

Citations within Discoverability Engine

All Brands today which are serious about digital marketing. are asking only one question - "Why are the LLMs not mentioning our brand?". The answer to this question lies in which domains the LLM model reaches out to for synthesising an answer. The domains could be both generic and industry specific niche domains. In this report we analyze top generic cross brand domains cited by LLMs within Discoverability Engine.

Report Type: Analytics and Insights

Industry: Cross-Industry

Geography: Global

Author / Research Team: Discoverability Engine Analyst Team

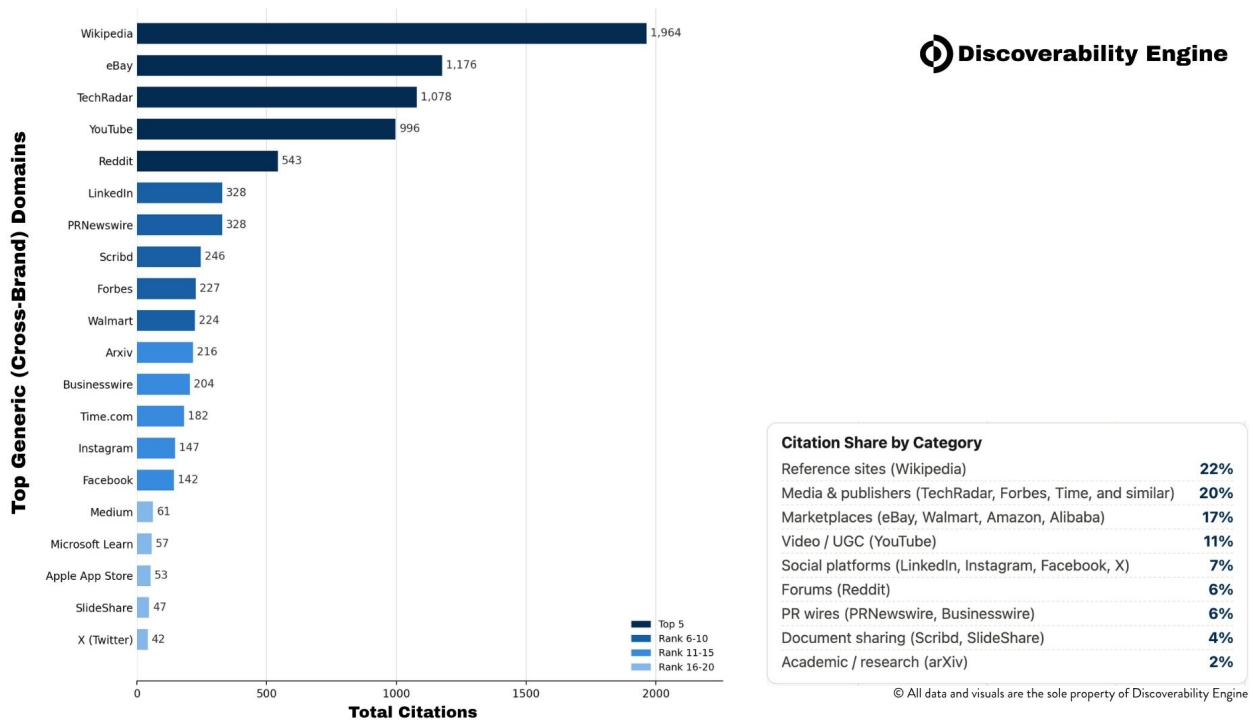
Publish Date: 5 September 2026



Our teams analysed various domains cited across brands and Industries within the Discoverability Engine, which is partly a Generative Engine Optimization tool.

The analysis was done across 26 brand spanning 11 industries, from E-commerce to finance to fitness, education and technology. We isolated those domains which we generic and showed up repeatedly across three or more different industries. These domains became the trustworthy source, regardless of what the question is about.

Isolating these domains, we had 8883 citations across 38 domains. These domains are then grouped into 9 buckets, giving us a clear understanding of why specific domains are picked by LLMs. This analysis will help brand marketers and content creators develop their strategy for discoverability and AI visibility.



Above graph shows the top 20 domains only for the sake of visual representation.

