



AraLith

www.aralith.com · SEO Orchestration & Content Operating System

| OPERATIONAL GUIDE · 2027 EDITION

AI Overview Optimization

How to structure your content to get cited in Google's AI Overviews: answer-first writing, semantic coverage, structured data and Generative Engine Optimization.

35 points

Complete operational checklist, to use on every article

Entity · RAG

Entity SEO, passage-level retrieval and multi-engine included

Schema.org

Markup examples ready to adapt and validate

Free guide by aralith.com

Full article version at
[aralith.com/en/guides/ai-optimization-overview-how-to-do-it-right-a-guide-to-2027](https://www.aralith.com/en/guides/ai-optimization-overview-how-to-do-it-right-a-guide-to-2027)

CONTENTS

What you will find in this guide

- 01 What the AI Overview is and why it changes everything
- 02 How it works: Gemini's fan-out queries
- 03 Classic SEO vs GEO: what really changes
- 04 The 5 key strategies to get cited
- 05 The step-by-step process (5 operational steps)
- 06 Schema markup: which one to use and when
- 07 E-E-A-T: measurable authority
- 08 Case study: medichub.it cited in 30 days
- 09 The AI Overview numbers in 2027
- 10 Entity SEO and the Knowledge Graph
- 11 Off-site authority: brand mentions and digital PR
- 12 How the AI reads your page: RAG, passages and grounding
- 13 Technical foundations: AI crawlers, rendering, speed
- 14 Original data as a citation asset
- 15 Measurement and beyond Google: ChatGPT, Perplexity, Bing
- 16 Complete operational checklist (35 points)
- 17 Frequently asked questions

How to use this guide

Every chapter opens with the direct answer to the question in its title: the same answer-first principle you will need to apply to your own content. You can read it in sequence, or jump straight to the checklist in chapter 16 and use it as the operational reference for every article you publish.

CHAPTER 1

What the AI Overview is and why it changes everything

The AI Overview is the AI-generated summary box that Google shows at the top of the SERP for informational queries, synthesizing answers from multiple web sources and citing them as references. It is the evolution of the Search Generative Experience and the deepest change in search of the last twenty years.

Unlike classic organic results, the AI Overview answers directly on the results page. For content creators the paradigm changes: ranking on the first page is no longer enough — your content must be clear, structured and semantically rich enough for the AI to extract the answer from it. This approach is called **Generative Engine Optimization (GEO)**, alongside Answer Engine Optimization (AEO).

Google does not select the most popular source: it selects the one that answers the user's question with the most clarity and completeness. The extractive formatting of your content determines who gets cited.

CHAPTER 2

How it works: Gemini's fan-out queries

The Gemini model powering AI Overviews expands the user's original query into a series of related sub-questions (fan-out), then synthesizes the answers choosing the best sources for each sub-question.

Practical example: for the query "what is the AI Overview", Gemini internally generates variants such as "how does the AI Overview work", "how to get cited in the AI Overview", "how long does it take to appear". Content that covers not only the main question but also its natural variants (who, what, how, why, when) is perceived as a high-density source and has a much higher probability of being cited.

Operational implication: every article must map the whole graph of questions around the topic, not a single keyword. Use Google Suggest, People Also Ask and your own search query data to reconstruct the real fan-out queries.

Classic SEO vs GEO: what really changes

Classic SEO optimizes for keywords, backlinks and CTR; GEO optimizes for answer clarity, semantic structure and source credibility. The two approaches do not exclude each other: they reinforce each other.

DIMENSION	CLASSIC SEO	GEO / AI OVERVIEW
Goal	Organic ranking in the SERP	Citation in the generative box
Optimization unit	Keyword and page	Question and extractable passage
Main signals	Backlinks, CTR, Core Web Vitals	Answer-first, topic coverage, Schema, E-E-A-T
Winning format	Long, comprehensive content	Self-contained blocks: 40-60 word answer per section
Measurement	Rankings and organic traffic	Presence and position of AI citations

What still holds from SEO

Organic ranking remains the prerequisite: about 90% of the sources cited in AI Overviews already come from the top 10. Technical health, speed and authority still count.

What GEO adds

At equal rankings, the source with the most extractive structure gets cited: immediate answer, self-contained blocks, correct markup. It is the layer that decides who enters the box.

