- Reddit, Quora, forums: For conversational patterns and user perspectives.
- Public datasets (Common Crawl, OpenWebText): Bulk of web scraping data.

See the list of most influential websites derived from our crawler and general knowledge at the bottom of this article.

Now, to help adapt your strategy for diverse AI bots, let's review the crawling bots—their strenghts, limitations and influences.

Crawling Bots

OpenAI (e.g., GPT-4, GPT-4 Browsing)

OpenAI bots, like GPT-4, are **language models** designed for natural language understanding and generation. Although they are not traditional search bots, they can generate responses based on their training data or live browsing when enabled.

Capabilities:

- Training Data: GPT-4 (and it's derivatives) is trained on a static, historical dataset that includes publicly available text from the web (e.g., books, articles, and forums). It does not actively crawl the web like Google or Bing bots.
- **Browsing Mode**: The browsing-enabled version (e.g., GPT-4 Browsing) can fetch real-time information by querying specific websites or APIs when live data is needed.
- Focus: Answers questions, generates content, and performs reasoning tasks based on user input, prioritizing conversational fluency and logic over traditional ranking systems.

Limitations:

- **Training Updates**: Model training updates are irregular (e.g., months or years apart), so knowledge is often outdated.
- **Browsing Behavior**: Queries authoritative or well-known websites, sometimes restricted by robots.txt or content availability.
- Limited Headlesss Capabilities: The crawling bot by OpenAI has limited headless crawling capabilities, which makes server rendered content more suitable for it's consumption.

Influence Factors:

- Historical content relevance (what was available at the time of training).
- Authority and consistency of the data across trusted sources (to ensure that the outputs are based on reliable and consistent information).

Google Bot (e.g., Google Search Crawler)

Google bots are search engine crawlers designed to index and rank web pages for the Google search engine. Their goal is to organize and provide the most relevant information to users.

Capabilities:

- **Web Crawling**: Actively crawls the web to discover and index content based on a website's structure, metadata, and user engagement metrics.
- Ranking Algorithms: Uses sophisticated algorithms like PageRank, BERT, and MUM to rank results based on relevance, authority, and quality.
- Freshness: Continuously updates its index to reflect the most current web content.

Limitations:

- Relies on website accessibility (robots.txt, sitemaps, and crawl allowances).
- May not always index dynamic or gated content unless explicitly permitted.

Influence Factors:

- SEO optimization like quality backlinks, keywords, and site structure.
- Content quality, mobile-friendliness, and page load speeds.
- User engagement metrics like click-through rates and time spent on the page.

Bing Bot (e.g., Microsoft Bing Crawler)

Similar to Google bots, Bing bots are **search engine crawlers** designed to index and rank content for Microsoft Bing's search engine.

Capabilities:

- **Web Crawling**: Operates like Google bots, discovering and indexing web content to provide search results.
- Al Integration: Bing increasingly integrates Al (like GPT-based models via OpenAl partnership) into its search results and chat-based search features (e.g., Bing Al).
- **Niche Data Sources**: Bing focuses on certain content types like academic or business data and has a unique emphasis on multimedia (images and videos).

Limitations:

- Smaller index compared to Google, leading to potentially less comprehensive coverage of the web.
- Ranking criteria slightly differ, with a stronger focus on specific schema implementations and user intent.

Influence Factors:

- Metadata optimization, such as structured data, plays a more significant role in Bing rankings.
- Authoritative and multimedia-rich content tends to perform better.

Key Differences Between OpenAI, Google, and Bing Bots

Feature	OpenAl (GPT-4)	Google Bot	Bing Bot
Primary Goal	Generate conversational Al responses	Index and rank content for search results	Index and rai
Training/Updates	Static historical snapshots, updated periodically	Dynamic real-time web crawling	Dynamic real crawling
Browsing	Limited to specific queries or browsing-enabled instances	Fully crawls the web automatically	Fully crawls t automatically
Scope of Data	Text-based datasets (no active scraping unless browsing enabled)	All accessible web content	All accessible content
Ranking	No ranking, relies on relevance during generation	Complex ranking algorithms	Ranking basi relevance, so multimedia
Freshness	Historical unless browsing is enabled	Continuously updated	Continuously
User Influence	Indirect via training data and authoritative content	SEO optimization and content updates	SEO optimiza

• Consider this TWO main takeaways:

- LLMs do not scrape live data; they are trained on historical snapshots, which means changes to your online presence may take months or years to impact their responses.
- Browsing-enabled AI tools or agents (e.g., ChatGPT with browsing capabilities)
 use live data when available but rely on highly ranked, authoritative sources.

