AI Visibility Report: Wikipedia, the free encyclopedia

Prepared for: Wikipedia, the free encyclopedia | **Website:** https://en.wikipedia.org/wiki/Main_Page | **Date of analysis:** 2026-09-27

EXECUTIVE SUMMARY

We analysed 8 pages on the en.wikipedia.org domain on 27 September 2026, assessing how ready your site is to be found, understood and cited by AI-driven search engines such as Google AI Overviews, ChatGPT and Perplexity. Your site earned a **GEO Readiness Score of 77 out of 100**, placing it in the **Good** tier, held back most heavily by content and metadata signals rather than technical access. The single most impactful Confirmed finding is that **all 8 analysed pages are missing a meta description**, which removes the short, controllable summary that AI engines and search snippets lean on to represent your pages. Our top three priority actions are: add unique meta descriptions and publication dates to every page, extend structured data to the 7 content pages that currently carry none, and strengthen your entity signals by adding og:site_name and more sameAs links. Fixing these gaps will make your pages easier for AI systems to summarise accurately and cite confidently, improving how your brand is represented in AI answers.

GEO READINESS SCORE: 77/100 - GOOD

Your overall readiness is solidly in the Good tier. Strong technical foundations and completely open AI crawler access are pulling the score up, while content quality and E-E-A-T signals are pulling it down.

Component	Score	Weight	Weighted Score
AI Platform Readiness	100	25%	25.0
Content Quality & E-E-A-T	45	25%	11.3
Technical Foundation	93	20%	18.6
Schema & Structured Data	70	15%	10.5
Brand Authority & Entity Presence	75	15%	11.3
Overall	**77**	**100%**	**77**

- **AI Platform Readiness (100):** Nothing was found to deduct. No AI crawlers are blocked in robots.txt, so every major AI platform can reach your content.
- **Content Quality & E-E-A-T (45):** This component lost 55 points across six issues: 6 for 3 of 8 content pages lacking an H1 or at least two H2 sections (/wiki/Wikipedia:About, /wiki/Wikipedia:Contact_us, /wiki/Help:Introduction), 10 because no page uses question-style headings, 9 for 5 of 8 substantial pages showing no author byline, 10 because 8 of 8 substantial pages show no publication date, 10 because 8 of 8 pages are missing a title or meta description, and 10 because 59 of 151 images have no alt text.
- **Technical Foundation (93):** This component lost 7 points: 5 for no XML sitemap found at /sitemap.xml or in robots.txt, and 2 because 1 linked page could not be fetched.
- **Schema & Structured Data (70):** This component lost 30 points: 5 because the homepage has no WebSite schema, 20 because 7 content pages carry no structured data, and 5 because no inner page has BreadcrumbList schema.
- **Brand Authority & Entity Presence (75):** This component is capped at 80 because off-site presence is not checked in this audit. It lost 25 points from that cap-adjusted base: 15 for having only 1 sameAs link to an external profile, and 10 because the homepage declares no og:site_name.

KEY FINDINGS

The issues below are ordered by impact. Each is numbered so the Prioritised Action Plan can refer back to it. These are problems only; genuine strengths are listed separately afterwards.

1. Every analysed page is missing a meta description

- **What we saw:** All 8 pages return `metaDescription: null`, including /wiki/Main_Page, /wiki/Wikipedia:About and /wiki/Wikipedia:Contact_us.
- **Confidence:** Confirmed.
- **Business impact:** High. The meta description is the short, controllable summary that AI engines and search snippets often quote or paraphrase, so without it you leave your representation to chance. This is part of the 10-point "missing a title or meta description" deduction in Content Quality & E-E-A-T.
- **Action:** Add a unique, factual 150-160 character meta description to every page, starting with /wiki/Main_Page, /wiki/Wikipedia:About and /wiki/Wikipedia:Contact_us.

2. No page shows a publication or updated date

- **What we saw:** All 8 substantial pages return `hasPublicationDates: false`, including /wiki/Main_Page, /wiki/Wikipedia:About and /wiki/Wikipedia:Contact_us.
- **Confidence:** Confirmed.
- **Business impact:** High. AI systems favour content they can date-stamp for freshness, and pages with no visible date are less likely to be trusted and cited for time-sensitive answers. This is the 10-point publication-date deduction in Content Quality & E-E-A-T.
- **Action:** Expose a visible "last updated" date and a machine-readable `dateModified` on every content page, prioritising fast-moving pages like /wiki/Portal:Current_events.