The 5 key strategies to get cited

Being cited in the AI Overview depends on how your content is built, not just on how many keywords it contains. These five strategies separate cited content from content the AI never sees.

1 Direct, immediate answers (answer-first writing)

Open every paragraph with the answer to the main question, no preambles. Google's AI extracts the first 2-3 sentences of each section: they must be self-contained and unambiguous, readable even out of context. The ideal lead answer is 40-60 words.

2 Deep semantic coverage

Anticipate the related sub-queries and the questions of the whole search funnel. An article that covers the entire topic, not just the main keyword, is more likely to be chosen as a source for multiple fan-out queries at the same time.

3 Extractive formatting

Use bullet lists, tables and clear H2/H3 headings. The AI prefers structures that let it isolate information without rewriting it: every heading should be a real question or a complete statement, every list item self-contained.

4 Structured FAQ blocks

Integrate Q&A sections with at least 5-8 real user questions (source: Google Suggest and People Also Ask). Each answer must be short, dense and self-contained (2-4 sentences). FAQPage markup amplifies extraction probability and PAA visibility.

5 E-E-A-T and structured data

Implement Schema.org markup (Article, HowTo, FAQPage) and demonstrate Experience, Expertise, Authoritativeness and Trust with verifiable data, real authors with Person markup and cited sources. One concrete, dated fact is worth more than ten generic claims.

The step-by-step process: 5 operational steps

Follow this sequence for every article. The order matters: never open a paragraph with context or the problem — always start with the answer, then argue it. Each step makes the content extractive: the AI must be able to take your answer as it is and use it in the generated box.

1 Analyze queries with informational intent

Use Google Suggest, People Also Ask and keyword research tools to find the questions your audience really asks. Focus on informational queries (who, what, how, why, when): that is where the AI Overview appears most often. Build a map of main question + fan-out queries.

2 Structure the article with question-answer headings

Turn every main section into an explicit question (H2/H3). The first sentence under the heading must be the direct, complete answer, readable out of context. Practical rule: every H2 = one sub-question, with a 40-60 word lead answer.

3 Add lists, tables and comparison blocks

Granular formatting (lists, comparison tables, explicit definitions) is easier for the model to extract. Every list item must be autonomous and understandable on its own, without depending on the previous paragraph.

4 Integrate an FAQ section with Schema markup

Add an FAQ section with at least 5-8 real questions. Each answer complete in 2-4 sentences. Implement Schema.org FAQPage markup to explicitly signal the structure to Google's AI (JSON-LD example in chapter 6).

5 Verify E-E-A-T and validate structured data

Sign the article with a real author (Person markup), cite verifiable sources, add up-to-date data with explicit dates. Before publishing, validate the markup with Google Rich Results Test. Plan a quarterly review updating dates and cited data.

Schema markup: which one to use and when

The three most effective Schema types for the AI Overview are **FAQPage**, **HowTo** and **Article/BlogPosting**. They explicitly signal the content structure to the AI and make answer extraction easier.

SCHEMA	WHEN TO USE IT	AI OVERVIEW ADVANTAGE
FAQPage	Articles with an explicit question-and-answer section	The AI extracts question-answer pairs directly as self-contained snippets
HowTo	Procedural guides with numbered sequential steps	Google shows individual steps in the box; increases citations for "how to" queries
Article / BlogPosting	General informational content, news, editorial posts	Signals author, publication and update dates: reinforces E-E-A-T and freshness

Combining HowTo + FAQPage on the same page

Implementing HowTo (for the operational steps) and FAQPage (for the FAQ section) simultaneously is still a rare combination: using it is an immediate competitive advantage. Minimal FAQPage markup example in JSON-LD:

```
<script type="application/ld+json">
{
  "@context": "https://schema.org",
  "@type": "FAQPage",
  "mainEntity": [{
    "@type": "Question",
    "name": "What is Google's AI Overview?",
    "acceptedAnswer": {
      "@type": "Answer",
      "text": "The AI Overview is the AI-generated block that Google
        shows at the top of the SERP for informational queries,
        synthesizing answers from multiple sources and citing them."
    }
  }]
}
</script>
```

Always validate the markup with Google Rich Results Test before publishing: malformed markup is ignored and wastes the structured signal.