Now that we've identified the main features, two questions got the interest of our team at Nightwatch:

- 1. Is it possible to impact the historical snapshots—which take months or years to impact the LLM response?
- 2. How to impact the live data?

All our findings indicate:

- LLMs prioritize **content relevance and authority**, which means historical references from highly authoritative sources (e.g., Wikipedia, government websites, scientific publications) remain valuable even if they are older.
- For evergreen content (e.g., foundational knowledge, general principles), historical snapshots provide a reliable and broad foundation.
- **LLMs prioritize occurrence** of data (brand, product names) on high authority websites with higher weight attributed to older references.

From this, we can conclude that historical data—particularly data older than two years, predating the widespread adoption of AI—plays a significant role in shaping the "stable" neural connections within the AI model's knowledge framework.

These neural links, formed during the training process, represent associations and patterns derived from vast amounts of data, which the AI relies on to generate responses. In other words, we can't impact the historical snapshots.

Older data often carries more weight in defining these stable connections because it has been repeatedly reinforced during the model's iterative learning cycles, creating a foundational layer of knowledge. Which answers our second question—live data looses, when competing to the historial snapshots.

What about Al Search with Browsing Agents?

Browsing agents simulate user-like web searches and visit websites, relying on factors like search engine optimization (SEO), site authority, and content quality to prioritize results.

There are 4 Steps in the AI Browsing Process:

- 1. Query understanding: The AI interprets user intent and determines keywords or topics.
- 2. Search initiation: The AI uses search engines like Google or Bing to gather initial results.
- 3. Data selection: It prioritizes sources with authority, relevance, and user-centric content.
- 4. Content extraction: Relevant information is extracted for summarization or direct answer generation.

Impact of Browsing Agents: Unlike traditional SEO that targets human readers, Al-driven search heavily relies on structured, consistent, and authoritative content, as well as metadata and schema.

By now, it might seem as if influencing higher ranking in LLM and AI search is highly nuanced or complex. But there are ways, and we put it in actionable steps below.

How to Rank Higher in LLMs for Specific Queries

Influencing LLM Responses

First we will overview 4 power moves to influence LLM responses, then we will break each down in detail. Several of the outlined actionable steps align with traditional SEO strategies, but when upgraded with LLM and AI friendly tactics, these can push you ahead of top ranking pages (who chose to block Open.AI instead of trying to rank higher in these searches).

To boost your ranking for specific questions about your solution or brand:

1. Focus On Ubiquity Across High-Value Sources:

- Get your brand mentioned in trusted sources like Wikipedia, industry blogs, forums (Reddit, Quora), and widely-read publications.
- Collaborate with technical communities to reference your tools or services (e.g., GitHub, Stack Overflow, Micro forums in your niche).

2. Leverage Backlinks and Mentions:

 Build "good old" backlinks from high-authority domains. These signals remain vital for training datasets like Common Crawl, which weigh web authority heavily.

3. Promote Consistent Branding:

- Use consistent language across your site and profiles to help LLMs connect mentions to your brand.
- Submit structured data (e.g., schema.org) to ensure clear contextual linking between your brand and relevant topics.

4. Create, Create, Create: Evergreen Content:

 Write detailed, solution-specific (best evergreen) content. All models are trained to respond to specific, context-rich queries; build content around user intent and specific pain points.

Now let's unpack each of the 4 tactics.

1. Focus On Ubiquity Across High-Value Sources

In Short:

Al models often learn from widely available and trusted sources. By ensuring your brand or content is ubiquitous across these sources, you improve the likelihood of it being recognized, cited, or utilized in training datasets like Common Crawl, which powers many

Al models. To leverage this Al feature, follow the simple but effective practical advice below.