3. 59 of 151 images have no alt text

- **What we saw:** Across the crawl, 59 of 151 images are missing alt text, concentrated on /wiki/Wikipedia:Community_portal (25 of 52 missing), /wiki/Help:Contents (11 of 17 missing) and /wiki/Portal:Current_events (9 of 27 missing).
- **Confidence:** Confirmed.
- **Business impact:** Medium. Missing alt text removes the text description AI systems use to understand and cite images, and weakens accessibility. This is the 10-point image alt-text deduction in Content Quality & E-E-A-T.
- **Action:** Add descriptive alt text to the missing images, starting with the heaviest offenders: /wiki/Wikipedia:Community_portal, /wiki/Help:Contents and /wiki/Portal:Current_events.

4. No page uses question-style headings

- **What we saw:** Every analysed page returns `questionStyleHeadings: 0`, despite several pages carrying many section headings (for example /wiki/Wikipedia:Contents has 10 H2s and 23 H3s).
- **Confidence:** Confirmed.
- **Business impact:** Medium. AI answer engines quote question-and-answer style content most readily because it maps directly onto how users phrase queries, so this format gap limits how often your content is lifted into answers. This is the 10-point question-heading deduction in Content Quality & E-E-A-T.
- **Action:** Reframe suitable sections as questions on help and about pages, for example turning "Edit an article" into "How do I edit an article?" on /wiki/Help:Contents.

5. Seven content pages carry no structured data

- **What we saw:** Only /wiki/Main_Page carries schema (Article, Organization, ImageObject). The other 7 analysed pages return an empty `schemaTypes` array, including /wiki/Wikipedia:About, /wiki/Wikipedia:Contact_us and /wiki/Wikipedia:Contents. There is no WebSite schema on the homepage and no BreadcrumbList on any inner page.

- **Confidence:** Confirmed.
- **Business impact:** High. Structured data is how you tell AI systems explicitly what a page is and who published it, and without it these pages rely on the engine guessing. This drives the Schema & Structured Data deductions: 20 points for the 7 pages with no structured data, 5 for the missing WebSite schema, and 5 for the absent BreadcrumbList.
- **Action:** Add appropriate schema (for example Article or WebPage, plus BreadcrumbList) to the 7 unmarked pages and add WebSite schema to /wiki/Main_Page.

6. Five of eight pages have no author byline

- **What we saw:** 5 of 8 substantial pages return `hasAuthorByline: false`, named in the audit as /wiki/Wikipedia:About, /wiki/Wikipedia:Contact_us and /wiki/Wikipedia:Contents.
- **Confidence:** Confirmed.
- **Business impact:** Medium. Author or maintainer attribution is a core E-E-A-T (Experience, Expertise, Authoritativeness, Trust) signal that AI systems weigh when deciding whether to trust and cite a page. This is the 9-point byline deduction in Content Quality & E-E-A-T.
- **Action:** Add a visible maintainer or editorial attribution to /wiki/Wikipedia:About, /wiki/Wikipedia:Contact_us and /wiki/Wikipedia:Contents.

7. Weak entity signals: one sameAs link and no og:site_name

- **What we saw:** Only 1 external profile is declared via `sameAs` (http://www.wikidata.org/entity/Q5296), and the homepage declares no `og:site_name`.
- **Confidence:** Confirmed.
- **Business impact:** Medium. sameAs links and og:site_name help AI systems connect your pages to a single, recognised brand entity, and thin signals here make that connection harder. These are the 15-point sameAs and 10-point og:site_name deductions in Brand Authority & Entity Presence, which is capped at 80 because off-site presence is not checked.
- **Action:** Add `og:site_name` to the homepage and expand `sameAs` to point to your recognised official profiles.

