Website Auditor Scoring Guide

Complete Reference — 12 Dimensions, 200+ Factors, Composite Algorithm

Version 3.0 — May 2026 | digitalmarketingco.org/free-website-auditor | 100% Deterministic (Zero LLM Credits)

Table of Contents

How Scoring Works

Each dimension starts at 100 points. Points are deducted for missing elements, poor implementations, and security gaps. The composite score is a weighted average of all 12 dimensions. Every check is 100% deterministic — computed from HTML source, HTTP headers, and robots.txt without any external API calls or AI inference.

1. SEO (Search Engine Optimization) 15%

SEO is the foundation of organic search visibility. This dimension evaluates on-page elements that search engines use to understand, index, and rank your content. Missing or misconfigured SEO elements directly reduce your visibility to billions of daily search queries.

FACTOR	MAX DEDUCTION	WHY IT MATTERS
Title Tag (50–60 chars)	–20	The title tag is the single most important on-page SEO element. Google displays it as the clickable headline in search results. A missing or poorly sized title means lower click-through rates and weaker ranking signals. The 50–60 character range reflects Google's 2025–2026 display limits.
Meta Description (140–160 chars)	–15	Meta descriptions serve as your page's advertisement in search results. While not a direct ranking factor, a compelling description within 140–160 characters dramatically improves click-through rates, which indirectly boosts rankings through user engagement signals.
H1 Heading (single, keyword-rich)	–15	The H1 is your page's main topic declaration. Search engines treat it as the strongest content signal after the title. Multiple H1s dilute topic focus; missing H1s leave engines guessing. One H1 per page with your primary keyword is the universal best practice.
Heading Hierarchy (H1→H6)	–10	A logical heading structure (H1→H2→H3 without gaps) creates a content outline that search engines and screen readers follow. Skipped levels confuse crawlers about content relationships and reduce the semantic value of your page structure.
Canonical URL	–8	The canonical tag tells search engines which version of a page is the "official" copy. Without it, duplicate content across URLs splits your ranking authority. Incorrect canonicals can accidentally de-index your preferred page version.
Open Graph Tags	–8	Open Graph tags control how your page appears when shared on Facebook, LinkedIn, and other platforms. Complete OG tags (title, description, image, URL, type) ensure your content looks professional in social feeds, driving referral traffic back to your site.
Twitter/X Card Tags	–5	Twitter Card meta tags determine your page's preview format on Twitter/X. The "summary_large_image" type generates the most engagement. Without these tags, Twitter displays a generic, unappealing link preview that users are less likely to click.
Primary Keyword Placement	–9	Your primary keyword should appear in the H1, first paragraph, and URL slug. This triple-placement strategy sends the strongest possible relevance signal to search engines, confirming what topic your page is about from three complementary HTML locations.
Keyword Density (0.5–2.5%)	–8	Keyword density measures how frequently your target keyword appears relative to total word count. Below 0.5% signals weak topical relevance; above 3% triggers keyword stuffing penalties. The 0.5–2.5% sweet spot confirms relevance without over-optimization.
E-E-A-T Signals	–9	Google's E-E-A-T (Experience, Expertise, Authoritativeness, Trustworthiness) framework evaluates content credibility. We detect 10 indicators including about pages, author bios, testimonials, trust badges, and Person schema. Weak E-E-A-T signals can tank rankings in YMYL (Your Money Your Life) topics.
Readability (Flesch Score)	–4	The Flesch Reading Ease score measures how accessible your writing is. A score of 60+ means a general audience can understand it. Difficult content (below 40) reduces

		dwell time and engagement, signals that Google increasingly uses to evaluate content quality.
Internal Link Density	**–5**	Internal links distribute PageRank authority across your site and help crawlers discover pages. The optimal density is 2–5 internal links per 1,000 words. Too few starves pages of authority; too many dilutes each link's value.
Dofollow Outbound Links (2–5)	**–5**	Linking to authoritative external sources with dofollow signals topical relevance and editorial confidence to search engines. Zero outbound dofollow links looks unnatural; more than 10 leaks excessive PageRank. The 2–5 range is optimal.
Passive Voice Ratio	**–4**	Active voice writing is clearer, more engaging, and easier for AI/voice assistants to extract. A passive voice ratio above 30% indicates dense, academic-style writing that reduces user engagement and AI comprehension. Under 15% is ideal.
Table of Contents	**–3**	For articles over 1,500 words, a Table of Contents with anchor links improves dwell time, enables jump-link rich results in Google, and helps AI systems identify content sections for citation.
Zero-Click Optimization	**–2**	Featured snippets and zero-click results now dominate SERPs. TL;DR blocks and definition-style lead paragraphs ("X is...") are the content patterns Google extracts for position-zero featured snippets.
Entity Salience	**–2**	Entity salience measures how prominently named entities (people, brands, places, concepts) appear in your content. Strong entity mentions build topical authority signals that Google's Knowledge Graph uses to connect your page to real-world entities.
Binge-Worthy Signals	**–2**	Related content sections, "Read Next" links, and engagement patterns keep users on your site longer. Google measures dwell time and pages-per-session as quality signals; binge-worthy content directly improves both metrics.
Social Share Buttons	**–3**	Social sharing widgets encourage content distribution across platforms. While social signals are not direct ranking factors, the referral traffic and brand mentions they generate contribute to overall domain authority.
OG Sharing Completeness	**–3**	Beyond basic OG tags, this checks og:locale, og:image dimensions (1200×630px), article timestamps, and WebP format — the full set needed for pixel-perfect previews across every social platform.
URL Length (<75 chars)	**–5**	Short, descriptive URLs are easier for users to share and remember. Google truncates URLs over ~75 characters in search results, and excessively long URLs correlate with deeper, less authoritative pages.
External Links without rel	**–6**	External links without rel="nofollow noopener noreferrer" leak PageRank authority to external sites and create security vulnerabilities (tabnabbing). Proper rel attributes protect both your SEO equity and your users.
Redirect Chains	**–6**	Each redirect in a chain adds 100–500ms of latency and dilutes PageRank by roughly 15% per hop. A direct URL resolution (zero or one hops) delivers maximum speed and full link equity transfer.
SPA/JS Rendering	**–6**	Single-page applications (React, Vue, Angular) that rely on client-side rendering may produce thin initial HTML that search engines cannot index. Server-side rendering (SSR) ensures crawlers see your full content without executing JavaScript.
Anchor Text Diversity	**–4**	Repetitive anchor text on internal links ("click here", "read more") wastes an opportunity to signal topic relevance. Diverse, descriptive anchors help search engines understand what each linked page is about.