Actionable Steps

• Wikipedia:

- Contribute to Articles: Create or edit Wikipedia pages with accurate information about your brand, tools, or services. Stick strictly to their content guidelines and cite reliable third-party references.
- Example: If you run an AI tool, contribute a section on a related Wikipedia page (e.g., "Applications of Natural Language Processing") mentioning your tool where applicable.

• Industry Blogs and Publications:

- Pitch guest posts to authoritative blogs in your niche. Offer expert insights or case studies to gain mentions.
- **Example**: If you're in fintech, contribute an article on, for eg.:"The Role of AI in Fraud Detection" to a trusted fintech blog, mentioning your brand/tool.

Forums and Communities:

- This one is often underestimated, but highly effective. Do engage actively in forums like Reddit, Quora, or niche-specific platforms. Provide real, value-driven answers and casually mention your solutions. Over time, it will pay off significantly.
- **Example**: On Reddit's r/SEO, answer a query about content optimization and mention your AI tool as part of your workflow.

• Technical Communities:

- Collaborate with developers on platforms like GitHub or Stack Overflow. Opensource some tools or contribute to ongoing projects with clear branding in documentation.
- **Example**: Publish a GitHub repo offering a free SEO audit tool with branding and links to your premium services.

2. Leverage Backlinks and Mentions

Explanation

Backlinks from high-authority domains act as endorsements, improving your site's credibility and boosting your visibility in Al datasets. When it comes to influencing LMM and Al search, there are a couple of nuanses to pay attention to.

Priorities:

- 1. Obtain backlinks and mentions with top 1000 authority websites that don't block OpenAI bots:
 - Find opportunities for placements of your brand mentions on top 1000 authority websites by the LLM model.
 - Among those—LMM friendly sources—we could mention wikipedia, reddit, quora.
- 2. Focus on newer and less established websites (tools, magazines, blogs)
 - For example, if we were an SEO business, we would focus on the authorities in the specific niche of SEO:
 - searchengineland.com
 - searchenginejournal.com
- 3. All other sites with high DR (domain rating)
 - The third group we focus on is all other websites—typically, these are the ones we
 would engage in SEO exchanges with.

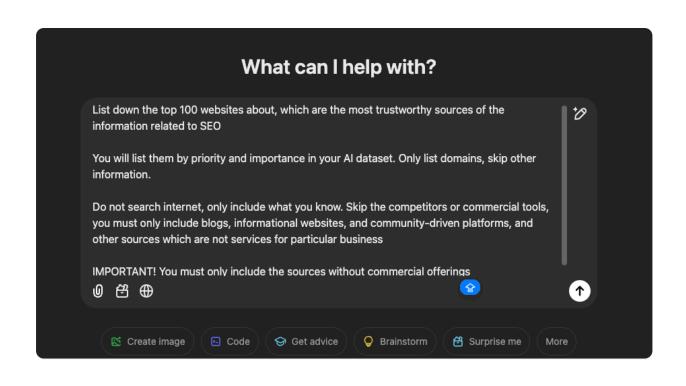
Here is little gift: a prompt for the AI to tell you the relevant websites for your service and niche that you are optimizing for:

List down the top 100 websites about, which are the most trustworthy sources of the information related to SEO

You will list them by priority and importance in your AI dataset. Only list domains, skip other information.

Do not search internet, only include what you know. Skip the competitors or commercial tools, you must only include blogs, informational websites, and community-driven platforms, and other sources which are not services for particular business

IMPORTANT! You must only include the sources without commercial offerings



3. Promote Consistent Branding

Explanation

Consistency in language and branding helps AI and users alike associate mentions across the web with your brand. Implement structured data like **schema.org** as it enhances this recognition.

Actionable Steps

Use Consistent Language and Messaging:

- Maintain a uniform tone and terminology across your website, social profiles, and external content.
- Example: If you describe your AI tool as "an NLP-powered writing assistant,"
 use that phrase consistently instead of alternating with "AI text editor."
- Always mention your brand name (for example: "At Nightwatch we found that instead of blocking AI to crawl your site, you may as well try ranking higher in LLM searches.)

