

2026 Minimum AI SEO / GEO

Website Technical Standard

Developer Implementation Specification | Version 1.0 | 2026

Document Purpose

This document defines the minimum technical requirements for AI-readable website architecture as required for citation eligibility, AI Overview inclusion, and LLM reference in 2026.

Scope

Machine readability, AI crawler access, structured data, schema markup, content architecture, and technical performance. Excludes marketing strategy, keyword strategy, backlink strategy, and future (2027+) frameworks.

1. Executive Summary

The 2026 Minimum AI SEO / GEO Website Technical Standard defines the lowest acceptable threshold of technical and structural implementation required for a website to be readable, extractable, and citable by AI-driven search systems. This is not an optimisation guide. It is a baseline specification.

Websites that do not meet this standard are structurally opaque to AI systems. They may still rank in traditional search engines but will be bypassed or deprioritised by AI Overviews, large language model citation engines, and generative search interfaces that require structured, unambiguous content to extract and attribute.

1.1 What This Standard Defines

The standard covers:

- The minimum structural requirements for every indexable page
- The schema markup types and fields required across page types
- The crawler access configuration required for AI bots
- The entity and knowledge architecture required for consistent AI attribution
- The technical performance baseline required for reliable AI crawling
- The internal linking structure required to signal content hierarchy to AI systems

1.2 Why This Standard Exists

Search behaviour is undergoing a structural change. Users increasingly receive answers from AI-generated summaries rather than clicking through to web pages. AI Overviews (Google), Perplexity, ChatGPT web search, and Claude all pull content from websites that meet specific structural criteria.

The AI systems generating these answers do not browse pages as humans do. They parse structured content, extract entities, validate schema, and assess the reliability and specificity of information. A website structured for human readability alone is insufficient for AI citation eligibility.

This standard exists to define what must be in place for a website to function as a machine-readable knowledge source in 2026.

Key Principle

- AI systems extract answers from content that is structured, factual, and unambiguous.
- Compliance with this standard is a prerequisite for citation eligibility, not a guarantee of it.
- Every requirement in this document is testable and binary: it is either present or absent.

2. Mandatory Page Structure Standard

Every indexable page on a compliant website must follow the structural sequence defined in this section, without exception. The order of elements is not advisory; it is required. AI extraction systems read content in document order and apply predictive parsing based on structural position.

2.1 Required Page Structure Sequence

The following sequence must be present on every indexable page, in this order:

Position	Element	Specification
1	H1 (Page Title)	One per page. Must exactly match the primary topic entity. Maximum 70 characters. Must be rendered as an HTML <h1> element. Must match the page's schema name field.
2	AI Summary Paragraph	First paragraph after H1. 40-80 words. Must answer the primary implicit or explicit question the page addresses. Written in declarative factual sentences. No promotional language. No rhetorical questions.
3	Key Facts Block	Follows the summary paragraph. Structured list of 3-7 factual statements about the page topic. Each fact must be independently verifiable. Presented as an HTML list or definition list. Mirrors fields in the page's schema markup.
4	H2 Structured Sections	Main content divided into H2-headed sections. Each section addresses a distinct sub-topic or user question. Minimum two H2 sections per page. Each H2 must be followed by at least one full paragraph before the next heading.
5	FAQ Section	Minimum five question-and-answer pairs. Questions must reflect real user queries. Answers must be 40-120 words. Must be marked up with FAQPage schema. Must appear before the summary/conclusion block.
6	Summary / Conclusion Block	Final section of the page. Restates the primary factual answer in 50-100 words. Uses a consistent H2 label such as 'Summary' or 'Key Takeaways'. Must not introduce new claims not covered elsewhere on the page.

2.2 H1 Rules

- Exactly one H1 per page. No exceptions.
- The H1 must be the first heading element in the page body.
- The H1 text must match the schema name or headline field on the same page.
- The H1 must not be used for decorative or navigational text.
- The H1 must be rendered in the HTML document as <h1>, not styled with CSS to visually resemble a heading.
- Maximum length: 70 characters including spaces.