Lang, Favicon, Robots Meta, Deprecated Tags	**–5 each**	These foundational HTML elements ensure proper language targeting, brand recognition in browser tabs, crawler directives, and modern standards compliance. Each missing element signals a neglected technical foundation.

2. Performance 14%

Page speed directly impacts user experience, conversion rates, and search rankings. Google's Core Web Vitals (LCP, CLS, INP) are official ranking factors. This dimension evaluates server response, asset optimization, and resource loading strategies.

FACTOR	MAX DEDUCTION	WHY IT MATTERS
Server Response Time	–20	Time to First Byte (TTFB) is the foundation of all page speed. A response over 2 seconds indicates server-side bottlenecks — poor hosting, missing caching, no CDN, or unoptimized database queries. Under 500ms is excellent; under 1 second is acceptable.
Estimated LCP (Largest Contentful Paint)	–10	LCP measures when the largest visible element (hero image, heading) finishes rendering. Google's threshold: under 2.5 seconds is "good," over 4 seconds is "poor." We estimate LCP from hero image size, render-blocking resources, and font loading strategy. Poor LCP is the #1 Core Web Vitals failure.
Estimated CLS (Cumulative Layout Shift)	–8	CLS measures unexpected visual movement during loading — images without dimensions, injected ads, font swaps. Google's threshold: under 0.1 is "good," over 0.25 is "poor." We estimate CLS from images missing width/height, dynamic content injection patterns, and web font loading strategies.
Estimated INP (Interaction to Next Paint)	–8	INP measures responsiveness — how quickly the page reacts to user input (clicks, taps, key presses). Google's threshold: under 200ms is "good," over 500ms is "poor." We estimate INP from JavaScript bundle size, event handler density, and main-thread blocking scripts.
HTML Document Size	–15	Excessively large HTML documents slow parsing and first paint. Modern SSR frameworks naturally produce larger HTML, but documents over 500KB (or 1.5MB for framework apps) indicate bloated markup, excessive inline data, or missing code splitting.
Render-Blocking Scripts	–12	Scripts in the <head> without async/defer block HTML parsing and delay first paint. Every render-blocking script adds to the critical rendering path. Converting to async/defer or deferring non-critical scripts is one of the highest-impact performance wins.
Render-Blocking CSS	–5	External CSS files block rendering until fully downloaded and parsed. Inlining critical CSS and loading non-critical stylesheets asynchronously eliminates the CSS bottleneck in the critical rendering path.
Image Optimization (lazy, WebP/AVIF, srcset)	–19	Images are typically 50–70% of page weight. Lazy loading defers off-screen images, WebP/AVIF formats reduce file size by 30–50%, and srcset serves appropriately sized images per device. Together these three techniques can cut page weight in half.
Image Dimensions (width/height)	–5	Images without explicit width and height cause layout shifts as they load, directly worsening your CLS score. Setting dimensions lets the browser reserve space before the image downloads.
Font Loading (preload, display, variable fonts)	–9	Web fonts cause Flash of Invisible Text (FOIT) or Flash of Unstyled Text (FOUT). Preloading primary fonts, using font-display:swap, limiting font families to 2–3, and using variable fonts minimizes text rendering delays.
CSS Bundle Size	–12	Large CSS payloads (>200KB) increase parsing time and delay first paint. PurgeCSS, tree-shaking, and critical CSS extraction reduce unused styles. Over 500KB indicates serious CSS bloat.

