

5W

T H E A I C O M M U N I C A T I O N S F I R M

R E S E A R C H R E P O R T / Q 1 2 0 2 6

The 5W Citation Source Audit

Where AI Gets Its Answers — and what it means
for communications strategy.

P U B L I S H E D
B Y
5W

R E S E A R C H
B A S E
9 datasets

C A D E N C E
Quarterly

C O N T E N T S

What's Inside

01	Executive Summary	03
02	The Core Insight	04
03	How This Was Built	05
04	Top Cited Domains	06
05	The Nine Findings	07
06	Platform Differences	12
07	Industry Patterns	13
08	Operator Playbook	14
09	References & Limitations	15
10	About 5W	16

The PR Tier Hierarchy No Longer Reflects How Influence Works

For decades, public relations operated on a stable hierarchy: Tier 1 media, then trade media, then blogs. That hierarchy no longer reflects how influence works.

When users ask ChatGPT, Claude, Perplexity, Gemini, or Google AI Overviews about a brand, a category, or an executive, those systems do not rely on the traditional PR tier structure. They pull from a fragmented, dynamic, and structurally different source ecosystem — where Wikipedia and Reddit dominate, LinkedIn and YouTube are rapidly rising, review platforms drive recommendations, and traditional Tier 1 media is underrepresented.

This synthesis report draws on nine independently published research datasets covering hundreds of millions of citations and prompts to offer a unified working model of AI citation behavior.

T H E F I V E C O R E F I N D I N G S

01 Wikipedia + Reddit = Structural Dominance. Together they account for over 25% of ChatGPT citations in the U.S. (Similarweb, Q1 2026) — more than any traditional media category.

02 The PR Tier System Is Misaligned With AI Reality. Reuters outranks Forbes. Forbes outranks most Tier 1 media. The Wall Street Journal and The New York Times often don't appear at all.

03 AI Citations Are Long-Tail, Not Winner-Take-All. Outside the dominant tier, distribution across many sources consistently outperforms concentration in a few.

04 Platforms Are Volatile. Reddit's share on ChatGPT collapsed from ~60% to ~10% of prompt responses in two weeks during September 2025 (SEMrush, Nov 2025). Static strategies fail.

05 Each AI Engine Is Different. There is no single “AI SEO” strategy. Each platform requires distinct optimization.

AI Engines Don't Rank Authority. They Assemble Answers.

This is the single biggest shift from traditional PR. AI engines don't rank — they assemble. They favor extractable content, prioritize consensus across sources, and reward repetition over prestige.

For 18 months, the industry has been asking: what is our AI strategy? Most answers have been vague. Create more content. Get on Reddit. Build thought leadership. The advice is directionally correct — but misweighted and incomplete.

Multiple large-scale datasets published in 2025 and 2026 — from Similarweb, SEMrush, Profound, Peec AI, SE Ranking, Goodie, Ahrefs, and Evertune — now allow the industry to move from anecdote to operational framework.

The data converges on a structural insight that traditional PR thinking has not absorbed:

T H R E E C O N S E Q U E N C E S

- AI engines favor extractable, structured content over narrative or prestige.
- AI engines prioritize consensus across many sources over a single authoritative one.
- AI engines reward repetition across the web over editorial endorsement.

Authority is no longer controlled by editors. It is distributed across platforms. The brands that win in AI visibility are not the ones with the most prestigious clip book. They are the ones whose name appears, consistently, in the structured surfaces the models actually pull from.

Nine Datasets. Hundreds of Millions of Citations and Prompts.

This is a synthesis report. 5W did not independently run the underlying primary research and did not independently verify the data. The report integrates findings from nine separately published studies covering January 2025 through April 2026, and surfaces patterns where the studies converge.

The studies use different units of measurement — some count citation events, others source domains, others unique prompts. The table below reports each in its native unit. Full URLs and publication dates appear in References & Limitations on page 15.