2.3 AI Summary Paragraph Requirements

The AI summary paragraph is the highest-priority extraction target for AI systems parsing the page. It must meet the following criteria:

- Position: immediately following the H1, before any other content block.
- Length: 40-80 words.
- Tense: present tense or timeless declarative statements.
- Content: must answer the question implied by the page title without requiring the reader to continue further.
- Language: factual, specific, and devoid of promotional modifiers such as 'leading', 'world-class', or 'innovative'.
- Format: single paragraph, no bullet points, no sub-headings within it.

2.4 Key Facts Block Requirements

- Minimum 3 facts, maximum 7 facts.
- Each fact must be a complete, standalone statement.
- Facts must be quantifiable, attributable, or directly verifiable where possible.
- Facts must correspond to fields present in the page's schema markup.
- Implemented as an HTML unordered list () or definition list (<dl>).
- Must include a visible label such as 'Key Facts' or 'At a Glance' using an appropriate heading or label element.

2.5 FAQ Section Requirements

- Minimum five question-and-answer pairs per page.
- Questions must be phrased in natural language, not keyword strings.
- Each answer must be 40-120 words.
- Answers must be self-contained and not require the reader to read surrounding content to be understood.
- The FAQ section must be marked up using FAQPage JSON-LD schema (see Section 4).
- FAQ must appear as visible on-page content, not hidden via CSS or JavaScript collapse by default.
- The FAQ heading must use an H2 element.

3. Content Architecture Rules

Content architecture defines how information is organised across the website and within individual pages. These rules govern the relationship between pages, the structure of content within them, and the constraints on writing style required for AI extraction.

3.1 Single Entity Per Page Rule

Each indexable page must address one and only one primary entity or topic. An entity is a discrete, nameable thing: a service, a product, a location, a person, a concept, or an organisation.

- A page must not attempt to cover multiple services or topics unless the page is explicitly an index or hub page.
- All schema markup on the page must reference the same primary entity.
- All H2 sections must be sub-topics of the primary entity, not separate entities.
- If a page is found to contain multiple primary entities, it must be split into separate pages.

3.2 Intent-Based Page Separation

Pages must be separated by user intent. A single URL must not serve multiple intent types simultaneously. The four required intent types are:

- Informational: the user seeks to learn or understand something.
- Transactional: the user seeks to take an action, make a purchase, or initiate a contact.
- Comparative: the user seeks to evaluate options.
- Local: the user seeks information about a specific geographic location.

A page optimised for informational intent must not include transactional CTAs as primary page elements. A transactional page must not lead with educational content.

3.3 Front-Loaded Answer Requirement

The primary answer to the implicit or explicit question of the page must appear within the first 150 words of the body content. This is a hard requirement, not a recommendation. AI extraction systems apply decaying weight to content positioned further from the start of the document body.

- The primary answer must appear in the AI summary paragraph (see 2.3).
- Supporting evidence, context, and elaboration follow the primary answer.
- Introductory paragraphs that do not contain the answer are not compliant.

3.4 Factual Writing Constraints

- All factual claims must be attributable to a named source, a defined standard, or the entity itself.
- Superlatives require evidence: 'the most common' must be supported by a cited source or data point.
- Dates, figures, and statistics must be current and accurate at the time of publication.
- Passive constructions that obscure the source of a claim are prohibited: 'It is said that' or 'Many experts believe' do not constitute factual writing.
- Promotional language is prohibited: words such as 'leading', 'cutting-edge', 'innovative', 'best-in-class', and 'world-class' are not permitted unless supported by verifiable evidence.

3.5 AI Answer Block Requirement

In addition to the AI summary paragraph, each page must contain at least one section structured explicitly for AI extraction. This is referred to as the AI Answer Block. It must:

- Begin with an H2 or H3 heading that frames a specific question.
- Immediately follow the heading with a direct, factual answer in the first sentence.
- Continue with supporting context in the following sentences.
- Contain no promotional language.
- Be implementable as a FAQPage schema entry (see Section 4).