CHAPTER 7

E-E-A-T: measurable authority

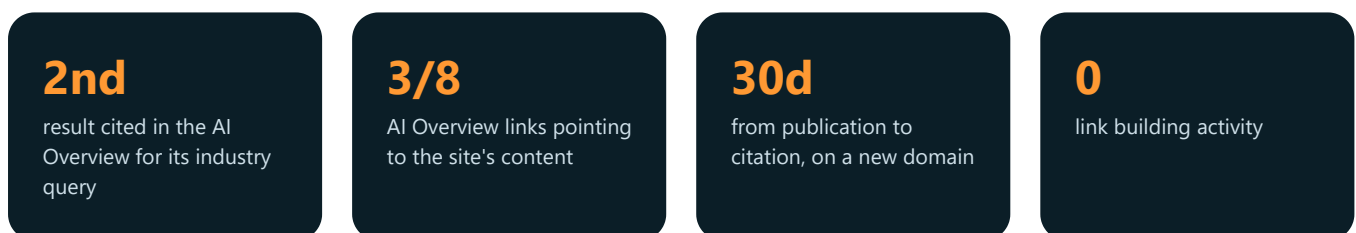
The AI Overview favors sources with verifiable data, dated case studies and authors with documentable expertise. E-E-A-T (Experience, Expertise, Authoritativeness, Trust) is not a theoretical signal: it is built with concrete elements.

- ✓ **Real, verifiable author:** byline with Person markup, bio with credentials, links to professional profiles (e.g. LinkedIn).
- ✓ **Explicitly dated data:** "July 2026" signals freshness; a fact without a date is a weak fact.
- ✓ **Primary sources cited:** official documentation (e.g. Google Search Central), studies, your own metrics.
- ✓ **Case studies with real metrics:** measurable results on real projects are the highest level of proof and the hardest for competitors to replicate.
- ✓ **Planned updates:** quarterly review at minimum; monthly cycle for fast-moving topics like AI.

CHAPTER 8

Case study: medichub.it cited in 30 days

A new domain can be cited in the AI Overview without link building: medichub.it achieved it in 30 days thanks exclusively to the semantic structure of its content (verified data, July 2026).



The method applied, replicable on any project, combines the principles of this guide: an above-the-fold "In short" section with answer-first bullets, H2/H3 headings written as real questions (to capture People Also Ask), fan-out query coverage, simultaneous HowTo + FAQPage markup and explicitly dated data. It is the SEO+AEO+GEO workflow integrated in AraLith COS.

CHAPTER 9

The AI Overview numbers in 2027

65%

of informational queries show an AI Overview

42%

average CTR for the #1 cited source

90%

of cited sources already rank in the top 10

3x

citations for content with FAQ schema

Operational reading: the correlation with the top 10 is strong but not absolute. An article in position 5 with FAQ structure can be cited before one in position 1 without extractive formatting. Organic ranking remains necessary; extractive structure decides who gets cited.

CHAPTER 10

Entity SEO and the Knowledge Graph

The AI Overview does not reason in keywords but in entities: Gemini maps mentions of people, organizations, products and places to unique identifiers in Google's Knowledge Graph. If the model cannot unambiguously connect your content to an entity or a specific fact, it does not cite it, no matter how well written it is.

Entity-based optimization works on two levels: **disambiguation** (who you are) and **topical authority** (what you are an expert in).

- ✓ **Complete Organization markup:** name, logo, description, address, contacts and above all sameAs properties pointing to consistent official profiles (LinkedIn, registries, industry directories). It is your "identity document" for the Knowledge Graph.
- ✓ **Person markup for authors:** the same logic applied to people: bio, credentials, jobTitle, sameAs to verifiable professional profiles.
- ✓ **Cross-platform consistency:** the brand's name, description and claims must be identical wherever they appear. Inconsistencies weaken entity consolidation.
- ✓ **Topical map and internal clusters:** build a cluster of pages covering all the sub-topics of the main topic, linked with descriptive anchors. Topical authority is proven at domain level, not single-page level.

Do not optimize a page for a keyword: optimize an entity (the brand, the author, the product) for a semantic territory. Every new cluster page reinforces the citation probability of all the others.

Off-site authority: brand mentions and digital PR

The AI evaluates an entity's authority across the whole web, not just your site: brand mentions on authoritative third-party sources count as trust signals even when they contain no link.