8. No XML sitemap and one unreachable linked page

- **What we saw:** No XML sitemap was found at /sitemap.xml or referenced in robots.txt, and 1 linked page could not be fetched: https://en.wikipedia.org/wiki/Wikipedia:About_Today.
- **Confidence:** Confirmed.
- **Business impact:** Low. A sitemap helps crawlers discover pages efficiently and a broken link wastes crawl effort, but neither blocks access outright. These are the 5-point sitemap and 2-point unreachable-page deductions in Technical Foundation.
- **Action:** Publish an XML sitemap and reference it in robots.txt, and fix or remove the link to /wiki/Wikipedia:About_Today.

What's already working

- **Every AI crawler is allowed in robots.txt.** The `aiCrawlersBlockedInRobotsTxt` list is empty, so no major AI platform is blocked (AI Platform Readiness scored 100).
- **Your homepage carries meaningful schema.** /wiki/Main_Page declares Article, Organization and ImageObject types, giving AI systems clear structured cues on your most important page.
- **The site is served securely and fast.** HTTPS is in place, the homepage responded in 161ms, and strict-transport-security, content-security-policy and x-content-type-options headers are all present.
- **Content depth is strong.** Substantial pages such as /wiki/Wikipedia:Community_portal (7,967 words) and /wiki/Portal:Current_events (7,187 words) give AI systems rich material to draw on.
- **Canonical tags are clean.** No canonical mismatches were found across the analysed pages, so AI systems are not sent conflicting signals about which URL is authoritative.

AI VISIBILITY DASHBOARD

The table below shows the state of play for each major AI platform: whether its crawler can reach your site, the main gap standing in the way, and the priority action to close it. Crawler access is taken directly from the audit data.

AI Platform	Crawler Access	Key Gap	Priority Action
Google AI Overviews (Googlebot)	Allowed	No meta descriptions or publication dates to summarise from (Findings 1, 2)	Add meta descriptions and visible dates to all pages
ChatGPT Web Search (GPTBot)	Allowed	7 of 8 pages carry no structured data (Finding 5)	Add Article/WebPage schema to unmarked pages
Perplexity AI (PerplexityBot)	Allowed	No question-style headings for answer extraction (Finding 4)	Reframe help sections as questions
Google Gemini (Google-Extended)	Allowed	Missing bylines and dates weaken E-E-A-T (Findings 2, 6)	Add author attribution and updated dates
Bing Copilot (Bingbot)	Allowed	No XML sitemap for efficient discovery (Finding 8)	Publish a sitemap and list it in robots.txt

This dashboard shows what currently stands between your site and each platform, not a measurement of how often each platform actually cites you. Actual AI citations are not checked in this audit. The good news is that access is open across the board, so the work is about making your content easier to understand and trust, not about unblocking crawlers.

AI CRAWLER ACCESS

This section confirms which AI and search crawlers can reach your content. Every crawler below is currently allowed, based on an empty `aiCrawlersBlockedInRobotsTxt` list and a robots.txt file that was found.

AI Crawler	Platform	Status	Impact	Recommendation
Googlebot	Google Search / AI Overviews	Allowed	Critical	Keep allowed; ensure sitemap aids discovery
GPTBot	ChatGPT	Allowed	High	Keep allowed; add structured data to aid understanding
Bingbot	Bing / Copilot	Allowed	High	Keep allowed; publish an XML sitemap
PerplexityBot	Perplexity AI	Allowed	Medium	Keep allowed; add question-style headings
Google-Extended	Google Gemini	Allowed	Medium	Keep allowed; strengthen bylines and dates
ClaudeBot	Anthropic Claude	Allowed	Medium	Keep allowed; maintain clean, structured content
Applebot-Extended	Apple Intelligence	Allowed	Medium	Keep allowed; keep metadata complete

In plain terms: blocking AI crawlers is like closing your store during business hours. Your doors are wide open to every major AI platform, which is exactly where you want to be. The opportunity now is to make sure that once these crawlers walk in, they find well-labelled, well-attributed content they can confidently quote.