JavaScript Bundle Size	**–10**	JavaScript is the most expensive resource — every KB must be parsed, compiled, and executed. Code splitting, tree shaking, and dynamic imports keep the initial JS payload manageable. Over 1MB triggers penalties; over 2MB for framework apps.
Resource Hints (preconnect, prefetch, preload)	**–7**	Resource hints tell the browser to start DNS lookups, TCP connections, and resource downloads earlier. Preconnect for critical third-party domains, preload for hero images/fonts, and prefetch for likely next navigations measurably reduce perceived load time.
HTTP Version (HTTP/2 or HTTP/3)	**–4**	HTTP/2 enables multiplexing (parallel resource loading over one connection) and header compression. HTTP/3 (QUIC) adds connection migration and zero round-trip handshakes. Sites still on HTTP/1.1 suffer from head-of-line blocking and sequential loading.
CDN Detection	**–5**	A Content Delivery Network caches your content at edge locations worldwide, reducing latency for distant users from seconds to milliseconds. Without a CDN, every request travels to your origin server, adding unnecessary round-trip time.
Compression (Brotli/gzip)	**–6**	Response compression reduces transfer sizes by 60–80%. Brotli is ~15% more efficient than gzip for text resources. Missing compression means users download raw, uncompressed HTML, CSS, and JavaScript.
Third-Party Scripts	**–6**	Each third-party script (analytics, ads, chat widgets) adds DNS lookups, connections, and main-thread execution. Over 10 third-party scripts from different domains can add 2–5 seconds to page load. Auditing and deferring non-critical trackers is essential.
Cache-Control Headers	**–5**	Cache-Control headers with proper max-age values let browsers store resources locally, eliminating redundant downloads on repeat visits. Missing cache headers force the browser to re-download everything every time.
Analytics Setup Quality	**–6**	Having no analytics means you cannot measure SEO or marketing performance. We evaluate GA4 presence, GTM configuration, and event tracking sophistication. Comprehensive analytics (GA4 + GTM + custom events) is the baseline for data-driven optimization.
Long Task Script Detection	**–6**	Scripts that block the main thread for over 50ms cause visible jank and poor INP scores. We detect patterns indicative of long tasks (large inline scripts, synchronous data processing) that should be moved to web workers or broken into smaller async chunks.

3. Security 10%

Security protects your users' data and your site's reputation. Google uses HTTPS as a ranking signal, and browsers actively warn users about insecure sites. This dimension evaluates encryption, security headers, and vulnerability indicators.

FACTOR	MAX DEDUCTION	WHY IT MATTERS
HTTPS Encryption	–25	HTTPS encrypts all data between user and server. Without it, browsers display "Not Secure" warnings, login credentials are transmitted in plaintext, and Google demotes your rankings. HTTPS is a non-negotiable baseline in 2026.
Mixed Content	–10	Loading HTTP resources (images, scripts, stylesheets) on an HTTPS page breaks the chain of trust. Browsers may block these resources or show security warnings, degrading both user experience and your SSL certificate's effectiveness.
Strict-Transport-Security (HSTS)	–13	HSTS forces browsers to always use HTTPS, preventing protocol downgrade attacks. Without HSTS, an attacker can intercept the initial HTTP request before the redirect to HTTPS. The preload directive adds your domain to browsers' built-in HTTPS-only list.
Content-Security-Policy (CSP)	–10	CSP is the primary defense against cross-site scripting (XSS) attacks. It defines which sources can load scripts, styles, and other resources. A missing CSP leaves your site vulnerable to injected malicious code that can steal user data.
CSP Strictness Score	–6	Not all CSPs are equal. We score strictness (0–100) based on directive count, use of nonces vs. unsafe-inline, default-src restrictiveness, and frame-ancestors. A weak CSP with unsafe-inline everywhere provides little real protection.
Permissions-Policy Depth	–7	Permissions-Policy restricts which browser APIs (camera, microphone, geolocation, payment) your page can access. Without it, malicious injected scripts could silently access sensitive device features. We measure both header presence and the number of restricted features.
X-Content-Type-Options	–5	The nosniff directive prevents browsers from MIME-type sniffing, which could execute a text file as JavaScript. This single header stops an entire class of content-type confusion attacks.
X-Frame-Options	–5	X-Frame-Options prevents your page from being embedded in iframes on malicious sites (clickjacking). Without it, attackers can overlay invisible iframes on top of legitimate UI to trick users into clicking hidden actions.
Referrer-Policy	–5	Controls how much URL information is sent to external sites when users click links. Without it, your full page URLs (which may contain sensitive query parameters) leak to every external site you link to.
Cross-Origin Policies (COOP/CORP/COEP)	–8	These three headers isolate your page's browsing context, preventing cross-origin attacks like Spectre side-channel data theft. They enable powerful features like SharedArrayBuffer while protecting user data from cross-origin reads.
Outdated Library Detection	–8	We detect jQuery <3.5, Angular <1.8, Bootstrap <5, and other outdated libraries with known CVE vulnerabilities. Running outdated libraries is one of the most common causes of website compromises.
Cookie Security (Secure, HttpOnly, SameSite)	–8	Cookies without Secure, HttpOnly, and SameSite attributes are vulnerable to interception (non-HTTPS), XSS theft (accessible to JavaScript), and CSRF

		attacks (sent on cross-site requests). All three attributes should be set on every session cookie.
Insecure Form Actions	**−8**	Forms that submit to HTTP endpoints transmit user data (passwords, credit cards, personal info) in plaintext. This is a critical security vulnerability that violates PCI-DSS and GDPR data protection requirements.
Subresource Integrity (SRI)	**−3**	SRI hashes verify that external scripts/stylesheets haven't been tampered with. If a CDN or third-party is compromised, SRI prevents the modified file from executing on your site — a critical supply-chain defense.
External Link Security	**−5**	External links with target="_blank" but missing rel="noopener noreferrer" enable tabnabbing attacks where the linked page can redirect your tab to a phishing site. This is a common security oversight with real attack potential.
Privacy Policy & Cookie Consent	**−6**	GDPR, CCPA, and ePrivacy regulations require visible privacy policies and cookie consent mechanisms. Missing these signals non-compliance that can result in fines up to 4% of annual revenue under GDPR.
Rate Limiting	**−2**	Rate limiting headers (X-RateLimit-Limit, Retry-After) indicate protection against brute-force login attempts, credential stuffing, and DDoS attacks on form endpoints. Sites with forms but no rate limiting are vulnerable to automated abuse.
Server/Technology Exposure	**−6**	Server and X-Powered-By headers that expose your technology stack (Apache version, PHP version) give attackers a roadmap of known vulnerabilities to exploit. CDN-proxied headers are an acceptable exception.