This logic differs from classic link building: what matters is the repeated, consistent association of the brand with a topic across multiple trustworthy, independent sources. A new domain can be cited without backlinks (chapter 8), but third-party validation accelerates and consolidates presence in AI answers.

The practical levers, in order of accessibility for an SMB

- ✓ **Targeted digital PR:** contributions and interviews on industry publications and blogs, even without links. The brand+topic co-occurrence is the signal.
- ✓ **Citable data:** publish original numbers that journalists and bloggers have a reason to pick up (see chapter 14).
- ✓ **Presence in vertical communities and directories:** authentic discussions on industry forums and communities, profiles in recognized directories.
- ✓ **Podcasts, webinars and video:** transcripts get indexed and associate the brand's people with the topic.

Warning

Authenticity above all: artificial mentions, duplicated press releases or coordinated community activity get recognized and can damage the entity instead of strengthening it.

How the AI reads your page: RAG, passages and grounding

Generative engines do not read pages the way people do: they split them into individual passages (chunks) and score each one for relevance, clarity and factual density. Every section must therefore stand on its own.

The mechanism is called **Retrieval-Augmented Generation (RAG)**: the model retrieves the most relevant passages from the index, then generates the answer anchoring it (**grounding**) to the retrieved documents. Every sentence of the AI Overview must be supported by a real passage: if your content is ambiguous or lacks sharp factual statements, the model cannot anchor the answer to your text and picks a clearer source.

Practical consequences of passage-level retrieval

- ✓ **Block autonomy:** every H2/H3 + first paragraph is an extractable unit; avoid references like "as seen above" that break the passage's self-sufficiency.
- ✓ **Factual density:** prefer verifiable, specific statements over vague wording; a passage with a number, a definition or a name is more "anchorable" than a discursive one.
- ✓ **TL;DR under key headings:** a 1-2 sentence summary right under the section title works as a perfect chunk for extraction.
- ✓ **One concept per passage:** if a paragraph mixes two answers, neither will be extracted well.

Write every section as if it had to be read alone, outside the article's context: that is exactly how the AI will read it.

Technical foundations: AI crawlers, rendering, speed

Perfect content that is technically inaccessible will never be cited. Three checks: crawler access, server-side rendered content, performance.

- ✓ **Crawler access:** check robots.txt. AI Overviews use the regular Googlebot; the Google-Extended token instead controls the use of your content for model training. Blocking it does not remove you from AI Overviews, but the configuration must be a conscious decision. Also evaluate the crawlers of other generative engines (e.g. OAI-SearchBot, PerplexityBot) based on your strategy.
- ✓ **Content visible without JavaScript:** critical text must not live only inside tabs, accordions or client-side rendered components. If the passage is not in the served HTML, it does not exist for retrieval. Check with "view source", not the inspector.
- ✓ **Performance and Core Web Vitals:** loading speed and rendering stability remain baseline factors for organic ranking, and since about 90% of cited sources come from the top 10, they remain indirectly decisive for AI citations too. Distrust the hyper-precise statistics circulating on this topic: many have no verifiable primary source.
- ✓ **Indexing hygiene:** correct canonicals, updated sitemap, no duplicate content diluting the authority of your best passage.

robots.txt

Googlebot free; Google-Extended and third-party AI crawlers configured consciously, not by default.

Served HTML

The text that matters is in the page source, not generated by the browser after load.

Core Web Vitals

LCP, CLS and INP in the "Good" ranges: the ranking prerequisite that feeds citations.

Original data as a citation asset

The highest level of E-E-A-T is becoming a primary source: publishing data nobody else has, so that others, including the AI, have to cite you.

A case study with real metrics (like medichub.it, chapter 8) is already original data. The principle generalizes: any number produced by your business can become a citation asset if published correctly.

- ✓ **Industry benchmarks:** anonymous aggregates from your clients or your product (e.g. average times, adoption rates, most frequent errors).
- ✓ **Vertical mini-surveys:** even 50-100 respondents from a specific niche produce a unique, citable data point.
- ✓ **Documented experiments:** A/B tests, before/after with explicit methodology, dates and absolute numbers.

Publication rules

Declared methodology, explicit sample and period, date in the title or first paragraph, one numeric takeaway that fits in a single sentence (that is what the AI will extract), and a stable dedicated page that accumulates citations instead of scattering them.