4. Schema Markup Requirements

All schema markup must be implemented as JSON-LD. Microdata and RDFa implementations are not accepted as compliant. JSON-LD blocks must be placed within a `<script type="application/ld+json">` tag in the `<head>` of the HTML document, or immediately before the closing `</body>` tag.

All schema must pass validation against the Google Rich Results Test and the Schema.org validator without errors. Warnings are acceptable. Errors are not.

4.1 Organization Schema

Required on: the homepage. Recommended on: the About page and Contact page.

- `@type`: Organization (or a more specific subtype such as LocalBusiness where applicable)
- `@id`: a canonical URL representing the organisation, e.g. `https://example.com/#organisation`
- `name`: legal or trading name of the organisation
- `url`: homepage URL
- `logo`: URL of logo image, minimum 112x112px
- `sameAs`: array of verified social media profile URLs and directory listings
- `contactPoint`: at minimum one ContactPoint with `contactType`, `telephone`, and `areaServed`

4.2 WebSite Schema

Required on: the homepage only.

- `@type`: WebSite
- `@id`: canonical site URL with fragment, e.g. `https://example.com/#website`
- `name`: site name
- `url`: site root URL
- `potentialAction`: SearchAction if site search is available

4.3 WebPage Schema

Required on: every indexable page.

- `@type`: WebPage (or a more specific subtype: AboutPage, ContactPage, etc.)
- `@id`: the canonical URL of the page
- `name`: must match the page H1 exactly
- `url`: canonical URL
- `description`: must match the meta description of the page
- `inLanguage`: ISO 639-1 language code, e.g. 'en-GB'
- `isPartOf`: reference to WebSite `@id`
- `about`: reference to the primary entity schema on the page
- `datePublished`: ISO 8601 date
- `dateModified`: ISO 8601 date, updated on any substantive content change

4.4 BreadcrumbList Schema

Required on: all pages except the homepage.

- @type: BreadcrumbList
- itemListElement: array of ListItem entries
- Each ListItem must include: @type ListItem, position (integer starting at 1), name (human-readable label), item (canonical URL)
- The breadcrumb schema must match the visible breadcrumb navigation on the page exactly.

4.5 Article Schema

Required on: all blog posts, news articles, guides, and long-form editorial content.

- @type: Article (or BlogPosting, NewsArticle as appropriate)
- headline: must match H1, maximum 110 characters
- author: Person schema reference
- datePublished and dateModified: ISO 8601 dates
- publisher: Organization schema reference
- image: ImageObject with url, width, and height
- mainEntityOfPage: WebPage @id reference

4.6 FAQPage Schema

Required on: any page containing a FAQ section. Must exactly mirror the visible on-page FAQ content.

- @type: FAQPage
- mainEntity: array of Question entities
- Each Question: @type Question, name (question text), acceptedAnswer containing @type Answer and text (answer text)
- Question text in schema must exactly match the visible H3 or question text on the page.
- Answer text in schema must exactly match or be a direct extract of the visible answer text.

4.7 LocalBusiness Schema

Required on: any website for a business that operates from a physical location or serves a defined geographic area.

- @type: LocalBusiness (or specific subtype such as AutoDealer, RealEstateAgent)
- name: trading name
- address: PostalAddress with streetAddress, addressLocality, addressRegion, postalCode, addressCountry
- telephone: in E.164 international format
- openingHoursSpecification: array with dayOfWeek, opens, closes
- geo: GeoCoordinates with latitude and longitude
- hasMap: URL of Google Maps or equivalent listing
- priceRange: string indicating price bracket

4.8 Service Schema

Required on: all service description pages.