4. AI Readiness 10%

AI Readiness measures how well your content can be discovered, parsed, and cited by AI systems like ChatGPT, Claude, Perplexity, and Google's AI Overviews. In 2026, AI-mediated search is the fastest-growing traffic channel.

FACTOR	MAX DEDUCTION	WHY IT MATTERS
Structured Data (Schema.org)	−15	Structured data gives AI models explicit, machine-readable context about your content's topic, entities, and relationships. Without schema, AI systems must infer meaning from unstructured HTML — a lossy, unreliable process.
AI Crawler Governance (meta + robots.txt)	−20	We check whether AI crawlers (GPTBot, ClaudeBot, PerplexityBot, Google-Extended, Amazonbot, Meta-ExternalAgent, Applebot, Bytespider, Cohere, YouBot, PhindBot, GeminiBot) are allowed or blocked in both meta tags and robots.txt. Blocking AI crawlers prevents your content from appearing in AI-generated answers.
llms.txt File & Quality	−3	llms.txt is an emerging standard (like robots.txt for AI) that tells language models how to access and interpret your content. Quality levels: basic (exists), detailed (structured instructions), comprehensive (with citations and content map).
RAG-Friendliness Score	−5	RAG (Retrieval Augmented Generation) systems chunk your content into sections for vector embedding. Pages with id attributes on headings and section elements allow precise chunk targeting. A high RAG score means AI can cite specific sections, not just the whole page.
Content Provenance Signals	−3	AI models evaluate content credibility through provenance — attributions like "according to [source]" and "research by [institution]" make your claims verifiable and citation-worthy. Content without provenance is treated as unverified opinion.
Semantic HTML Elements	−12	Semantic elements (article, section, aside, figure, nav, main) provide structural meaning that AI can parse without heuristics. A page built with generic divs forces AI to guess your content hierarchy; semantic HTML makes it explicit.
FAQ/Q&A Content Patterns	−8	AI models are trained heavily on question-answer pairs. FAQ sections are the most directly extractable content format for AI citation. Question-headed content with direct answers mirrors the exact query pattern AI users ask.
Freshness Signals (dates, timestamps)	−5	AI models heavily weight content recency. dateModified, datePublished metadata, and <time> elements signal when content was last updated. Stale content without timestamps ranks lower in AI citation selection.
Social Links & Entity Signals	−5	Links to social profiles help AI systems confirm your brand's entity identity in the Knowledge Graph. Without them, AI cannot cross-reference your website with your social presence, weakening entity disambiguation.
HTML-to-Text Ratio	−5	Pages with very low text-to-HTML ratios (under 10%) are mostly markup and scripts with little substantive content for AI to extract. A healthy 20–70% ratio indicates content-rich pages that AI crawlers can efficiently process.
JS Content Dependency	−5	If most of your content requires JavaScript execution to render, AI crawlers (which generally do not execute JS) will see empty pages. Server-side rendering ensures AI models access your full content without running your application code.

5. AIO (AI Optimization) 9%

AIO measures content quality and conversion signals that make your page valuable for AI model training and citation. It combines content depth, entity markup, E-E-A-T schema signals, conversion rate optimization, and citation readiness.

FACTOR	MAX DEDUCTION	WHY IT MATTERS
Content Depth (word count)	–15	AI models value comprehensive, authoritative content. Thin pages (under 300 words) are rarely cited. Pages over 1,500 words with thorough topic coverage are 3x more likely to appear in AI-generated responses than thin content.
Structured Q&A Content	–10	FAQ and Q&A sections are the highest-value content for AI citation because they directly match the question-answer format of user queries. More FAQ items means more potential citation surfaces.
Entity Markup Strength	–8	Strong, em, and mark tags highlight key entities that AI models use for knowledge extraction. Without explicit entity emphasis, AI must use NLP to identify key terms — a process that can miss or misidentify important concepts.
E-E-A-T Schema Signals	–8	Person, Article, Review, and Service schemas provide machine-readable E-E-A-T signals. Two or more types demonstrate that your page has identifiable authors, structured content, user feedback, and defined services.
CTA Elements & Above-Fold CTA	–8	Clear call-to-action elements guide visitors toward conversion. Three or more CTAs with action-oriented text, plus a prominent CTA above the fold, indicate an optimized conversion funnel. Zero CTAs means visitors have no clear next step.
Trust Signals & Social Proof	–7	Testimonials, trust badges, guarantees, client logos, and social proof elements build conversion confidence. Without them, visitors must take your claims on faith — resulting in lower conversion rates and higher bounce rates.
Citation Readiness (outbound refs)	–5	Pages with 5+ external citations demonstrate research depth. AI models are more likely to cite content that itself cites authoritative sources — it signals editorial rigor and factual grounding.
Statistics/Data Point Density	–4	Specific numbers, percentages, and data points are highly extractable by AI and memorable for users. Content with zero data points reads as opinion; content with concrete statistics reads as authoritative analysis.
Author/Expertise Signals	–6	Visible author bylines with bios and credentials signal content accountability. AI models evaluate authorship as part of E-E-A-T scoring. Anonymous content is treated as less credible than attributed expertise.
Speakable Schema	–3	Speakable schema tells voice assistants (Google Assistant, Alexa, Siri) which parts of your page are suitable for text-to-speech. Without it, voice assistants cannot confidently read your content aloud in voice search results.
Value Proposition in Headings	–3	Your H1 or hero section should clearly communicate what you offer, for whom, and why it matters. A missing value proposition means visitors must hunt for your relevance — most won't bother and will bounce.