BRAND AUTHORITY

We start with what the crawl can confirm about your brand as an entity. Your homepage declares Organization schema, which is a positive signal that helps AI systems recognise you as a defined organisation. However, entity linking is thin: the crawl found only

one `sameAs` link, pointing to Wikidata (http://www.wikidata.org/entity/Q5296), and the homepage declares no `og:site_name`. Your brand name is broadly consistent, with the homepage title reading "Wikipedia, the free encyclopedia" and inner pages following a "[Page] - Wikipedia" pattern, but the missing og:site_name means one important consistency signal is absent. On navigation, an About page (/wiki/Wikipedia:About) and a Contact us page (/wiki/Wikipedia:Contact_us) were both crawled, though the audit's agent-readiness checks flag the contact path as not clearly linked, which is worth verifying.

Platform	What the crawl found	Why it matters for AI
Wikipedia	Not checked in this audit	Presence here is a strong entity anchor AI systems cross-reference
Wikidata	Linked from site (sameAs)	Confirmed structured entity link that ties your brand to a known ID
LinkedIn	Not checked in this audit	Corroborates organisation legitimacy and profile details
YouTube	Not checked in this audit	Video presence adds another authoritative reference point
Reddit	Not checked in this audit	Community discussion signals real-world relevance
Google Knowledge Panel	Not checked in this audit	A panel reflects strong entity recognition by Google
Crunchbase	Not checked in this audit	Business profile data reinforces organisation facts
GitHub	Not checked in this audit	Technical/project presence can support authority in relevant sectors

We recommend you check and claim your presence on the top two platforms for your sector, then link to them from your site so the connection is machine-readable. AI platforms build trust in a brand by cross-referencing it across multiple authoritative sources, so a single sameAs link to Wikidata is a good start but leaves most of that corroboration invisible to them. Note that we did not search any of these external platforms in this audit; the entries above reflect only what your own pages declare.

CITABILITY ANALYSIS

Citability measures how easily an AI engine can lift a clear, attributable answer from a page. The signals that matter most are depth, clear headings, question-style structure, structured data, visible authorship and dates. Because 8 pages were crawled, every analysed page is covered below with no page appearing in both lists.

Top Most Citable Pages

1. https://en.wikipedia.org/wiki/Main_Page Why it is citable: at 1,923 words with an H1, 8 H2 sections, an author byline and the only structured data in the crawl (Article, Organization, ImageObject), this is your best-labelled page. Improvement: add a meta description and a visible updated date, since both are currently absent.

2. https://en.wikipedia.org/wiki/Portal:Current_events Why it is citable: exceptionally deep at 7,187 words, with 7 H2s, 21 H3s and an author byline, giving AI systems rich, well-sectioned material. Improvement: add publication/updated dates and Article schema so this fast-moving page can be trusted for time-sensitive answers.

3. https://en.wikipedia.org/wiki/Wikipedia:Community_portal Why it is citable: the longest page at 7,967 words, with 6 H2s and an author byline. Improvement: add alt text to the 25 of 52 images that lack it and add structured data, as it currently carries none.

4. https://en.wikipedia.org/wiki/Help:Contents Why it is citable: 1,597 words with 8 clear H2 sections that map onto common tasks. Improvement: reframe those sections as questions (for example "How do I edit an article?") and add the missing alt text on 11 of 17 images.

5. https://en.wikipedia.org/wiki/Wikipedia:Contents Why it is citable: 1,057 words with a strong structural skeleton of 10 H2s and 23 H3s. Improvement: add structured data and an author byline, both of which are currently absent, and add a meta description.

Least Citable Pages

1. https://en.wikipedia.org/wiki/Wikipedia:Contact_us Why it is less citable: at 635 words it is the thinnest page, with no H2 sections, no structured data, no author byline and no meta description. Improvement: add two or more descriptive H2 sections and a meta description so AI systems can summarise how to reach you.

2. https://en.wikipedia.org/wiki/Help:Introduction Why it is less citable: 908 words with no H2 sections, no structured data and no author byline, so it lacks the scannable structure AI engines prefer. Improvement: introduce clear H2 headings, ideally in question form, and add Article or WebPage schema.

3. https://en.wikipedia.org/wiki/Wikipedia:About Why it is less citable: despite 1,204 words, it has no H2 sections, no structured data and no author byline, which weakens both structure and trust signals on a key brand page. Improvement: add section headings, an author or maintainer byline and Organization/WebPage schema, plus a meta description.