- `@type`: Service
- `name`: service name
- `description`: factual description, 50-160 characters
- `provider`: Organization schema reference
- `areaServed`: GeoCircle, GeoShape, or named Place
- `serviceType`: string defining the category of service

4.9 Product and Offer Schema

Required on: all product pages and e-commerce listings.

- `@type`: Product
- `name`, `description`, `image` (ImageObject), `sku`
- `offers`: Offer with `@type` Offer, `price`, `priceCurrency` (ISO 4217), `availability` (schema.org/InStock etc.), `url`
- `brand`: Brand or Organization reference
- `aggregateRating`: if reviews are present, include `ratingValue`, `bestRating`, `worstRating`, `ratingCount`

4.10 Person Schema

Required on: author profile pages and author bylines on Article schema.

- `@type`: Person
- `name`: full name
- `url`: profile page URL
- `sameAs`: LinkedIn profile URL at minimum
- `jobTitle`: current role
- `worksFor`: Organization schema reference

4.11 Schema Placement and Validation Rules

- All JSON-LD must be placed in the `<head>` or immediately before `</body>`. It must not be injected after page load via client-side JavaScript that runs asynchronously.
- Each JSON-LD block must be independently valid. Do not nest multiple `@type` definitions in a single block unless they are genuinely nested entities.
- The `mainEntity` relationship must connect the `WebPage` schema to the primary content schema on the page.
- All schema must pass the Google Rich Results Test (<https://search.google.com/test/rich-results>) with zero errors before deployment.

5. AI Crawler Access Layer

AI systems access websites through dedicated crawlers that respect robots.txt directives and, increasingly, llms.txt guidance files. A compliant website must be explicitly configured to permit access to these crawlers and must provide structured metadata files that communicate the site's purpose and content structure.

5.1 robots.txt Requirements

The robots.txt file must be present at the root of the domain (<https://example.com/robots.txt>). The following crawlers must be explicitly allowed:

```
# AI Search Crawlers - Required Allowances
User-agent: GPTBot
Allow: /

User-agent: ChatGPT-User
Allow: /

User-agent: Google-Extended
Allow: /

User-agent: PerplexityBot
Allow: /

User-agent: ClaudeBot
Allow: /

User-agent: anthropic-ai
Allow: /

User-agent: CCBot
Allow: /

User-agent: Applebot-Extended
Allow: /

# Sitemap Declaration
Sitemap: https://example.com/sitemap.xml
```

5.1.1 Required Exclusions

The following paths must be explicitly disallowed for all crawlers including AI crawlers:

- /admin/ and all administrative back-end paths
- /wp-admin/ or CMS-equivalent administration directories
- /checkout/ and payment processing paths
- /user/ and /account/ paths containing personal user data
- /api/ endpoints not intended for public consumption
- Any path containing session tokens or query parameters with personal data

5.1.2 Crawl Hygiene Rules

- The robots.txt file must not use wildcard disallow rules that could inadvertently block AI crawlers.
- Crawl-delay directives must not exceed 10 seconds.
- The robots.txt file must be served with HTTP 200 status. A 404 or 500 response is non-compliant.
- The Sitemap declaration within robots.txt must reference a valid, accessible sitemap URL.

5.2 llms.txt Requirements

The llms.txt file must be present at <https://example.com/llms.txt>. It must be served as plain text (Content-Type: text/plain) and must follow the emerging llms.txt standard format.

Required fields:

- # [Site Name]: the site name as an H1-level Markdown heading
- > [Site purpose]: a blockquote summarising the site's purpose in one sentence
- A prose section defining the primary topics and entities covered by the site
- A ## Pages section listing the key URL paths with a one-line description of each
- A ## Content Guidelines section specifying any access restrictions or usage notes

```
# Example Business Name

> A UK-based provider of [service type] for [target audience].

This site covers [primary topic], [secondary topic], and [tertiary topic].
Content is factual, professional, and updated regularly.

## Key Pages

- /services/: Overview of all services offered
- /services/[service-name]/: Detailed specification of [service name]
- /about/: Organisation background and credentials
- /contact/: Contact information and enquiry form

## Content Guidelines

All content on this site may be used for AI training and citation.
Attribution to [Business Name] (https://example.com) is requested.
```

5.3 llms-full.txt (Recommended)

The llms-full.txt file provides an expanded, structured site map for AI systems that request additional detail. It is not currently a mandatory requirement but is strongly recommended for sites with more than 20 indexable pages.