6. Accessibility 8%

Accessibility ensures your website is usable by people with disabilities — visual, auditory, motor, and cognitive. Beyond ethical obligation, accessibility improves SEO (semantic markup), reduces legal risk (ADA/EAA compliance), and expands your audience by ~15% of the global population.

FACTOR	MAX DEDUCTION	WHY IT MATTERS
Language Declaration (lang)	−10	The lang attribute on <html> tells screen readers which language to use for pronunciation. Without it, screen readers may mispronounce every word on the page, making content incomprehensible to blind users.
Skip Navigation Link	−8	A "Skip to main content" link as the first focusable element allows keyboard users to bypass navigation menus. Without it, users who cannot use a mouse must tab through every menu item on every page before reaching content.
Semantic Landmarks (main, nav, footer)	−12	Screen readers use landmark regions to let users jump between page sections. Without main, nav, and footer landmarks, blind users must read the entire page linearly — like reading a book without a table of contents.
Image Alt Text	−15	Images without alt text are invisible to screen reader users and provide no SEO value. Alt text should describe the image's content and purpose. Use empty alt="" only for purely decorative images.
Form Labels	−8	Form inputs without associated <label> elements are unnamed to screen readers. A blind user encountering an unlabeled text field has no idea what information to enter, making forms completely unusable.
Focus Styles	−8	Focus indicators (outlines on buttons, links when tabbed to) are the only way keyboard users can see where they are on the page. Removing :focus-visible styles renders your site unnavigable for the 2% of users who rely on keyboard navigation.
ARIA Implementation	−5	ARIA attributes provide additional accessibility information for interactive elements that native HTML cannot express. Zero ARIA attributes on a site with interactive components indicates missing screen reader support for custom widgets.
Button Labels	−8	Buttons without text content or aria-label are announced as "button" with no context by screen readers. Users cannot know what an unlabeled button does. Icon-only buttons must have aria-label attributes.
Color Contrast	−5	WCAG requires 4.5:1 contrast ratio for normal text and 3:1 for large text. Low contrast makes text difficult or impossible to read for users with low vision, color blindness, or in bright lighting conditions.
Video Captions	−8	Videos without captions exclude deaf and hard-of-hearing users, people in sound-sensitive environments, and non-native speakers. Captions also improve SEO by providing indexable text content for video.
CAPTCHA Accessibility	−5	Visual CAPTCHAs without audio alternatives completely block blind users from completing forms, registrations, or purchases. Accessible alternatives include audio CAPTCHAs or invisible challenge-response systems.
Duplicate IDs	−6	Duplicate element IDs break ARIA references (aria-describedby, aria-labelledby), form label associations, and fragment navigation. Screen readers may announce the wrong content or skip elements entirely.

HTML Validity / Nesting Errors	**–6**	Invalid HTML nesting (e.g., <p> inside <button>) causes assistive technology parsing failures. Screen readers rely on valid DOM structure to build their accessibility tree; errors create gaps in that tree.
Tables with Headers, Iframes with Titles	**–5 each**	Data tables without <th> headers are unreadable by screen readers — users cannot determine which column or row data belongs to. Similarly, iframes without title attributes provide no context about embedded content.

7. GEO (Generative Engine Optimization) 8%

GEO focuses on content formatting that AI-powered search engines (Google AI Overviews, Perplexity, ChatGPT Search) are most likely to extract and cite. It rewards structured, data-rich, citable content with clear question-answer patterns.

FACTOR	MAX DEDUCTION	WHY IT MATTERS
Question-Answer Patterns	–10	Generative engines extract content by matching user questions to on-page answers. Question-based H2/H3 headings followed by concise answer paragraphs are the #1 content pattern cited in AI Overviews and Perplexity answers.
Data/Comparison Tables	–8	Generative AI engines frequently cite tabular data for comparison queries ("X vs Y", "best options for..."). Tables provide structured, easily extractable information that AI can directly embed in responses.
Content Depth (>1000 words)	–10	Generative engines cite comprehensive, authoritative sources. Thin content rarely appears in AI-generated responses. Research shows pages over 1,000 words with unique insights are 5x more likely to be cited by AI search engines.
List-Based Content	–5	Numbered and bulleted lists are the second most-extracted content format (after Q&A) by generative engines. Step-by-step instructions, feature lists, and comparison points are all best presented as scannable lists.
Hreflang (with self-ref & x-default)	–18	Hreflang tags tell generative engines which language/region each page version targets. Missing self-referencing tags and x-default fallbacks cause AI to serve wrong-language content. Bidirectional validation ensures complete geo-targeting.
Statistics/Data Density	–4	Generative engines prefer content with verifiable data points. Pages with specific numbers, percentages, and cited statistics are treated as more authoritative than opinion-based content without supporting data.
Original Research Signals	–3	Charts, figures, and data tables that suggest original research signal first-party data. Generative engines cite original research over derivative content because it offers unique information not available elsewhere.
Content Freshness	–5	Stale content (over 1 year since last update) is progressively deprioritized by generative engines that favor recency. Fresh content with visible update dates signals active maintenance and current accuracy.
Unique Insight Patterns	–3	Phrases like "our research found" or "based on our analysis" signal first-hand expertise. Generative engines recognize these patterns as original insights worth citing, versus content that merely aggregates others' findings.