Improving your least citable pages is the highest-ROI content investment you can make for AI visibility. These pages already have traffic-worthy content but are missing the low-effort structural and metadata signals, headings, schema, bylines and descriptions, that decide whether an AI engine can confidently pull an answer from them and attribute it to you.

TECHNICAL HEALTH

Your technical foundation is one of the strongest parts of this audit. Your site is served securely, responds quickly and presents substantial content in the raw HTML that crawlers receive, which is exactly what AI systems need in order to read and quote your pages without running a browser. The two issues we found here are minor and low-impact: a missing XML sitemap and one linked page that could not be fetched. The table below translates each technical area into what we observed and what it means for your business.

Area	What we saw	Confidence	Business Impact
Core Web Vitals	Not measured by this crawl. Field data (real-user loading, interactivity and layout stability) is not captured here. Check in Google PageSpeed Insights and Search Console.	Unverified	Medium - slow or unstable pages can reduce how favourably both search and AI systems treat your content, so this should be confirmed independently.
Server-Side Rendering	Substantial content is present in the raw HTML of every analysed page (for example 7,967 words on /wiki/Wikipedia:Community_portal and 1,923 on /wiki/Main_Page), indicating content is served rather than assembled in the browser.	Likely	Low - this is a positive: AI crawlers can read your content directly without executing JavaScript, which removes a common barrier to AI visibility.
Mobile Optimisation	Every analysed page returns `hasViewport: true`, so all 8 pages declare a mobile viewport. Rendering quality on real phones is not tested in this audit.	Confirmed (viewport) / Unverified (real-device rendering)	Medium - a declared viewport is the baseline for mobile readiness, but you should still confirm layout and tap targets on real devices.
Security (HTTPS + Headers)	HTTPS is in place, and strict-transport-security, content-security-policy and x-content-type-options headers are all present.	Confirmed	Low - security is fully in order; this is a trust signal that is already working in your favour.
Page Speed (homepage response)	The homepage responded in 161ms.	Confirmed	Low - a fast server response helps crawlers fetch more of your pages efficiently; this is a strength.
Indexing Signals (sitemap + canonical)	No XML sitemap was found at /sitemap.xml or referenced in robots.txt. Canonical tags are clean with no mismatches across the 8 pages.	Confirmed	Low - clean canonicals prevent conflicting signals; the missing sitemap makes page discovery less efficient but does not block access (Finding 8).
IndexNow Protocol	Not measured by this crawl. Whether your site pings IndexNow when content changes is not something this audit can see. Confirm in Bing Webmaster Tools.	Unverified	Low - IndexNow can speed up how quickly Bing and Copilot pick up changes; verify and enable it if not already active.

What this means for your business: Technically, your site is in good shape and this is reflected in the Technical Foundation score of 93, which lost only 7 points (5 for the missing sitemap, 2 for the one unreachable page, /wiki/Wikipedia:About_Today). Because

your content is present in the raw HTML, you do not face the single most damaging technical problem for AI visibility, which is content that only appears after JavaScript runs. Your priorities here are small housekeeping tasks: publish a sitemap, fix the broken link, and independently confirm Core Web Vitals and IndexNow using the free tools named above.

SCHEMA & STRUCTURED DATA

Structured data is the machine-readable label that tells AI systems exactly what a page is, who published it and how it connects to your wider brand. Your homepage does this well, but the rest of your analysed pages carry nothing, which is the main reason the Schema & Structured Data component scored 70. The table below maps each schema type to where we found it and why it matters for AI.