Format requirements:

- Must include all indexable URLs with their intent type and primary entity.
- Must include the date of last substantive content update for each page.
- Must be formatted as structured plain text or Markdown.

- Must be referenced in the llms.txt file with a link.

6. Entity and Knowledge Structure Requirements

6.1 Single Primary Entity Per Page

Each page must have a defined primary entity. The primary entity is the main subject of the page, corresponding to the H1, the schema @type, and the predominant content.

- The primary entity must be named in the H1.
- The primary entity must be defined in the schema markup on the page.
- The primary entity must be referenced consistently throughout the page using the same name, not variations or synonyms, unless synonyms are explicitly introduced and defined.
- No page may have more than one primary entity. Secondary entities may be referenced but must not be given equal structural prominence.

6.2 Supporting Entity Rules

- Supporting entities may appear in body content, schema markup, and internal links.
- Each supporting entity must be referenced by a consistent canonical name.
- Supporting entities referenced multiple times on a site should link to a dedicated page for that entity where one exists.
- Supporting entities must not be given H1 or H2 prominence on a page where they are not the primary entity.

6.3 Site-Wide Entity Consistency Rules

- Entity names must be identical across all pages, schema markup, meta titles, meta descriptions, and navigation elements.
- If an entity name changes (e.g. a rebrand), all references across the site must be updated in the same deployment.
- Organisation name must be consistent across: the Organization schema, the WebSite schema, the footer, the About page, and all LocalBusiness schema instances.
- A master entity list should be maintained and referenced during content authoring. This list defines the canonical name for every named entity on the site.

6.4 Schema and Visible Content Alignment

- The name field in schema must exactly match the H1 text on that page.
- The description field in schema must exactly match the meta description.
- FAQ question and answer text in FAQPage schema must exactly match the visible on-page text.
- Address, telephone, and opening hours in LocalBusiness schema must exactly match the visible content on the Contact page.
- Discrepancies between schema content and visible page content will cause AI extraction errors and may trigger trust penalties from AI systems.

6.5 Entity Clarity Requirements for AI Extraction

- Pronouns that refer to the primary entity should be avoided where they could create ambiguity. Prefer named references.
- The primary entity's full canonical name must appear at least once in the AI summary paragraph.
- If the primary entity has an acronym or abbreviation, the full name must be used on first reference, with the abbreviation in parentheses.
- The entity's relationship to other entities must be defined explicitly: 'X is a service provided by Y' is compliant; 'our service' is not, unless Y has been established earlier in the paragraph.

7. Intent-Based Website Architecture

Pages must be designed around a single user intent. The intent determines the required page structure, content expectations, and schema priority.

7.1 Informational Intent Pages

Definition: pages where the user's primary goal is to understand a topic, concept, or entity.

Required Structure

- H1: topic name as a noun phrase (not a question)
- AI summary paragraph: defines the topic in 40-80 words
- Key Facts block
- Minimum four H2 sections covering: definition, context, application, and considerations
- FAQ section with minimum five questions
- Summary block

Content Expectations

- Content must be authoritative and factual throughout.
- No calls to action above the FAQ section.
- References and sources cited where specific data is used.

Schema Requirements

- WebPage (or Article for long-form), FAQPage, BreadcrumbList

7.2 Transactional Intent Pages

Definition: pages where the user's primary goal is to initiate a commercial action: purchase, enquiry, booking, or sign-up.