8. Mobile & PWA 8%

Mobile-first indexing means Google primarily uses the mobile version of your content for ranking. Over 60% of web traffic is mobile. This dimension evaluates responsive design, Progressive Web App readiness, and mobile-specific performance and usability.

FACTOR	MAX DEDUCTION	WHY IT MATTERS
Viewport Meta Tag	–25	The viewport meta tag is the foundation of mobile rendering. Without width=device-width and initial-scale=1, mobile browsers render your page at desktop width (typically 980px) and scale down, making text unreadably small and buttons untappable.
Viewport Zoom Restriction	–8	maximum-scale=1 and user-scalable=no prevent users from zooming in. This violates WCAG 1.4.4 and hurts users with low vision who need to zoom to read text. Google specifically penalizes zoom-restricted viewports.
Touch Target Size	–8	Interactive elements smaller than 44x44px (Google recommends 48x48px) are difficult to tap on mobile. Small touch targets cause accidental clicks, frustration, and increased bounce rates. We detect elements with undersized tap areas.
Fixed-Width Elements	–8	Elements with fixed pixel widths over 800px cause horizontal scrolling on mobile devices. This forces users to scroll sideways to read content — a severe usability failure that Google's mobile-first algorithm penalizes.
Web App Manifest	–8	The web app manifest enables PWA features: add-to-homescreen, app-like appearance, and splash screens. It's required for the PWA install prompt and defines your app's name, icons, theme color, and display mode.
PWA Install Readiness	–3	The browser's PWA install prompt requires three elements: a valid manifest, a registered service worker, and HTTPS. A manifest without a service worker means users cannot install your site as an app, missing a significant engagement opportunity.
Manifest Icon Coverage	–3	A manifest with fewer than 4 icon sizes means your app icon will be scaled (and distorted) on many devices. Recommended sizes: 192x192, 384x384, 512x512, plus a maskable icon for Android adaptive icons.
Service Worker	–5	Service workers enable offline capability, background sync, and push notifications. They also allow caching strategies that make repeat visits nearly instant. Missing a service worker means your site stops working completely when users go offline.
Responsive Images (srcset)	–8	Without srcset, mobile devices download the same full-resolution desktop images, wasting bandwidth and slowing load times. Srcset serves appropriately sized images per screen, reducing data usage by 40–70% on mobile.
Popup/Interstitial Detection	–5	Google's Page Experience update specifically penalizes intrusive interstitials (full-screen popups) on mobile. They block content access, increase bounce rates, and can result in ranking demotion for the entire page.
Safe-Area Insets	—	Modern phones with notches, dynamic islands, and rounded corners clip content that doesn't use env(safe-area-inset-*) CSS. This is a recommendation rather than a deduction — but failing to handle it means content hidden behind hardware elements.

Touch Feedback	**–2**	CSS :active states and touch-action properties provide visual feedback when users tap interactive elements. Without touch feedback, users cannot confirm their taps registered, leading to double-taps and frustrated interactions.
Page Weight (Sustainability Proxy)	**–5**	Pages over 3MB are heavy for mobile data plans — especially in regions with metered or slow connections. Under 1MB is considered lightweight. Page weight is also a sustainability indicator; lighter pages consume less energy in transmission and rendering.
i18n Signals	**—**	Internationalization signals (hreflang, lang, dir, locale meta) indicate a site prepared for global audiences. Three or more i18n signals earn positive recognition; RTL support is noted separately for right-to-left language audiences.

9. AEO (Answer Engine Optimization) 6%

AEO optimizes for direct answer results — featured snippets, People Also Ask boxes, voice search results, and knowledge panels. These position-zero results capture clicks before traditional organic listings.

FACTOR	MAX DEDUCTION	WHY IT MATTERS
FAQPage Schema	–15	FAQPage schema is the most direct path to expandable FAQ rich results in Google. Without it, even well-structured FAQ content won't earn the enhanced SERP treatment that dramatically increases visibility and click-through rates.
Featured Snippet Readiness	–10	Position-zero featured snippets capture ~35% of all clicks. They require question-based headings (H2/H3) followed by concise 40–60 word answer paragraphs. Content not formatted this way is invisible to the snippet algorithm.
Speakable Schema	–8	Voice assistants use Speakable schema to identify which page sections are suitable for text-to-speech. As voice search grows to ~50% of queries, Speakable markup ensures your content is selected for voice answers.
HowTo Schema	–5	HowTo schema enables rich step-by-step results in Google with expandable instructions, images, and time estimates. It's the highest-impact schema for instructional content — "how to" queries are among the most common search patterns.
Concise Answer Blocks	–4	Google's snippet algorithm specifically targets paragraphs between 10–50 words that directly answer questions. Pages with 5+ concise answer blocks have multiple opportunities to capture featured snippets across different queries.
Direct Answer Patterns	–4	Definitional statements ("X is defined as...", "X refers to...") are the #1 content pattern extracted for knowledge panel answers and zero-click results. They provide the exact sentence structure answer engines are trained to find.
Breadcrumb Schema	–5	BreadcrumbList schema enables breadcrumb rich results that show your site hierarchy in SERPs. This improves click-through rates by giving users context about where the page sits in your site structure.
Paragraph Length	–5	Voice assistants and answer engines prefer paragraphs under 30 words. Long, dense paragraphs (over 40 words) are too long for voice reading and too complex for snippet extraction. Short, direct paragraphs optimize for both channels.
Video Content + VideoObject Schema	–3	Video results appear in ~25% of featured snippet positions. VideoObject schema enables video rich results with thumbnails, duration, and upload dates directly in SERPs. Video content also captures image-pack and video-carousel placements.