SOURCE	DATASET	COVERAGE
Similarweb	~600,000 events	ChatGPT, Google AI Mode (Jan–Feb 2026, U.S.)
Peec AI	30M sources	ChatGPT, AI Mode, Gemini, Perplexity, AI Overviews
SEMrush	325K + 230K prompts	13-week cross-platform tracking
Profound	1.4M citations	Six AI models tracked
SE Ranking	129K domains	Domain-level correlation analysis
Goodie	5.7M citations	Feb–Jun 2025, four engines
Ahrefs	75K brands	December 2025 correlation study
Evertune	200M prompts	Long-tail distribution analysis
Passionfruit	12-month synthesis	March 2026 cross-study review

T H E Q 2 2 0 2 6 P R I M A R Y R E S E A R C H R U N

The next edition will layer 5W's own primary research on top of this baseline. We will run 1,500 fixed prompts (600 branded, 600 category, 300 executive) across ChatGPT, Claude, Perplexity, Gemini, and Google AI Mode in a single calendar week, classify every citation against a 12-bucket taxonomy, and publish the dataset and methodology for public replication.

The Top 20 Domains ChatGPT Cites

Similarweb's January–February 2026 dataset of approximately 600,000 citation events provides the cleanest single-platform leaderboard available. Three patterns stand out before the table loads:

- Structured and community sources dominate. Wikipedia, Reddit, YouTube, LinkedIn, GitHub, and Fandom collectively exceed every traditional news outlet in the top 20.
- WSJ, NYT, Bloomberg, and FT do not appear at all. Forbes is the only U.S. business publication on the list.
- ChatGPT cites OpenAI itself third — ahead of Reuters and every news outlet measured. Google AI Mode does the same with Google properties.

#	DOMAIN	SHARE
01	wikipedia.org	13.15%
02	reddit.com	11.97%
03	openai.com	6.21%
04	walmart.com	2.90%
05	youtube.com	2.67%
06	linkedin.com	2.42%
07	reuters.com	2.27%
08	nih.gov	2.22%
09	google.com	2.17%
10	amazon (media-amazon)	1.94%
11	wikimedia.org	1.93%
12	facebook.com	1.76%
13	ebay.com	1.75%
14	amazon.com	1.71%
15	github.com	1.62%
16	apple.com	1.48%
17	yahoo.com	1.44%
18	forbes.com	1.38%
19	fandom.com	1.29%
20	squarespace-cdn.com	1.29%

Source: Similarweb AI Citation Analysis, January–February 2026 (U.S.). Full URL on page 15.

Nine Patterns That Define AI Citation Behavior

Each finding below is what the integrated dataset shows. Each is followed by what it means for communications strategy and the source(s) the finding rests on.

01

Reddit Is Infrastructure, Not a Channel

Reddit ranks #1 across every major AI engine measured.

Reddit ranks first in citation share across most major AI engines. Peec AI's 30-million-source analysis ranks Reddit number one across ChatGPT, Google AI Mode, Gemini, Perplexity, and AI Overviews. On Perplexity specifically, Evertune found Reddit accounts for as many as one in five of all citations.

The mechanism is structural. OpenAI announced a content licensing partnership with Reddit in 2024; Google has its own data agreement. SE Ranking's domain-level analysis found brands with millions of Reddit mentions averaged seven ChatGPT citations versus 1.8 for brands with minimal presence — a 3.9x multiplier.

What most people get wrong: this is not about posting. It is about presence and credibility over time. The platform's culture rewards substance, the LLMs subsequently cite the substance, and promotional behavior is filtered out within hours.

Sources: Peec AI 30M-source analysis; Evertune 200M-prompt analysis; SE Ranking 129K-domain study.

02

Wikipedia Is the Ground Truth Layer

The single most influential document in any brand's AI visibility profile.