Required Structure

- H1: service or product name with implicit or explicit action framing
- AI summary paragraph: states what the page enables the user to do
- Key Facts block: price, availability, process, timeframe
- H2 sections covering: what is included, eligibility or requirements, the process
- FAQ section focused on pre-purchase or pre-enquiry questions
- Summary with primary call to action

Content Expectations

- Pricing must be stated where applicable. 'Contact us for pricing' without any indicative figure reduces AI citation eligibility.
- The process for transacting must be explained in explicit steps.

Schema Requirements

- WebPage, Service or Product/Offer, FAQPage, BreadcrumbList

7.3 Comparative Intent Pages

Definition: pages where the user's primary goal is to evaluate two or more options, products, services, or approaches.

Required Structure

- H1: must name both or all subjects being compared
- AI summary paragraph: states the outcome of the comparison in factual terms
- Comparison table or structured list with consistent criteria for each option
- H2 sections for each option being compared
- FAQ section addressing common comparison questions
- Summary with explicit recommendation or verdict where appropriate

Content Expectations

- Criteria used for comparison must be consistent across all options.
- Comparisons must be factual and evidence-based, not opinion-led.
- The page must not advocate for one option unless the page is explicitly branded as an opinion piece with an identified author.

Schema Requirements

- WebPage, FAQPage, BreadcrumbList; ItemList if comparing discrete products or services

7.4 Local Intent Pages ---

Definition: pages where the user's primary goal is to find information relevant to a specific geographic location.

Required Structure

- H1: primary service or entity name plus location name
- AI summary paragraph: states what the business or entity provides and in which specific location
- Key Facts block: address, telephone, hours, service area
- H2 sections: services available at this location, accessibility information, how to reach the location
- FAQ section: location-specific questions
- Summary with address and contact details restated

Content Expectations

- Location must be named in the first sentence of the AI summary paragraph.
- Address must appear as visible text on the page, not only in schema.
- Opening hours must appear as visible text on the page.

Schema Requirements

- LocalBusiness, WebPage, FAQPage, BreadcrumbList

8. Technical SEO and Performance Baseline

Technical performance directly affects AI crawler accessibility. Slow or unstable pages may be deprioritised or incompletely indexed by AI crawlers. The requirements in this section define the minimum technical performance threshold for compliant operation.

8.1 Core Web Vitals Targets

Metric	Good (Pass)	Needs Improvement	Poor (Fail)
LCP (Largest Contentful Paint)	Under 2.5 seconds	2.5s - 4.0s	Over 4.0 seconds
INP (Interaction to Next Paint)	Under 200 milliseconds	200ms - 500ms	Over 500 milliseconds
CLS (Cumulative Layout Shift)	Under 0.1	0.1 - 0.25	Over 0.25
TTFB (Time to First Byte)	Under 800 milliseconds	800ms - 1800ms	Over 1800 milliseconds

Core Web Vitals must be measured using Google PageSpeed Insights (<https://pagespeed.web.dev>) against the mobile assessment. 75th percentile performance across the site must meet the Good threshold for all three primary metrics (LCP, INP, CLS).

8.2 Rendering Requirements

- Server-Side Rendering (SSR) or Static Site Generation (SSG) is required. Client-Side Rendering (CSR) alone is non-compliant.
- All content that must be read by AI crawlers must be present in the initial HTML response. Content injected by JavaScript after page load is not guaranteed to be accessible to all AI crawlers.
- Hydration frameworks (Next.js, Nuxt.js, SvelteKit, Astro) are acceptable if they produce server-rendered HTML for the initial response.
- Lazy-loaded content that is part of the main page body must have a non-JavaScript fallback or must be present in the initial HTML.

8.3 CDN Requirements

- All production websites must be served via a Content Delivery Network (CDN).
- CDN must serve pages from a point of presence within the UK for UK-targeted websites.
- CDN must support HTTP/2 or HTTP/3.
- Edge caching must be configured for static assets with a minimum cache TTL of 31536000 seconds (one year) for versioned assets.

8.4 Image Optimisation Standards

- All images must be served in WebP format. AVIF is acceptable as an alternative or supplement.