Schema Type	Present	Where found	AI Impact
Organization	Yes	/wiki/Main_Page	High - drives entity recognition, helping AI systems identify you as a defined organisation.
Article + Author	Partial	Article present on /wiki/Main_Page only; no Author/authorship schema found on any page	High - authorship signals are a core trust cue; without an author property, AI systems cannot attribute your content confidently.
sameAs entity links	Yes (1 link)	/wiki/Main_Page, pointing to http://www.wikidata.org/entity/Q5296	High - sameAs links tie your pages to a recognised entity ID; one link is a start but leaves most corroboration invisible (Finding 7).
WebSite	No	Not found on /wiki/Main_Page	Medium - WebSite schema helps define your site as a whole and enables sitelink search features; this is the 5-point homepage deduction.
WebPage / Article on inner pages	No	Absent on all 7 inner pages (/wiki/Wikipedia:About, /wiki/Wikipedia:Contact_us, /wiki/Wikipedia:Contents, /wiki/Portal:Current_events, /wiki/Help:Contents, /wiki/Help:Introduction, /wiki/Wikipedia:Community_portal)	High - this is the 20-point deduction for 7 pages with no structured data; these pages rely on the engine guessing what they are.
BreadcrumbList	No	Not found on any inner page	Low - breadcrumbs help AI systems understand your site hierarchy; this is the 5-point deduction.

What this means for your business: Your best-labelled page, /wiki/Main_Page, shows what good structured data looks like, but the benefit stops there. The 20-point deduction for the 7 unmarked pages (Finding 5) is the largest single gap in this component, and it affects important pages such as your About and Contact us pages. Adding Article or WebPage schema to those seven pages, WebSite schema to the homepage, and BreadcrumbList across inner pages would move this component close to full marks. This is high-value work because structured data is the most direct way to control how AI systems classify and attribute your content.

EXPERIMENTAL SIGNALS (NOT SCORED)

The signals below are emerging conventions that no major AI platform has publicly confirmed it uses, so we report them purely for your awareness. They do not affect your GEO Readiness Score, they are not part of the Prioritised Action Plan, and their absence is not penalised anywhere in this report.

llms.txt

Status: /llms.txt not present. /llms-full.txt not present. `llms.txt` is a proposed plain-text file that would give AI tools a curated guide to your most important content. Adoption by AI platforms is unconfirmed, so this is optional and low priority.

Agentic Resource Discovery (ARD)

Status: Not implemented (no /.well-known/ai-catalog.json found). **Priority:** Low (Monitor / Early adopter). ARD is a proposed manifest that would let AI agents discover a site's capabilities and APIs. Real-world adoption is very low, so its absence carries no penalty and there is no urgency to implement it.

Agent Readiness

This looks at whether an autonomous AI agent could actually complete useful work on your site, such as finding pricing, identifying who runs the site, reading your policies, using an API or downloading resources. Your **indicative Agent Readiness sub-score is 38 out of 100**. This is an indicative measure only and does not feed the GEO Readiness Score.

Signal	Status	Why it matters to an agent
Contact path discoverable	Not detected	An agent cannot reliably find how to reach you, limiting tasks that need human contact.
Machine-readable ownership (contactPoint)	Not detected (no schema contactPoint)	Without a structured contact point, an agent cannot confirm who to contact programmatically.
Machine-readable ownership (sameAs)	Present	An agent can link your site to a known entity ID, which helps confirm identity.
Privacy policy linked	Not detected	An agent cannot check how data is handled before acting.
Terms linked	Present	An agent can locate the terms that govern use of your content.
Documentation linked	Present	An agent can find help/documentation to understand how the site works.
API spec discoverable (openapi.json)	Not detected	An agent has no machine-readable interface to interact with your services directly.
Downloadable resources	None found (0)	An agent has no files it can retrieve to complete a task.
Breadcrumb navigation schema	Not detected	An agent has weaker structural cues for navigating your site hierarchy.

Recommendations (useful regardless of AI adoption): First, make sure your contact page (/wiki/Wikipedia:Contact_us) is clearly and consistently linked from your main navigation, since a discoverable contact path helps human visitors as much as agents. Second, add a machine-readable contactPoint to your Organization schema, which reinforces the entity signals that already benefit your normal search presence. Neither of these is urgent and neither affects your score.

PRIORITISED ACTION PLAN

This is the part of the report to act on. Every action below traces back to a Key Finding or a specific audit data point, and names the exact pages or files involved. Experimental signals are deliberately excluded. Actions are grouped by how quickly they can be done, and each is rated by Impact (High, Medium, Low) and Effort.

Quick Wins (This Week)

Each of these takes under four hours and removes a specific scoring deduction.