Wikipedia is the most-cited single domain in ChatGPT (13.15% of U.S. citations) and a top source on every other major engine measured. It is widely documented across published research as a major training and citation source for the leading LLMs, and the most consistently retrieved authoritative source at inference time when models ground a factual answer.

If your Wikipedia page is weak, AI answers are weak. If it is missing, AI fills the gap, often incorrectly.

Correction to industry thinking: Wikipedia is not optional. The path to a strong page is not direct editing — Wikipedia's notability and reverter rules punish that. The path is earning citation-eligible secondary coverage that other editors then use to build the page.

Sources: Similarweb (Q1 2026); cross-referenced across Goodie, SEMrush, Spotlight.

03

LinkedIn Is the Fastest-Growing Signal

From rank #11 to #5 on ChatGPT in three months — the largest shift Profound observed all year.

LinkedIn climbed from approximately #11 on ChatGPT in November 2025 to #5 by February 2026 (Profound). SEMrush's 325,000-prompt study found LinkedIn cited in 14.3% of ChatGPT Search responses, 13.5% of Google AI Mode responses, and 5.3% of Perplexity responses. For B2B and software queries, Profound found LinkedIn is now the #1 most-cited domain across all six major AI platforms.

Critical nuance: ChatGPT and Google AI Mode pull approximately 59% of LinkedIn citations from individual member content. Perplexity inverts this, pulling about 59% from Company Pages. Both the leadership-publishing effort and the company-page operation matter — and they compound.

Leadership visibility is now a ranking factor. Most communications programs underinvest in named-leader publishing because it does not produce traditional earned-media metrics. The AI citation data overrules the traditional metric.

Sources: Profound (Feb 2026); SEMrush 325K-prompt study; ALM Corp synthesis.

04

YouTube Is the Hidden Power Signal

0.737 correlation with AI visibility — the strongest single predictor in any 2025–2026 study.

Ahrefs' December 2025 study of 75,000 brands found YouTube mentions correlated at 0.737 with appearances in ChatGPT, AI Mode, and AI Overviews — the strongest single correlation in their dataset.

AI engines read transcripts. Mentions persist indefinitely. The video itself is incidental — the transcript is the asset. A single ten-minute video with a substantive brand mention can generate citation lift for months.

The insight: a strong creator-led video mention can outperform a major media hit in AI visibility. Most communications programs do not budget against this. They should.

Source: Ahrefs 75K-brand correlation study, December 2025.

05

Forbes Is the Editorial Exception

The most-cited U.S. business publication on ChatGPT. WSJ, NYT, and Bloomberg do not appear in the top 20.

Forbes ranks 18th in Similarweb's ChatGPT dataset at 1.38% of all citations. The Wall Street Journal, The New York Times, Bloomberg, and Financial Times — all marquee Tier 1 PR targets — do not appear in the top 20 at all in this dataset.

Three structural reasons: paywalls limit body-text extraction; licensing disputes

between LLM platforms and major news publishers have reduced indexing; long-form narrative features produce less clean factual extraction than tighter trade or contributor pieces.

Prestige does not equal extractability. Extractability does not equal citation.

This is not an argument to stop pitching Tier 1. Mainstream coverage retains its value for reputation, financial credibility, and as upstream feedstock to Wikipedia. It is an argument that an earned-media strategy concentrated in Tier 1 only is structurally underweighted on the AI citation layer.

Source: Similarweb AI Citation Analysis, Q1 2026.

06

Review Platforms Drive Decision Citations

Brands across G2, Capterra, Trustpilot, and Yelp see a 3x citation multiplier.

SE Ranking found brands listed on multiple review platforms averaged 4.6 to 6.3 ChatGPT citations versus 1.8 for absent brands. Peec AI confirmed Yelp and G2 specifically appear frequently in recommendation queries. Passionfruit's March 2026 synthesis found brands with G2, Capterra, Trustpilot, and Yelp profiles have approximately 3x higher citation probability than brands without them.