- JPEG and PNG are not acceptable as the primary served format for new implementations.
- All images must have descriptive alt text. Alt text must describe the image content factually, not serve as a keyword insertion vehicle.
- Hero images and images above the fold must use the loading='eager' attribute. All other images must use loading='lazy'.
- All images must include explicit width and height attributes to prevent CLS.

8.5 Indexation Control

- Every page must have a canonical URL declared in the <head> as: <link rel='canonical' href=[canonical-url]>
- The canonical URL must be an absolute URL including the protocol.
- Paginated content must use rel='next' and rel='prev' pagination signals.
- The XML sitemap must include all indexable pages and must exclude non-canonical URLs, paginated pages (except page 1), and pages with noindex directives.
- The sitemap must be updated within 24 hours of any new page being published or any existing page being substantially modified.
- The sitemap must include lastmod dates in ISO 8601 format.

8.6 Duplicate Content Prevention

- www and non-www versions of the domain must 301 redirect to a single canonical version.
- HTTP must 301 redirect to HTTPS.
- Trailing slash and non-trailing slash variants must be consistent and must redirect non-canonical variants.
- URL parameters used for tracking (utm_source, fbclid, etc.) must not create indexable duplicate URLs. Implement parameter handling via Google Search Console or equivalent.
- Printer-friendly or PDF versions of pages must use noindex or canonical directives pointing to the original page.

9. Internal Linking and Information Architecture

9.1 Semantic Linking Rules

- Internal links must be placed within body content, not only in navigation menus or sidebars.
- Every internal link must use descriptive anchor text that identifies the destination page's primary entity. Generic anchor text such as 'click here', 'read more', or 'learn more' is non-compliant.
- Anchor text must match or closely reflect the H1 of the destination page.
- Links to the same destination within a single page must use consistent anchor text.

9.2 Hierarchical Structure Requirements

- The site must have a defined content hierarchy with a maximum of three levels: root, section, and page.
- The homepage must link to all primary section pages.
- Each section page must link to all child pages within that section.
- The breadcrumb navigation must accurately reflect the content hierarchy and must match the BreadcrumbList schema on every page.
- Hub pages (section-level pages) must exist for every defined content category.

9.3 Orphan Page Prevention

- No indexable page may have zero internal links pointing to it.
- Every page must be reachable from the homepage within three clicks.
- Pages that are removed from navigation must be assessed for redirect or consolidation. They must not remain indexable with no internal links pointing to them.
- A site crawl using a tool such as Screaming Frog, Sitebulb, or equivalent must be run at minimum quarterly to identify and resolve orphan pages.

9.4 Contextual Linking Rules

- Internal links must appear within contextually relevant body content. Links placed in footer link blocks are supplementary and do not satisfy the contextual linking requirement.
- A page must link to a related page only when the linked page is directly relevant to the surrounding content at the point of the link.
- Links inserted purely to distribute link equity without contextual relevance are non-compliant.
- The maximum number of internal links in the body content of a single page is 25. Navigation and footer links are excluded from this count.

10. Minimum Compliance Checklist

A website is considered compliant with the 2026 Minimum AI SEO / GEO Website Technical Standard only if every item in the following checklist is satisfied. Each item is binary. Partial compliance does not constitute compliance for that item.

10.1 Page Structure

#	Requirement	Pass Condition
1	Single H1 per page, matching schema name field, maximum 70 characters	Verified via HTML audit; zero pages with multiple H1 elements
2	AI summary paragraph present as first body paragraph, 40-80 words, declarative and factual	Word count and position verified; no promotional language present
3	Key Facts block present with 3-7 factual items, implemented as HTML list	Visible on-page; items match schema fields
4	Minimum two H2 sections per page, each followed by at least one paragraph	Verified via heading structure audit
5	FAQ section with minimum five Q&A pairs, each answer 40-120 words	Word count verified; FAQPage schema present and matching
6	Summary block present as final content section, 50-100 words, no new claims	Position and word count verified