Measurement and beyond Google: ChatGPT, Perplexity, Bing

Every generative engine retrieves sources differently: optimizing only for Google covers just part of the AI traffic. The shared foundation (extractable passages, Schema, entity authority) holds everywhere; per-engine tactics layer on top.

ENGINE	WHERE IT RETRIEVES FROM	PRIORITY LEVER
Google AI Overview	Top of the classic SERP (about 90% from the top 10)	Organic ranking + extractive formatting + FAQ/HowTo schema
ChatGPT Search	Own index + Bing index	Bing rankings, short sharp answers, citations on authoritative pages
Perplexity	Live web search on every query	Freshness and update cadence, dense citations, presence on news and communities
Bing Copilot	Bing index	Technical SEO on Bing (often neglected), Schema markup, consistent entities

How to measure AI presence

- ✓ **Systematic manual checks:** target queries in incognito browsing, with dated screenshots to build a history.
- ✓ **Automatic monitoring:** tools that track citations and position in AI answers for a set of queries (a built-in feature in modern CMSs like AraLith COS), with alerts on new or lost citations.
- ✓ **AI referrals in analytics:** segment traffic from chatgpt.com, perplexity.ai, copilot.microsoft.com and from AI Overview citation clicks.
- ✓ **AI share of voice:** for strategic queries, track how often your brand appears versus competitors in generated answers: the metric replacing "average ranking" in the generative era.
- ✓ **Zero-click management:** accept that part of the queries end inside the AI answer; include elements that earn the click (tools, templates, full data, downloadable checklists) and measure conversions, not just sessions.

Complete operational checklist (35 points)

Use this checklist on every article before publishing. Content that satisfies every point is structurally ready to be extracted by the AI Overview.

A. Research and planning

- ☐ Main question identified with informational intent (who/what/how/why/when)
- ☐ At least 5-8 fan-out queries mapped with Google Suggest and People Also Ask
- ☐ Current AI Overview presence verified for the target queries (incognito search)
- ☐ Already-cited content analyzed: structure, format, gaps to fill
- ☐ Differentiating angle defined (own data, case study, direct experience)

B. Content structure

- ☐ "In short" section above the fold with answer-first bullets
- ☐ Every H2/H3 written as a real question or a complete statement
- ☐ First sentence of every section = direct, self-contained answer (40-60 word lead)
- ☐ No paragraph opening with context or preamble before the answer
- ☐ Full topic coverage: every fan-out query has a dedicated section

C. Extractive formatting

- ☐ Bullet lists with autonomous, individually understandable items
- ☐ At least one comparison table or explicit definition where relevant
- ☐ Procedural guides structured in numbered sequential steps
- ☐ Short paragraphs: one concept per paragraph
- ☐ FAQ section with 5-8 real questions, complete 2-4 sentence answers

D. Structured data

- ☐ FAQPage markup implemented on the FAQ section
- ☐ HowTo markup implemented on procedural steps (if present)
- ☐ Article/BlogPosting markup with author, datePublished and dateModified
- ☐ Person markup on the author with verifiable credentials
- ☐ Validation passed with Google Rich Results Test

Checklist: E-E-A-T, entities, technical

E. E-E-A-T and maintenance

- ☐ Real author byline, with bio and link to a professional profile
- ☐ Concrete facts cited with an explicit date (e.g. "July 2026")
- ☐ Primary sources linked (official documentation, studies, own metrics)
- ☐ Update date visible on the page
- ☐ Planned review: quarterly (minimum) or monthly for volatile topics

F. Entities and off-site authority

- ☐ Organization markup with sameAs to consistent official profiles
- ☐ Brand name, description and claims identical on every platform
- ☐ Page part of a topical cluster with descriptive internal links
- ☐ At least one mention or contribution on an authoritative third-party source
- ☐ Original data published (benchmark, case study, survey) with explicit methodology

G. Technical and measurement

- ☐ robots.txt checked: Googlebot access guaranteed, Google-Extended decided consciously
- ☐ Critical text present in the served HTML (checked with "view source")
- ☐ No key content inside client-side rendered tabs/accordions
- ☐ Target queries monitored: AI citation checks with a dated history
- ☐ AI referrals segmented in analytics (chatgpt.com, perplexity.ai, copilot)

Print these two pages and keep them next to every publication: an article that passes all 35 points is structurally ready for AI extraction.