#	Action	Addresses	Impact	Effort
1	Add a unique 150-160 character meta description to all 8 pages, starting with /wiki/Main_Page, /wiki/Wikipedia:About and /wiki/Wikipedia:Contact_us	Finding 1	High	Low
2	Expose a visible "last updated" date and machine-readable `dateModified` on all pages, prioritising /wiki/Portal:Current_events	Finding 2	High	Low
3	Add `og:site_name` to /wiki/Main_Page	Finding 7	Medium	Low
4	Expand `sameAs` beyond the single Wikidata link to your recognised official profiles	Finding 7	Medium	Low

5	Fix or remove the broken link to https://en.wikipedia.org/wiki/Wikipedia:About_Today	Finding 8	Low	Low
6	Publish an XML sitemap and reference it in robots.txt	Finding 8	Low	Low

Medium-Term Improvements (This Month)

Each of these takes roughly one to five days.

#	Action	Addresses	Impact	Effort
1	Add Article or WebPage schema to the 7 unmarked pages and WebSite schema to /wiki/Main_Page	Finding 5	High	Medium
2	Add descriptive alt text to the 59 images missing it, starting with /wiki/Wikipedia:Community_portal (25), /wiki/Help:Contents (11) and /wiki/Portal:Current_events (9)	Finding 3	Medium	Medium
3	Add author or maintainer attribution to /wiki/Wikipedia:About, /wiki/Wikipedia:Contact_us and /wiki/Wikipedia:Contents	Finding 6	Medium	Medium
4	Add H1/H2 structure to the pages that lack it: /wiki/Wikipedia:About, /wiki/Wikipedia:Contact_us and /wiki/Help:Introduction	Finding (H1/H2 deduction), Content	Medium	Medium
5	Verify and enable IndexNow in Bing Webmaster Tools, and check Core Web Vitals in PageSpeed Insights	Technical Health (Unverified items)	Low	Medium

Strategic Initiatives (This Quarter)

Each of these takes weeks to months.

#	Action	Addresses	Impact	Effort
1	Reframe suitable sections as question-style headings across help/about pages, for example "How do I edit an article?" on /wiki/Help:Contents	Finding 4	Medium	High
2	Add BreadcrumbList schema across all inner pages to expose site hierarchy	Finding 5	Low	High
3	Check, claim and link your presence on the top external platforms for your sector so entity corroboration becomes machine-readable	Finding 7 (off-site, not checked in this audit)	Medium	High

Estimated Score Impact

Every point in your score comes from a listed deduction, so the projections below are arithmetic, not a guess.

Quick Wins remove the following deductions:

- **Content Quality & E-E-A-T:** meta description (10 points, via #1) and publication date (10 points, via #2) are removed. Component rises from 45 to 65, restoring 20 points. Weighted effect: 20 × 0.25 = **+5.0**.
- **Technical Foundation:** the unreachable page (2 points, via #5) is removed. Component rises from 93 to 95. Weighted effect: 2 × 0.20 = **+0.4**. (The sitemap deduction of 5 is also addressed by #6 and is counted under full implementation below for clarity, but if #6 lands this week it lifts Technical to 100 and adds a further 1.0.)
- **Brand Authority & Entity Presence:** og:site_name (10 points, via #3) and sameAs (15 points, via #4) are removed, a nominal 25 points. However, Brand is capped at 80 because off-site presence is not checked, so only 5 effective points count, taking the component from 75 to 80. Weighted effect: 5 × 0.15 = **+0.75**.

Applying the Quick Wins as listed (items 1-5), the new component scores are: Platform 100, Content 65, Technical 95, Schema 70, Brand 80. The weighted total is 25.0 + 16.25 + 19.0 + 10.5 + 12.0 = **82.75, which rounds to 83**. Your readiness score moves from **77 to 83**.

Full implementation of all three action groups removes every remaining deduction that this crawl can score, subject to the Brand cap:

- Content rises to 100 (removing the H1/H2, question-heading, byline, date, meta description and alt-text deductions).

- Technical rises to 100 (removing the sitemap and unreachable-page deductions).
- Schema rises to 100 (removing the WebSite, 7-page and BreadcrumbList deductions).
- Brand rises to its cap of 80; it cannot exceed 80 without verified off-site presence, so the external-platform work in the strategic plan supports that ceiling but cannot lift the on-site score past it.