Review platforms function as third-party validation that AI engines treat as authoritative for vendor-comparison and recommendation queries. They provide structured ratings, comparative data, and clear extraction signals.

Action: claim and complete profiles on the three major platforms for the category. Encourage structured reviews — star ratings combined with specific use cases and pros/cons. Generic five-star reviews carry less signal than detailed mid-range reviews.

Sources: SE Ranking 129K-domain study; Peec AI; Passionfruit synthesis (Mar 2026).

07

AI Visibility Is Long-Tail, Not Winner-Take-All

Outside Wikipedia and Reddit, no domain exceeds 3% of ChatGPT citations.

Wikipedia and Reddit sit in their own tier on ChatGPT, at 13.15% and 11.97% of citations respectively. Below them, the distribution flattens dramatically: in the Similarweb data, no other domain exceeds 3% of ChatGPT citations except OpenAI's own properties (6.21%). The remaining seventeen domains in the top 20 together account for roughly 32%, and the rest of the citation share spreads across thousands of long-tail sources. Evertune's separate tracking across 200 million prompts confirms the broader pattern — outside the dominant tier, citation share is broadly distributed

rather than concentrated.

This is a fundamentally different distribution from traditional SEO, where the top 10 results capture roughly two-thirds of clicks. AI search citations are a long tail with a few outliers — not a winner-take-all market.

Traditional SEO rewards rank concentration. AI visibility rewards distribution across many sources.

The strategic consequence: getting mentioned across many high-citation third-party domains is more valuable than ranking your own .com higher. Distributed mentions produce measurable lift in three to six weeks.

Sources: Similarweb (Q1 2026); Evertune 200M-prompt analysis.

08

Depth Beats Authority

Fandom outranks Wikipedia in Google AI Mode. Structure plus depth beats brand authority.

Fandom.com leads Google AI Mode's citation list at 7.16% — ahead of Wikipedia (5.21%), YouTube (4.91%), and Reddit (4.19%). The reason is not simply that AI Mode sees lots of entertainment queries. It is that Fandom pages are structurally optimized for what AI engines prefer.

Fandom pages run thousands of words covering one specific subject, organized under precise headings, maintained by communities with encyclopedic precision. Each page exists to answer one question about one thing.

Generalizable lesson: any brand publishing deep, single-topic reference pages on its area of expertise is building the structure AI engines reward. AI rarely cites homepages — most citations come from pages several folders deep. Specific beats broad. Deep beats wide.

Source: Similarweb AI Citation Analysis (Google AI Mode), Q1 2026.

09

Volatility Is Structural

Reddit's ChatGPT share collapsed from ~60% to ~10% of prompt responses in two weeks (Sept 2025).

The biggest shift of 2025 was the September collapse. Across SEMrush's 230,000-prompt 13-week tracking study, ChatGPT's citation share for Reddit dropped from approximately 60% of prompt responses to roughly 10% in a two-week window. Wikipedia followed a similar pattern, falling from roughly 55% to under 20%. Both partially recovered.

Forbes doubled its ChatGPT citation share in the same period. LinkedIn trended

upward. Some weight redistributed; some collapsed entirely.

Annual AI visibility audits are obsolete. Quarterly is the floor. Monthly is competitive advantage.

The platforms tune retrieval behavior aggressively. Rankings shift meaningfully on a multi-week timescale. Brands measuring annually are reporting against citation patterns that no longer exist.

Source: SEMrush 230K-prompt 13-week tracking study, November 2025.

Five Engines. Five Different Citation Patterns.

There is no single “AI SEO.” Each engine sources differently. A strategy that produces results on one platform is not transferable to another.