10. Schema.org 6%

Schema.org provides the vocabulary that search engines and AI use to understand your content explicitly. This dimension evaluates schema implementation depth, type diversity, property completeness, and rich results eligibility.

FACTOR	MAX DEDUCTION	WHY IT MATTERS
Schema Presence	−30	Zero structured data is the single largest Schema.org penalty. Without any schema, search engines have no machine-readable understanding of your content type, business, or entity relationships. JSON-LD is the recommended format.
Tier 1 Schema Types	−8	Tier 1 types (Organization, WebSite, Article, Product, FAQPage, BreadcrumbList, LocalBusiness, BlogPosting, NewsArticle) deliver the highest SEO impact and richest search features. Three or more top-tier types indicate mature implementation.
Tier 2 Schema Types	−3	Tier 2 types (Person, HowTo, Service, Event, Review, AggregateRating, VideoObject, Offer) extend rich result eligibility. They add star ratings, event details, product offers, and authorship signals to your search listings.
Property Completeness	−8	Schema types with few properties provide minimal value. An Organization schema with only a name is far less useful than one with address, phone, logo, sameAs, and foundingDate. We measure average property completeness across all types.
JSON-LD Format	−3	Google recommends JSON-LD over Microdata and RDFa. JSON-LD is easier to implement, maintain, and debug — it sits in a script tag separate from your HTML markup, avoiding the complexity of inline annotations.
Schema Validity	−8	Invalid JSON-LD blocks (parse errors, missing required fields) are silently ignored by search engines. Your schema provides zero value if it contains syntax errors. We detect parse failures and sparse implementations.
sameAs Entity Linking	−4	sameAs links (to Wikipedia, Wikidata, social profiles) help search engines disambiguate your brand entity from others with similar names. Without sameAs, your Knowledge Graph entry may be confused with unrelated entities.
SearchAction Schema	−2	SearchAction on your WebSite type enables the sitelinks search box in Google — a prominent search field directly in your branded SERP listing that lets users search your site without visiting it first.
BuyAction / SubscribeAction	—	Action schemas (BuyAction, OrderAction, SubscribeAction, JoinAction) signal conversion intent to search engines, potentially enabling direct purchase or signup rich results in future Google Shopping and AI-powered interfaces.
Rich Results Eligibility	−5	We evaluate which rich result types your schema qualifies for: FAQ dropdowns, breadcrumbs, star ratings, how-to steps, product cards, and more. Zero eligibility means your SERP listing will always be a plain blue link.
Nested Schema Depth	−2	Nested schemas (author within Article, offers within Product) define entity relationships that flat schemas cannot. Deeper nesting enables richer Knowledge Graph connections and more specific rich result features.

11. Local SEO 4%

Local SEO determines your visibility in location-based searches ("near me" queries, Google Maps, local packs). It evaluates business information consistency, local schema markup, and geographic signals.

FACTOR	MAX DEDUCTION	WHY IT MATTERS
NAP (Name, Address, Phone)	**–15**	NAP consistency is the #1 local ranking factor. Your business name, address, and phone number must be prominently displayed and match your Google Business Profile exactly. Inconsistencies across the web confuse Google about your location.
LocalBusiness Schema	**–15**	LocalBusiness JSON-LD provides Google with structured business data: name, address, phone, hours, geo coordinates, and service area. Without it, Google must scrape your page to infer business details — a lossy process that often misses information.
Phone Number	**–8**	A clickable phone number (tel: link) is essential for mobile users and signals local business legitimacy. Missing phone numbers reduce trust and make it impossible for nearby customers to call directly from search results.
Google Maps Embed	**–5**	An embedded Google Map provides visual location confirmation for both users and search engines. It signals a physical business presence and helps Google associate your website with a specific geographic point.
Address Microdata	**–5**	PostalAddress schema makes your address machine-readable. Without it, search engines parse your address from plain text, which can misidentify suite numbers, city names, or multi-line formatting.
Geo Meta Tags	**–5**	geo.region, geo.placename, and geo.position meta tags explicitly declare your business's geographic target. They supplement schema data with header-level location signals.
Opening Hours	**–5**	Business hours are a top local ranking factor and appear prominently in local pack results. LocalBusiness schema without openingHoursSpecification misses one of the most-searched local business attributes.
Review/Rating Schema	**–5**	AggregateRating schema displays star ratings in search results, dramatically improving click-through rates. Reviews are the #2 local ranking factor (after NAP consistency). Missing review schema means no stars in SERPs.
Service Area Schema	**–3**	The areaServed property in LocalBusiness schema defines your service radius. Without it, Google cannot determine which geographic queries your business is relevant for beyond your physical address location.

12. WCAG (Web Content Accessibility Guidelines) 2%

WCAG extends basic accessibility into deep compliance with the W3C's Web Content Accessibility Guidelines 2.2. It evaluates motion sensitivity, contrast, cognitive load, keyboard accessibility, and the newest 2.2 criteria.

