GEO Technical Reference

Complete System Documentation, Scoring Formulas & Research Data

Version 1.0 | February 2026
Internal technical documentation
Based on 14 primary research sources

Contents

1. System Architecture

The GEO audit system consists of an orchestration layer and five specialized analysis modules that run in parallel for performance.

Component	Function	Dependencies
Orchestrator	Page crawling, data collection, score aggregation	All modules
Citability Module	Content block analysis, quotability scoring	Page data
Brand Module	Cross-platform presence scanning	WebSearch, WebFetch
Technical Module	robots.txt, llms.txt, headers, rendering	Page data
Content Module	E-E-A-T signal evaluation	Page data
Schema Module	JSON-LD detection, validation, gap analysis	Page data

Data Flow

1. Orchestrator fetches homepage, detects business type, crawls sitemap (max 50 pages)
2. Page-level data collected: URL, title, meta, headings, word count, schema, links, images
3. Five modules analyze data in parallel, each producing score (0-100) + findings
4. Orchestrator computes weighted composite score
5. Issues classified by severity (Critical/High/Medium/Low)
6. Report generated with scores, findings, and prioritized action plan

2. Scoring Formulas & Weights

Composite Score Formula

GEO_Score = (Citability x 0.25) + (Brand x 0.20) + (EEAT x 0.20) + (Technical x 0.15) + (Schema x 0.10) + (Platform x 0.10)

Weight Rationale

Dimension	Weight	Rationale
AI Citability	25%	Citation is the fundamental output. Content that cannot be cited fails at GEO regardless of other sig
Brand Authority	20%	Entity recognition determines whether AI trusts a source enough to cite. High-E-E-A-T at #6-10 bea
Content E-E-A-T	20%	Quality signals (author credentials, verifiable claims, depth) are the gatekeeper for citation selectior
Technical GEO	15%	Crawler access is a prerequisite. If AI cannot crawl, nothing else matters. But once accessible, dimi
Schema	10%	Schema contributes ~10% of Perplexity ranking. Critical but enables understanding rather than dire
Platform	10%	Off-site presence matters (48% of citations from community platforms) but is slower to influence.

Score Thresholds

Range	Label	Color Code	Action Required
90-100	Excellent	Green	Maintain and iterate
75-89	Good	Blue	Targeted improvements
60-74	Fair	Blue	Significant optimization needed
40-59	Poor	Yellow	Major restructuring required
0-39	Critical	Red	Fundamental rebuild needed

3. Audit Module Specifications

3.1 Citability Module

- Analyzes content blocks for AI quotability
- Scores passage self-containment (can it answer a question standalone?)
- Measures statistical density (fact-to-word ratio, target > 1:80)
- Evaluates answer block quality (direct answers in first 200-300 words)
- Checks heading hierarchy (H1>H2>H3 structure)
- Identifies high-value pages that could be reformatted for better citation
- Optimal citation passage length: 134-167 words

3.2 Brand Module

- Scans YouTube for brand presence and video content
- Checks Reddit for brand mentions (subreddit activity, comments)
- Searches Wikipedia/Wikidata for entity presence
- Evaluates LinkedIn company profile completeness
- Assesses third-party mention volume and sentiment
- Measures cross-platform NAP (Name, Address, Phone) consistency
- Evaluates review volume and velocity across platforms

3.3 Technical Module

- Analyzes robots.txt for AI crawler access (9 crawlers, 2 tiers)
- Checks for llms.txt presence, validity, and completeness
- Verifies server-side rendering for all public content
- Tests page load speed (<3 seconds target)
- Checks meta tags, canonical URLs, Open Graph, Twitter Cards
- Validates HTTP headers for security and caching
- Tests IndexNow protocol support

3.4 Content Module

- Evaluates Experience signals (case studies, real examples, photos)
- Measures Expertise indicators (author credentials, certifications, how-tos)
- Assesses Authoritativeness (awards, press, third-party mentions, reviews)
- Checks Trustworthiness (source citations, real data, transparent business info)
- Evaluates content freshness (publication and update dates)
- Measures content depth (word count, topic coverage, comprehensiveness)

• Checks author attribution and Person schema

3.5 Schema Module

• Detects all JSON-LD schema types present on each page

• Validates schema against schema.org specifications

• Identifies missing schema opportunities by business type

• Checks for essential types: Organization/LocalBusiness, FAQ, Article

• Evaluates schema completeness (required vs optional properties)