P L A T F O R M	T O P 5 C I T E D D O M A I N S (I N O R D E R)
ChatGPT	Wikipedia, Reddit, OpenAI, Walmart, YouTube
Google AI Mode	Fandom, Wikipedia, YouTube, Reddit, Google
Gemini	Reddit, YouTube, Wikipedia, Medium, Forbes
Perplexity	Reddit, LinkedIn, NIH, Microsoft, G2
AI Overviews	YouTube, Reddit, Forbes, LinkedIn, Wikipedia

Sources: Similarweb (Jan–Feb 2026); Peec AI 30M-source analysis; Profound; SEMrush. Full URLs on page 15.

K E Y P A T T E R N S

- Reddit appears in the top five on every platform. It is the universal channel.
- Google's surfaces favor Google-owned properties — YouTube, Google.com, Fandom (Google has long-standing partnerships with Wikia/Fandom).
- Perplexity skews toward research-credible sources: NIH, G2, structured B2B data. Most footnote-explicit of the engines.
- ChatGPT is the most Wikipedia-heavy. Strong Wikipedia content disproportionately moves ChatGPT visibility.
- Gemini integrates Google search results directly. Strong traditional SEO converts to Gemini visibility more than to any other engine.

Citation Behavior Varies Sharply by Vertical

The presence and quality of vertical trade media is one of the strongest predictors of how a category is described in AI. Categories with strong specialized trades see those trades dominate. Categories without them see Reddit, Wikipedia, and review sites fill the vacuum.

INDUSTRY	DOMINANT CITATION SOURCES
B2B SaaS	G2, Capterra, Reddit (r/SaaS, r/sysadmin), GitHub, LinkedIn, vertical trades
Beauty	Reddit (r/SkincareAddiction), Glossy, WWD, Allure, dermatologist sources
Fintech	American Banker, Banking Dive, Reddit (r/personalfinance), .gov, SEC filings
Healthcare	NIH, .gov, .edu, peer-reviewed journals, STAT, Healthcare Dive
Travel	TripAdvisor, Yelp, Reddit (r/travel), Skift, Hotel Management
Cannabis	MJBizDaily, Marijuana Moment, Leafly, Reddit (r/trees, state subreddits)
Legal	Law360, Above the Law, ALM properties, bar associations, case law
CPG	Modern Retail, Retail Dive, Food Dive, AdAge, Reddit, review aggregators

Synthesis based on patterns observed across the nine source datasets and 5W's industry experience.

THE OVERARCHING PATTERN

In high-stakes verticals — healthcare, finance, legal — government, academic, and authoritative sources carry disproportionate weight. The models recognize where source authority matters most.

In consumer-facing verticals — beauty, travel, CPG — community platforms, review aggregators, and influencer-published content lead. Trust signals are distributed across many sources rather than concentrated in editorial brands.

In B2B — SaaS, professional services — review platforms (G2, Capterra) and LinkedIn lead, with vertical trades providing the editorial layer. Wikipedia matters less than in consumer categories.

What This Means for Brands

AI engines reward distribution, not concentration. The brand that appears in many places consistently beats the brand that appears in one place authoritatively.

To win in AI visibility, brands must execute against four mutually reinforcing levers. None is optional. Each one feeds the others.

01 Control Your Ground Truth

- Wikipedia page — accurate, complete, well-sourced to citation-eligible publications.
- Owned site — About page, leadership bios, product pages, and newsroom written in the language you want repeated by AI.
- Schema markup and structured data on every page that matters.
- Press releases and corporate communications consistent with your ground-truth language.

02 Build Distributed Authority

- Reddit — active brand presence, founder/operator participation, AMAs, expert contribution.
- LinkedIn — named-leader publishing on a weekly cadence; active company page.
- YouTube — seeded mentions in category creator content, reviews, comparisons, tutorials.
- Review platforms — G2, Capterra, Trustpilot, Yelp profiles claimed and structured.

03 Create Extractable Content

- Case studies with specific numbers, named clients, and structured outcomes.
- FAQs that answer one question per page in clear, extractable language.
- Deep vertical reference pages that own a single topic decisively.
- Original research and proprietary data — citations compound on findings nobody else has.