FACTOR	MAX DEDUCTION	WHY IT MATTERS
Reduced Motion Support	–10	Users with vestibular disorders experience nausea, dizziness, and disorientation from animations. @media (prefers-reduced-motion: reduce) respects their OS setting by disabling motion. This is a WCAG 2.3.3 requirement.
Reduced Motion Enforcement	—	Detecting the media query is not enough — animations must actually be disabled (animation: none, transition: none). We check for enforcement to ensure the preference is honored, not just declared.
High Contrast Mode	–8	Users with low vision often enable forced-colors or high-contrast mode in their OS. Websites that don't handle this mode may become unusable — invisible text, missing borders, lost icons. Custom @media (forced-colors: active) styles prevent this.
Link Distinguishability	–8	WCAG 1.4.1 requires links to be distinguishable from surrounding text by more than just color (for color-blind users). Underlines or other visual indicators ensure all users can identify clickable elements.
Text Resize (rem/em)	–8	WCAG 1.4.4 requires text to be resizable to 200% without loss of content. Pixel-based font sizes don't scale with user preferences; rem, em, and clamp() respect the user's base font size setting.
Focus Not Obscured (WCAG 2.4.11)	–5	Sticky headers, fixed footers, and floating elements can cover the keyboard focus indicator when users tab through content. WCAG 2.2's new 2.4.11 criterion requires that focused elements never be fully hidden by sticky/fixed overlays.
Accessible Authentication (WCAG 3.3.8)	–5	WCAG 2.2 requires login forms to not demand cognitive tasks (memorizing codes, solving puzzles) as the only auth method. Forms must support autocomplete, password managers, and copy-paste in password fields.
Dragging Alternatives (WCAG 2.5.7)	–5	Drag-and-drop interfaces are inaccessible to users with motor impairments. WCAG 2.2 requires click/tap alternatives (up/down buttons, select controls) for every drag operation.
ARIA Live Regions	–3	Dynamic content updates (form feedback, notifications, chat messages) are invisible to screen readers unless announced via aria-live regions (role="status" or role="alert"). Without them, blind users miss critical state changes.
Color Contrast Issues	–15	WCAG 1.4.3 requires 4.5:1 contrast for normal text, 3:1 for large text. We detect potential contrast issues from CSS analysis. Poor contrast affects an estimated 300 million color-blind users and anyone in bright sunlight.
Cognitive Load Score	–5	Auto-playing carousels, excessive animations, notification badges, and complex layouts increase cognitive burden, particularly affecting users with ADHD, autism, and cognitive disabilities. WCAG 2.2 specifically addresses cognitive accessibility.
Heading Level Skips	–4	Skipping heading levels (H1→H3, missing H2) breaks screen reader navigation. Users who navigate by headings land at unexpected positions, losing context about where they are in the content hierarchy (WCAG 1.3.1).

ARIA Role Validity	**−8**	Invalid ARIA roles (typos, deprecated values) are ignored by assistive technology. Worse, some invalid roles can override native semantics, making a button announced as a generic "group" or making a link unclickable by screen reader commands.
Autoplay Media	**−10**	Autoplay audio/video violates WCAG 1.4.2 and can be disorienting, especially for users with cognitive disabilities or screen reader users whose audio is drowned out. Media must not play automatically or must provide immediate pause controls.
Timing Adjustable	**−3**	WCAG 2.2.1 requires users to be able to extend, adjust, or disable time limits on interactive content. Countdown timers, auto-refreshing content, and session timeouts without extension options violate this criterion.

Composite Scoring Algorithm

Each of the 12 dimensions starts at 100 and loses points for each failing factor. Scores are clamped to 0–100. The composite score is a weighted average reflecting 2026 priorities.

```
// Each dimension: start at 100, subtract per-factor deductions, clamp to [0, 100]
dimensionScore = clamp(100 - sumOfDeductions, 0, 100)

// Composite Score — weighted average (weights sum to 1.00)
compositeScore = round(
  SEO × 0.15 +
  Performance × 0.14 +
  Security × 0.10 +
  AI_Readiness × 0.10 +
  AIO × 0.09 +
  Accessibility × 0.08 +
  GEO × 0.08 +
  Mobile × 0.08 +
  AEO × 0.06 +
  Schema.org × 0.06 +
  Local_SEO × 0.04 +
  WCAG × 0.02
)

// Final composite clamped to [0, 100]
finalScore = clamp(compositeScore, 0, 100)
```

Letter Grade Scale

Score Range	Grade
97–100	**A+**
93–96	**A**
90–92	**A–**
87–89	**B+**
83–86	**B**
80–82	**B–**
77–79	**C+**
73–76	**C**
70–72	**C–**
67–69	**D+**
63–66	**D**
60–62	**D–**
0–59	**F**

Impact / Effort Ratings

Every recommendation in the audit report includes an Impact/Effort rating to help you prioritize fixes:

RATING	MEANING
Impact: High	Fixing this will produce a large, measurable improvement in score, rankings, or user experience
Impact: Medium	Noticeable improvement; recommended but not urgent
Impact: Low	Minor improvement; address when convenient
Effort: High	Requires significant development time, architecture changes, or specialist skills
Effort: Medium	A few hours of work for a competent developer
Effort: Low	Quick fix — often a single line of HTML or a config change

Best Strategy: Start with **High Impact / Low Effort** fixes first (title tags, meta descriptions, missing headers). These deliver the biggest score improvements for the least work. Save **Low Impact / High Effort** items (full PWA implementation, CSP nonce migration) for later phases.

Methodology Note: All scores are computed deterministically from HTML source code, HTTP response headers, and robots.txt content. No external APIs, no AI inference, and no Lighthouse simulations are used. Core Web Vitals metrics (LCP, CLS, INP) are proxy estimates based on HTML structure analysis — for lab measurements, use Google PageSpeed Insights or Chrome DevTools.