• Checks for Product schema with pricing and availability

• Validates Review and aggregateRating markup

4. Issue Severity Classification

Severity	Timeline	Examples
Critical	Fix immediately	All AI crawlers blocked; no indexable content; site errors on key pages; no structured data; bran
High	Fix within 1 week	Key crawlers blocked; no llms.txt; zero Q&A content; missing Org schema; no author attribution
Medium	Fix within 1 month	Partial crawler blocking; incomplete llms.txt; low citability scores; missing FAQ schema; thin aut
Low	When possible	Minor schema errors; some missing alt text; content freshness issues; missing OG tags; subopt

5. Business Type Detection

The audit automatically classifies websites to apply business-type-specific scoring adjustments and recommendations.

Type	Detection Signals	Key Schema	Extra Weight On
SaaS	Pricing page, sign up CTAs, app subdomain	SoftwareApplication, FAQ, HowTo	Feature comparison, docs, integrations
Local Business	Physical address, Maps embed, hours	LocalBusiness, GeoCoordinates, Hours	NAP consistency, GBP, local schema
E-commerce	Product listings, cart, prices	Product, AggregateRating, Offer	Product descriptions, buying guides
Publisher	Blog-heavy nav, article schema, authors	Article, NewsArticle, Person	Author credentials, original research
Agency	Case studies, portfolio, client logos	Organization, Service, Person	Thought leadership, case study depth

6. Citation Statistics Database

Compiled from 14 primary research sources. All statistics have been verified across at least two independent sources where possible.

Statistic	Value	Source	Confidence
Fact-to-word ratio > 1:80 citation boost	4.2x	Incremys	High
Statistics in content visibility boost	30-40%	Profound	High
Semantic completeness > 8.5/10 boost	4.2x	Wellows	High
H1>H2>H3 hierarchy citation boost	2.8x	Incremys	High
Strong E-E-A-T at rank #6-10 vs weak #1	2.3x	Wellows	High
Topic clusters citation rate boost	30%	Wellows	Medium
AI referral traffic growth (SMBs)	+123%	HubSpot	High
Schema in Perplexity ranking	~10%	Qwairy	Medium
Domain Authority correlation w/ AI	r=0.18	Wellows	High
ChatGPT sources per query	~8	Qwairy	High
Perplexity sources per query	~22	Qwairy	High
Wikipedia in ChatGPT top 10	47.9%	Profound	High
Reddit in Perplexity top 10	46.7%	Profound	High
.com domains in all citations	80.41%	Profound	High
Citations from owned sites	44%	Profound	High
Citations from community platforms	48%	Profound	High
ChatGPT searches with local intent	59%	HubSpot	Medium
Gen Z using AI as primary search	31%	HubSpot	Medium
GBP photos engagement boost	35%	HubSpot	Medium
Consumers using review-responsive biz	88%	HubSpot	Medium
Brands without GEO strategy	47%	Foundation Inc	Medium

7. Platform Citation Behavior (Detailed)

ChatGPT

Metric	Value
Sources per query	~8
Top cited domain	Wikipedia (7.8% of all citations)
Wikipedia in top 10	47.9%
Most cited page type	"Best X" blog lists (43.8%)
Content preference	Encyclopedic, factual, authoritative
Domain type distribution	.com (80.41%), .org (11.29%)

Perplexity

Metric	Value
Sources per query	~22 (2.76x more than ChatGPT)
Top cited domain	Reddit (6.6% of all citations)
Reddit in top 10	46.7%
YouTube in top 10	13.9%
Content preference	Community-driven, recency-focused
Key differentiator	Broader citation (22 vs 8 sources)

Google AI Overviews

Metric	Value
Citation mix	Balanced social-professional
Reddit presence	2.2% of citations
YouTube presence	1.9% (but 18.8% of top 10)
Quora presence	1.5%
LinkedIn presence	1.3%

Key factor	Existing ranking + E-E-A-T signals

8. Schema Implementation Guide

Complete JSON-LD schema implementation reference for GEO-optimized websites. All schema should be embedded as <script type="application/ld+json"> in the page head.