Canonical URLs:

- Make sure that all your content points back to a single source of truth (e.g., your primary website) to avoid content fragmentation.
- Always use a single, preferrably short and consistent domain (for example: nightwatch.io)

Structured Data:

- Structured data provides clear signals for search engines and datasets about your identity.
- Example: Use Organization schema to provide details like your name, logo, and website links, and Product schema to describe your tools.

4. Create Evergreen Content

Explanation

Al models favor content that is timeless and answers specific, actionable queries. Plus, evergreen content works in your favour for the long run, as Al prioritizes historical data. Creating solution-oriented, evergreen content makes your site a valuable resource for users and Al training datasets.

Practical Advice

• Focus on User Intent:

- Understand your audience's pain points and tailor your content to address them.
- Example: If your audience is developers, create content like "How to Integrate an AI Chatbot into Your Web App."

• Deep Dives Over Superficial Coverage:

- Write in-depth, specific content rather than general overviews.
- Example: Instead of "What is SEO?", write "Step-by-Step Guide to Optimizing Ecommerce Sites for Local SEO."

• Regular Updates:

- Refresh content to keep it relevant. This signals to search engines that your site is actively maintained.
- Example: Update your "Top Al Tools for 2023" post to "Top Al Tools for 2024."

Formats that Work:

- Use tutorials, FAQs, case studies, and guides that are highly useful for both users and AI training.
- Example: Publish a series like "Al for Beginners" with hands-on examples.

3 Main Strategies for Long-Term Influence

- Work Towards Becoming a Cited Source: Publish research papers, technical guides, or open data on platforms like arXiv or GitHub, which are frequently used in LLM training.
- Contribute to Authoritative Platforms: Build a presence on Q&A platforms or contribute to thought-leadership articles on trusted industry sites.
- Monitor Model Training Updates: Stay informed about the datasets used by prominent LLMs (e.g., Common Crawl updates) and adapt your strategy accordingly.

How to Rank Higher in Al Search (Browsing-Based Al)

Importance of Websites in AI Browsing

Here's a breakdown of which websites are favored by browsing agents and why:

Website Type	Examples	Importance
High Authority Sources	Wikipedia, Britannica	Primary data validation, often the first stop.
Technical Sources	GitHub, Stack Overflow	Referenced for solutions, technical validation.
News & Media	NYT, TechCrunch, BBC	Used for current events and trending topics.
Q&A Communities	Reddit, Quora	Conversational queries and real-world use cases.
Industry Blogs	Moz, HubSpot, niche blogs	Niche expertise; valuable for long-tail keywords.
Corporate Websites	Official brand pages	Direct product or service information.

6 Steps to Prepare for Al Browsing Agents

1. Optimize for E-A-T (Expertise, Authoritativeness, Trustworthiness):

- Use detailed author profiles to establish expertise link with their Linkedin profiles, expose their additional info if they have a website or any other published content.
- Obtain reviews or mentions on other authoritative sites mentioned above.

2. Implement Structured Data:

 Use schema markup for FAQs, how-tos, and product details. This helps AI extract and interpret your data more accurately.

3. Write Al-Friendly Content (but don't forget on humans too):

- Avoid fluff; prioritize concise, factual, and structured answers,
- Provide subjective point of view that is unique and enriched contains facts, new information or angles not available on other sources,
- Add summaries, tables, and bullet points to improve data extraction.

4. Focus on Domain Authority:

- Continuously improve your backlink profile—never gets old.
- Update your content frequently to remain relevant and visible.

5. Target Featured Snippets:

 Optimize for snippets by directly answering common queries with concise and structured answers. A good tactic is to use FAQ sections under blog posts, landing pages and other content on your website.

6. Stay Active on Q&A Platforms:

 Engage with discussions about your niche on Reddit, Quora, and forums. Use these platforms to position your solution as the go-to answer.