04 Increase Repetition Across Sources

- Mentions are more important than backlinks.
- Distribution across many sources is more important than ranking in any one.
- Earned coverage in citation-eligible publications feeds Wikipedia upstream.
- Repetition across the surfaces AI engines pull from creates the consensus signal that drives citation.

Sources and What 5W Did Not Verify

P R I M A R Y S O U R C E S T U D I E S

- 01** Similarweb (Apr 2026). The Most Cited Domains by LLMs. similarweb.com/blog/marketing/geo/most-cited-domains-llms
- 02** SEMrush (Nov 2025). The Most-Cited Domains in AI: A 3-Month Study. semrush.com/blog/most-cited-domains-ai
- 03** Goodie (Sep 2025). What Are the Most Cited Domains in LLMs? higoodie.com/blog/most-cited-domains-in-llms
- 04** Contently (Apr 2026). Top 10 Sources LLMs Cite Most in 2026. contently.com/2026/04/29/top-sources-llms-cite
- 05** Passionfruit (Mar 2026). How LLMs Search for Citations. getpassionfruit.com/blog/how-llms-search-for-citations-what-they-look-for-and-what-they-actually-find
- 06** Wellows (Nov 2025). Cited by ChatGPT: 7K Queries, 485K Citations. wellows.com/insights/chatgpt-citations-report
- 07** xSeek. AI Source Radar: Track What Sources LLMs Cite. xseek.io/sources
- 08** Profound (via ALM Corp synthesis). LinkedIn rank shift on ChatGPT, Nov 2025–Feb 2026. almcorp.com/blog/linkedin-ai-search-citations-2026
- 09** Ahrefs (Dec 2025). 75K-brand correlation study (via BrandMentions synthesis). brandmentions.link/ahrefs-brand-mentions

W H A T 5 W D I D N O T V E R I F Y

- 5W did not run the underlying primary research in this Q1 edition. The Q2 edition will include 5W's own primary research run as described on page 5.
- Where studies disagree at the margin, the largest dataset is generally weighted most heavily; specific disagreements are surfaced in the body of each finding.
- All percentages, rankings, and correlations are reported as published by the original researchers.

L I M I T A T I O N S

- This is a U.S.-focused synthesis. International citation patterns are not covered in this edition.
- The September 2025 volatility event documented in Finding 9 is a clear warning that citation patterns shift on multi-week timescales. Findings holding in Q1 2026 may shift by Q2.
- Some claims about LLM training data composition are widely discussed in the trade press but not fully documented by the model developers themselves. Where this is the case, the report uses cautious language and stops short of asserting specific training-weight figures.

About 5W

5W is the AI Communications Firm, building brand authority across the platforms where decisions now happen — ChatGPT, Claude, Perplexity, Gemini, and Google AI Overviews — alongside earned media, digital, and influencer channels. 5W combines public relations, digital marketing, Generative Engine Optimization (GEO), and proprietary AI visibility research, helping clients measure and grow their presence in AI-driven buyer research.

Founded more than 20 years ago, 5W has been recognized as a top U.S. PR agency by O’Dwyer’s, named Agency of the Year in the American Business Awards®, and honored as a Top Place to Work in Communications in 2026 by Ragan. 5W serves clients across B2C sectors including Beauty & Fashion, Consumer Brands, Entertainment, Food & Beverage, Health & Wellness, Travel & Hospitality, Technology, and Nonprofit; B2B specialties including Corporate Communications and Reputation Management; as well as Public Affairs, Crisis Communications, and Digital Marketing, including Social Media, Influencer, Paid Media, GEO, and SEO. 5W was also named to the Digiday WorkLife Employer of the Year list.

For more information, visit www.5wpr.com.

The 5W Citation Source Audit — Q1 2026

Published by 5W. Quarterly cadence. Q2 2026 publishes July with 5W's first primary research run.