Frequently asked questions

What is Google's AI Overview and how does it work?

It is an AI-generated block that Google shows at the top of the SERP for informational queries. It synthesizes answers from multiple web sources, citing them as references, and selects the content with the clearest structure, the widest topical coverage and the highest level of authority (E-E-A-T).

Do you need to rank on the first page to get cited?

Not strictly, but the correlation is strong: about 90% of cited sources already come from the top 10. You need a good organic position, but also extractive formatting: an article in position 5 with FAQ structure can be cited before one in position 1 without structure.

How do I check whether my site is already cited in an AI Overview?

Manually, by searching your target queries on Google in incognito mode (to avoid personalization). Alternatively, AI Overview monitoring tools built into modern CMSs track citations and positions automatically, with real-time alerts.

Does structured data really help with citations?

Yes, significantly. FAQPage, HowTo and Article explicitly signal the content structure to the AI and make extraction easier. They are not a guarantee, but content with correct markup has up to 3 times higher citation probability.

How often should I update my content?

Google favors freshness, especially for evolving topics like AI. A quarterly review updating dates and cited data is the recommended minimum; for highly volatile topics a monthly check cycle is preferable.

Is AI Overview optimization different from traditional SEO?

Yes, complementary but different. Traditional SEO optimizes for organic ranking (keywords, backlinks, Core Web Vitals); GEO optimizes for semantic extraction: immediate answers, topic coverage, structured formatting and demonstrated authority. The two strategies reinforce each other.

Frequently asked questions

What are fan-out queries and why do they matter?

They are the secondary questions users ask after the main one. Gemini automatically expands the original intent into these sub-questions to build the answer: content that covers them all has a much higher probability of being cited.

What is the ideal answer length for extraction?

The lead answer in the first paragraph should be 40 to 60 words: complete enough to be extracted on its own, short enough not to be cut by the AI. The following paragraphs go deeper.

What is Entity SEO and why does it matter for the AI Overview?

It is the optimization of entities (brands, authors, products) instead of keywords alone. Gemini maps mentions to identifiers in Google's Knowledge Graph: if your content cannot be unambiguously connected to a recognized entity, citation probability collapses. It is built with Organization/Person markup, consistent sameAs properties and internal topical clusters.

Are brand mentions without links worth anything?

Yes. The AI evaluates entity authority across the whole web: the repeated association of your brand with a topic on trustworthy third-party sources is a trust signal even without backlinks. Digital PR, citable data, podcasts and presence in vertical communities strengthen the entity beyond your site's perimeter.

Is optimizing for Google enough for ChatGPT and Perplexity too?

Partially. The foundation is shared (extractable passages, Schema, entity authority), but each engine retrieves differently: ChatGPT also relies on the Bing index, Perplexity runs live searches rewarding freshness and dense citations, AI Overviews draw from the top of Google's SERP. You need a common base plus per-engine tactics.

Can a new site without authority get cited?

Yes, as the medichub.it case shows (July 2026): a new domain cited within 30 days without link building. The AI Overview does not rely only on domain authority but on the extractive quality of the individual content: direct answers, complete coverage, correct markup and verifiable E-E-A-T signals.

WHAT NOW?

From the guide to the practice

You have the complete method: answer-first, fan-out query coverage, Schema markup, consistent entities, original data and multi-engine measurement. The next step is applying it consistently to every piece of content you publish, and monitoring what gets cited.

1. Apply

Use the 35-point checklist on every article, new or existing, starting from the queries where the AI Overview already appears.

2. Monitor

Build your citation history: who gets cited, with what structure, and when your page enters the box.

3. Iterate

Update content with fresh, dated data, widen the cluster's coverage and consolidate the brand entity.

Want to be cited in Google's AI Overviews?

We analyze your content, identify high-potential queries and structure your articles to maximize citation probability, with planned quarterly updates, correct Schema markup and fan-out query coverage.

Request a free analysis · aralith.com

Discover AraLith AI Overview · www.aralith.com

© 2026 AraLith · Operational guide to AI Overview optimization, 2027 edition. Full article version available at www.aralith.com/en/guides/ai-optimization-overview-how-to-do-it-right-a-guide-to-2027