Here's a table ranking **semantic HTML tags** by importance, based on their role in helping Alcrawlers and models extract and understand content effectively:

Rank	Tag	Purpose	Best Practice
1	<title></td><td>Defines the title of the document (appears in browser tabs and search results).</td><td>Include concise, keyword-rich titles (50-60 characters).</td></tr><tr><td>2</td><td><meta></td><td>Provides metadata about the page, like description</td><td>Write clear meta descriptions (under 160 characters) and use</td></tr></tbody></table></title>		

		and keywords.	relevant keywords.
3	<h1></h1>	Main heading, defines the primary topic of the page.	Use one <h1> per page with the most relevant keywords.</h1>
4	<header></header>	Groups introductory and navigational content.	Place your <h1> and navigation links here.</h1>
5	<article></article>	Denotes self-contained content like blog posts or news articles.	Use for distinct, standalone pieces of content.
6	<section></section>	Organizes content into thematic blocks.	Use for grouping related content with proper <h2> or <h3> headings.</h3></h2>
7	<nav></nav>	Represents the navigation menu.	Include links to key sections of the site for better structure.
8	<footer></footer>	Contains information about the page, like copyright or contact details.	Include structured links or metadata for extra relevance.
9	<aside></aside>	Highlights secondary or complementary content (e.g., sidebars).	Use for related or supplementary information.
10	<h2> to</h2>	Subheadings, define the hierarchy of subtopics.	Organize headings logically under

15	<figcaptio< th=""><th>Provides a caption for <figure> content.</figure></th><th>Write descriptive captions that enhance context.</th></figcaptio<>	Provides a caption for <figure> content.</figure>	Write descriptive captions that enhance context.
16		Structures tabular data.	Include <aption> to describe the table, and use properly labeled rows and columns.</aption>
17	<caption></caption>	Describes a table's purpose or content.	Use for clarity when presenting tabular data.
18	<main></main>	Identifies the main content of the document.	Helps crawlers distinguish primary content from secondary elements.
19	<time></time>	Marks dates or times.	Use for events, blog posts, or any time-specific data with proper formatting.
20	<details></details>	Provides collapsible content.	Use with <summary> to make additional information accessible without cluttering the layout.</summary>

Key Benefits of LLM and AI Search Optimization

Aspect	Benefit
LLM Optimization	Long-term influence on AI models and consistent query ranking.
Browsing Al Search	Immediate visibility for real-time AI-driven queries.
Brand Trust	Building credibility across both traditional and Al-driven mediums.
Content Versatility	Content optimized for AI appeals to both users and machines.

Easy Tips to Sustaining a Long-Term LLM Ranking Strategy

5.1 Immediate Actions

- Identify and target high-authority platforms for backlinks and mentions.
- Audit your website for structured data, content clarity, and relevance.
- Begin contributing to niche forums and authoritative blogs.

5.2 Long-Term Initiatives

- 1. **Data Partnerships**: Provide data or insights that can be incorporated into public datasets.
- 2. **Monitor AI Tools**: Stay updated on LLM capabilities and browsing agent algorithms to refine your strategy.
- 3. **Iterative Improvements**: Regularly update your content and SEO efforts to maintain relevance in AI searches.

You can establish a dominant presence across Al-driven platforms, and now it's the best time to work towards that goal. Make sure to keep your focus on long-term initiatives and short-term benefits simultanously.

Why Tracking AI Rankings is Challenging (and how to tackle it)

1. Lack of Standardization:

Al models like GPT-4 or Google's Al Overview lack a unified or transparent ranking system. These models do not display "results" in the traditional sense but rather generate responses based on internal reasoning and their understanding of relevance and context.

2. Dynamic Nature of Al Outputs:

Responses from LLMs and browsing-enabled tools can vary significantly based on phrasing, context, or even the specific AI agent. This variability makes tracking performance across these platforms more complex than traditional search engines.

3. Limited Tools in the Market:

Most SEO tools today are not designed to monitor AI responses. Traditional or specialized rank trackers (like SEMrush, Ahrefs, or even STAT) do not natively

support the tracking of AI-driven rankings or experimentation with long-term AI-specif strategies.	