Schema Type	Required Properties	GEO Impact
LocalBusiness	name, address, telephone, openingHours, geo, url, sameAs	Entity recognition, NAP consistency
Organization	name, url, logo, sameAs, contactPoint	Brand entity identity
FAQPage	mainEntity[].name, mainEntity[].acceptedAnswer	Direct Q&A extraction by AI
Service	name, description, provider, areaServed	Service-specific AI citations
Product	name, description, offers.price, brand	Product comparison citations
Review	author, reviewBody, reviewRating	Trust and social proof signals
AggregateRating	ratingValue, reviewCount, bestRating	Summarized trust at a glance
Article	headline, author, datePublished, publisher	Content attribution and freshness
HowTo	name, step[].name, step[].text	Process/tutorial AI extraction
Person	name, jobTitle, worksFor, sameAs	Author authority and E-E-A-T
ContactPoint	telephone, email, contactType, areaServed	Contact accessibility signal
BreadcrumbList	itemListElement[].name, .item	Navigation hierarchy understanding

9. Local Business GEO Playbook

Local businesses have unique GEO opportunities. 59% of ChatGPT searches involve local intent, and local businesses can dominate geographic queries more easily than national competitors.

Priority Actions

Priority	Action	Impact	Timeline
#1	Claim/optimize Google Business Profile (all locations)	Highest single factor	Week 1
#2	Launch review campaign (target 50+ Google reviews)	Trust + velocity signal	Weeks 1-8
#3	Cross-platform NAP consistency (Houzz, Yelp, BBB, Angi)	Entity recognition	Week 2
#4	Add photos to GBP (35% more engagement)	Visual trust signal	Week 2
#5	Respond to all reviews (88% consumer preference)	Active business signal	Ongoing
#6	Create location-specific content pages	Geographic entity signals	Weeks 3-4
#7	YouTube channel with project walkthroughs	#2 cited source on AI Overviews	Month 2
#8	Reddit/community presence (local subreddits)	Perplexity citation source	Month 2
#9	Local digital PR (newspaper features, industry pubs)	Earned media trust	Months 2-6
#10	Author page with Person schema for owner/experts	E-E-A-T authority signal	Week 3

Timeline expectations: Initial traction in 4-8 weeks. Foundation work (schema, content) in 4-8 weeks. Authority building (cross-platform) in 3-6 months. Measurable citation improvements within 90 days.

10. Strategic Shifts: SEO to GEO

The transition from traditional SEO to GEO represents five fundamental shifts in how digital visibility is achieved:

SEO Era	GEO Era	Implication
Link equity	Citation authority	Third-party mentions outweigh backlinks
Keyword rankings	Answer inclusion	Success = being the cited source in AI responses
Single platform (Google)	Multi-engine optimization	Must optimize for ChatGPT, Google AI, Perplexity, Claude, Copilot
Content marketing	Digital PR as core lever	Earned media directly influences AI citation patterns
Static content	Content refresh cadence	Regular updates with data signal active authority

11. Investment Benchmarks

Category	Investment Range	Timeline
Foundation work (schema, content restructuring)	Varies by scope	4-8 weeks
Mid-market GEO program (annual)	$75k-$150k/year	Ongoing
Enterprise GEO program (annual)	$250k+/year	Ongoing
Authority building (cross-platform presence)	Varies	3-6 months
Measurable citation improvements	N/A	Within 90 days

47% of brands still lack a deliberate GEO strategy, creating a significant first-mover advantage for organizations that invest now.

12. Research Sources

All data in this reference is sourced from the following 14 primary sources, researched February 2026:

#	Source	Topic
1	SearchEngineLand	Mastering GEO 2026 (comprehensive guide)
2	Profound	AI Platform Citation Patterns (data study)
3	HubSpot	GEO for Small Business (SMB focus)
4	Firebrand Marketing	GEO Best Practices 2026
5	Go Fish Digital	GEO Strategies
6	SchemaApp	AI Search and Schema Markup
7	Hook Agency	Brand Authority in AI Search
8	Wellows	AI Overviews Ranking Factors (data study)
9	Averi	Track AI Citations (metrics guide)
10	Qwairy	Perplexity vs ChatGPT Citation Study
11	Foundation Inc	GEO Metrics
12	ALM Corp	Schema Markup 2026
13	Revved Digital	E-E-A-T and AI Search 2026
14	LinkBuildingHQ	llms.txt evaluation

This technical reference document is based on automated analysis, industry research, and verified data sources. Statistics should be revalidated periodically as the AI search landscape evolves rapidly.