The new component scores would be: Platform 100, Content 100, Technical 100, Schema 100, Brand 80. The weighted total is 25.0 + 25.0 + 20.0 + 15.0 + 12.0 = **97.0**, rounding to **97**.

To be clear: these figures describe the change in your GEO Readiness Score based on removing observed, scored deductions. They are not a prediction of traffic, AI citations or revenue, and no results are guaranteed.

APPENDIX

Methodology

We analysed 8 pages on the en.wikipedia.org domain on 27 September 2026:

- https://en.wikipedia.org/wiki/Main_Page
- https://en.wikipedia.org/wiki/Wikipedia:About
- https://en.wikipedia.org/wiki/Wikipedia:Contact_us
- https://en.wikipedia.org/wiki/Wikipedia:Contents
- https://en.wikipedia.org/wiki/Portal:Current_events
- https://en.wikipedia.org/wiki/Help:Contents
- https://en.wikipedia.org/wiki/Help:Introduction
- https://en.wikipedia.org/wiki/Wikipedia:Community_portal

One further linked page, https://en.wikipedia.org/wiki/Wikipedia:About_Today, could not be fetched during the crawl and is recorded as unreachable. No page carried a truncation note.

What was checked: robots.txt and AI crawler rules, XML sitemap presence, canonical tags, page titles and meta descriptions, heading structure (H1/H2/H3 and question-style headings), word counts, structured data types, images and alt text, author bylines and publication dates, security headers, and homepage response time.

What was NOT checked: off-site brand presence (Wikipedia, Wikidata, LinkedIn, YouTube, Reddit, Knowledge Panel, Crunchbase and similar), Core Web Vitals field data, backlinks, keyword rankings, and actual AI citations. These require separate tools and are marked Unverified throughout this report.

How the score is calculated: each of the five components starts at 100 and loses points only for problems observed in the crawl. The weighted components (AI Platform Readiness 25%, Content Quality & E-E-A-T 25%, Technical Foundation 20%, Schema & Structured Data 15%, Brand Authority & Entity Presence 15%) sum to the overall score. The same site always produces the same score. Brand Authority is capped at 80 because off-site presence is not checked. Experimental signals are never scored.

Date of analysis: 27 September 2026.

Confidence levels

- **Confirmed:** directly present in the audit data (for example, a page returns no meta description).
- **Likely:** a reasoned inference from the audit data, with the basis stated (for example, substantial raw HTML suggests server-side rendering).
- **Unverified:** something this crawl cannot see (for example, Core Web Vitals field data or off-site brand presence); confirm with the named external tool.

References

- Google Search Central documentation (structured data, crawling and indexing guidance).
- Schema.org (vocabulary definitions for Organization, WebSite, Article, BreadcrumbList, sameAs and related types).
- AI crawler user-agent documentation published by each platform (Google, OpenAI, Microsoft Bing, Perplexity, Anthropic and Apple).

Glossary

Term	Definition
GEO	Generative Engine Optimisation: preparing a site to be found, understood and cited by AI-driven search and answer engines.
AI Overviews	Google's AI-generated answer summaries shown above traditional search results.
E-E-A-T	Experience, Expertise, Authoritativeness and Trust: signals used to judge content credibility.
SSR	Server-Side Rendering: delivering fully-formed HTML from the server so crawlers can read content without running JavaScript.
Core Web Vitals	Google's real-user metrics for loading speed, interactivity and visual stability.
JSON-LD	A format for embedding structured data (Schema.org) in a page as machine-readable code.
sameAs	A Schema.org property linking an entity to its authoritative profiles elsewhere.
Canonical tag	An HTML signal declaring the preferred URL for a page to avoid duplicate-content confusion.
IndexNow	A protocol that instantly notifies search engines when content changes.
llms.txt	A proposed plain-text file guiding AI tools to a site's key content; adoption is unconfirmed.
Topical Authority	The depth and breadth of trusted coverage a site has on a subject area.
Agent Readiness	How well an autonomous AI agent could complete tasks on a site, such as finding contact details or using an API.

GEO is a rapidly evolving field; recommendations should be reviewed monthly as AI platform requirements evolve.