6.2 Nightwatch: The First Industry Tool for Tracking Al Rankings

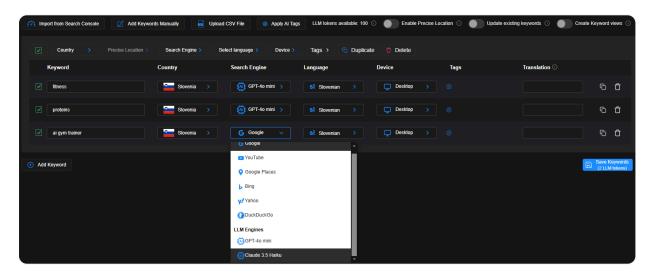
Nightwatch has pioneered a solution by introducing tools specifically designed to track and experiment with AI-driven rankings. This capability provides unmatched insights into how your brand, website, or solutions perform in various AI-generated responses.

Gain Support With Tracking 4 AI Platforms

Nightwatch currently supports ranking and experimentation tracking for the following systems:

- **GPT-4 (Optimized for Browsing)**: Monitors how your content appears in responses generated by GPT-40 with browsing capabilities.
- Haiku: Tracks rankings and mentions in AI responses driven by this rising LLM competitor.
- **Bing Al Chat**: Tracks brand presence and specific mentions in Bing's conversational search interface.
- **Google Al Overview**: Measures performance in Google's Al-generated search summaries, which are becoming increasingly common in search result previews.

Here is how it looks:



Users can now besides the traditional search engines track rankings across various AI models and AI searches in Nightwatch (sign-up here).

PS: Currently we are in early stages of testing the LLM tracking capabilities (supporting GPT-4o and Anthropic Haiku — the most widely used models. If you would like to get an early access, please fill out the form here.

6.4 Alternative Option: Home-Made Scripts

Tracking AI rankings often involves home-made scripts. While these can provide basic insights, they require technical expertise and resources to implement:

- Manual Prompting: Using GitHub scripts or custom headless scrapers to send queries to AI platforms. These methods can be inefficient and time-consuming, but they can still get the job done on a smaller scale.
- Response Parsing: Analyzing and storing AI responses to specific prompts over time.
- Data Visualization: Building dashboards to track trends and evaluate changes.

Although effective, these scripts are labor-intensive, lack the scalability and analytical depth of dedicated tools like Nightwatch.

Why You Should Track AI Rankings

1. Early-Mover Advantage:

As AI-driven search interfaces become more prevalent, brands with visibility in these systems will gain a competitive edge. As we found 25% of top tanking pages are blocking **Open.AI**, opening room for other pages to take their space. This means that businesses starting to optimize for AI rankings and track results now, will be better positioned to capture audience attention and drive traffic to their websites before the market becomes saturated with competitors.

2. Fine-Tuned Strategies:

By understanding how AI rankings respond to content updates, you can craft a strategy tailored to these systems, complementing your traditional SEO efforts. This involves analyzing the specific algorithms and ranking factors that AI systems prioritize, allowing you to make informed decisions about content creation, keyword targeting, and overall digital marketing strategies that align with the evolving landscape.

3. Comprehensive Insights:

Combining traditional SEO tracking with AI ranking analysis provides a holistic view of your digital performance across all platforms. This comprehensive approach enables you to identify trends, measure the effectiveness of your strategies, and make data-driven adjustments to improve your online presence. By leveraging insights from both traditional and AI-driven metrics, you can ensure that your brand remains competitive and relevant in an increasingly digital world.

Focus on dual strategy: long-term and immediate efforts

The historical nature of LLM training creates a solid foundation for stable, comprehensive knowledge, but it requires consistent, long-term effort to influence.

In contrast, browsing-enabled AI tools offer opportunities to rank for real-time queries but rely heavily on SEO and domain authority. By understanding how both systems work and focusing on effective actionable steps listed above, you can craft a dual strategy that leverages the stability of historical data and the immediacy of browsing AI for maximum impact.

Contact:

☑ aljaz@nightwatch.io

Sign-up for LLM tracking: nightwatch.io

Connect on Linkedin