10.2 Schema Markup

#	Requirement	Pass Condition
1	Organization schema on homepage with all required fields	Passes Google Rich Results Test with zero errors
2	WebSite schema on homepage	Passes validation; @id present
3	WebPage schema on every indexable page with name matching H1	Automated validation across all pages
4	BreadcrumbList schema on all non-homepage pages, matching visible breadcrumb	Schema text matches visible navigation text exactly
5	FAQPage schema on all pages with visible FAQ sections, matching visible text exactly	Character-level match between schema and visible content

6	Article schema on all editorial/blog content with author and publisher references	Passes Rich Results Test; dates present
7	LocalBusiness schema where applicable, with full address, phone, and hours	Matches visible page content exactly; passes validation
8	Service schema on all service pages	Present with provider reference
9	All JSON-LD in <head> or before </body>, not injected asynchronously	Confirmed via HTML source; not dependent on JS execution

10.3 AI Crawler Access

#	Requirement	Pass Condition
1	robots.txt present at root domain, HTTP 200 status	Verified via direct URL request
2	All required AI bots explicitly allowed: GPTBot, ChatGPT-User, Google-Extended, PerplexityBot, ClaudeBot, anthropic-ai, CCBot, Applebot-Extended	robots.txt audit; each bot name present with Allow: /
3	Sitemap declared in robots.txt, pointing to valid URL	Sitemap URL returns HTTP 200 with valid XML
4	llms.txt present at root domain, served as text/plain	URL accessible; Content-Type verified; all required sections present
5	Administrative and private paths explicitly disallowed	robots.txt review; admin paths blocked

10.4 Entity and Content Architecture

#	Requirement	Pass Condition
1	Single primary entity per page, named in H1 and schema	Entity audit across all pages
2	Entity names consistent across all pages, schema, navigation, and meta data	Cross-reference audit; zero discrepancies
3	Schema name and description fields match visible H1 and meta description exactly	Automated comparison check
4	Front-loaded answer within first 150 words of body content	Word position audit on all pages

5	Intent-based page separation: no page serves multiple intent types	Intent classification audit across site
---	--	--

10.5 Technical Performance

#	Requirement	Pass Condition
1	LCP under 2.5 seconds on mobile, measured at 75th percentile	PageSpeed Insights; field data where available
2	INP under 200 milliseconds on mobile, measured at 75th percentile	PageSpeed Insights; CrUX data
3	CLS under 0.1 on mobile, measured at 75th percentile	PageSpeed Insights; zero layout shift on load
4	SSR or SSG rendering; all indexable content in initial HTML response	View Source confirms content present without JS execution
5	All images in WebP or AVIF format with descriptive alt text, explicit width and height	Image audit; format and attribute verification
6	Canonical URL declared on every indexable page as absolute URL	HTML audit; no relative canonical URLs
7	XML sitemap present, valid, and updated within 24 hours of page changes	Sitemap validation; lastmod dates present
8	www/non-www 301 redirect in place; HTTP to HTTPS 301 redirect in place	Redirect chain verification
9	CDN in place serving from appropriate geographic PoP	DNS and header audit; CDN provider confirmed

10.6 Internal Linking

#	Requirement	Pass Condition
1	All internal links use descriptive anchor text matching destination H1	Link audit; zero generic anchor text in body content
2	No indexable page with zero internal links pointing to it (no orphan pages)	Site crawl; all pages reachable

		within three clicks from homepage
3	Breadcrumb navigation present on all non-homepage pages, matching BreadcrumbList schema	Visual and schema audit
4	Maximum 25 internal body-content links per page	Link count audit; navigation links excluded from count

End of Standard Document

2026 Minimum AI SEO / GEO Website Technical Standard

Version 1.0 | State of the Art Digital | stevecoulter.co.uk

This document is a developer implementation specification. Requirements are explicit, testable, and binary.