The nine buckets along with their percentage share in generic domain citations are as below

- Reference Websites and Encyclopedias (E.g. Wikipedia) — **22%**
- Media & Publisher Websites (E.g. TechRadar, Forbes, Time & similar outlets) — **20%**
- Marketplaces (E.g. eBay, Walmart, Amazon, Alibaba) — **17%**
- Websites hosting Video / User Generated Content (E.g. YouTube) — **11%**
- Social platforms (E.g. LinkedIn, Instagram, Facebook, X) — **7%**
- Discussion Forums (E.g. Reddit) — **6%**
- PR wires (E.g. PRNewswire, Businesswire) — **6%**
- Document sharing websites (E.g. Scribd, SlideShare) — **4%**
- Academic and Research Websites (E.g. arXiv) — **2%**

LLMs while synthesising an answer to a user's prompt query, do not build them exclusively from brand owned websites. When LLMs provide citations their sources come from a much broader information eco-system - Brand's entire digital footprint. A brand can have an excellent website and still have a weak AI visibility if the broader web does not have a footprint with corroborating information about it.

These nine buckets listed in our report, gives us well categorised layers for which content and it's distribution strategies can be formed. The domains listed within these categories are just the base layer and not the ceiling. Every brand should try to build their digital footprint on these generic, AI trusted sources along with their industry specific sources.

Three buckets — Reference websites, Media and Publisher websites and Marketplaces, together make up roughly 60% of the generic citation share. This is clear indication that LLMs default to encyclopedic sources, reputed media coverage and trusted marketplace listings before they reach for anything else.

Social media, forums and document sharing websites combine for a share under 20%. Most brand marketing teams are focused on this layer by default. This is a clear indication that marketing efforts should be split wisely across the layers defined creating a strong digital footprint which translates into AI visibility and commercial returns.

Let's look at each of the buckets in detail:

Reference Websites and Encyclopedias (E.g. Wikipedia) — 22%

LLMs consider a brand or a person or an object as an entity. Reference websites and Encyclopedias tell AI "What is this entity?".

In our report, reference sources account for approximately 22% of the citation share. This is a very insightfully disproportionate metric. Platforms like Wikipedia are not trying to sell something. Information present there is backed by references. It provides true information about a company or a brand establishing right entity identity and context. It will tell the user (Human or AI) what the company is about, when was it founded, notable products and services, history including highs and lows, competitors, events and more.

This is important for generative engine optimisation because it is not just about what the brand says about itself but what the broader web says the brand is.

Questions a brand marketer must consider - Does your brand have entries on websites like Wikipedia? Does your brand qualify to have a wikipedia entry? If it already has that entry, how accurate, current and well-sourced is it?

It is important for brand marketers to maintain accurate and consistent information across authoritative reference sources to turn them into an effective component of AI visibility.

Media & Publisher Websites (E.g. TechRadar, Forbes, Time and similar outlets) – 20%

Editorial coverage with digital media and publishers provide independent validation. A company may make a certain claim, but when an independent publication validates or denies it, then that is significant. It provides verification and contextual signal to the company's claim.

Brand marketers should build media coverage around their best rated products and services, product comparisons, industry trends, market rankings, expert opinions, product reviews, positive developments, technology trends and more.

Editorials on trusted media websites build contextual authority around the brand and makes it a part of the content ecosystem that AI trusts and uses to answer questions. It should become an AI visibility strategy, not simply a traditional brand awareness exercise.

Questions a brand marketer must consider - Does your brand have earned coverage on established digital media sources and publishers? Are your subject matter experts well credited along with their correct credentials in order to establish their authority in the industry?

Marketplaces (E.g. eBay, Walmart, Amazon, Alibaba) – 17%

For B2C brands, marketplaces tell important information about what this business sells and how it is positioned commercially.

In our analysis market places have been cited more than often for businesses. These include platforms like amazon, eBay, Walmart and Alibaba.

Why do LLMs cite marketplaces?, This is mainly because there is enormous amount of product level information. Marketplaces can become an excellent source of AI visibility. They provide information like product names, specifications, prices, categories, review and ratings.

For marketplace listing to be cited, it is important that the information is consistent across all product listings on the web, be it brand's own website, google merchant centre or marketplaces. Should AI detect discrepancies in the product information across these touch points, it might drop your brand from recommendations citing reasons such as bias and hallucination.

At times, it may also happen that your brand has excellent reviews, but they are not on your website. They might be on the marketplace listing or on Google reviews. Cross referencing marketplaces and Google review links on your website is essential so that AI can easily find them and attribute them to your brand.

Questions a brand marketer must consider - Is your brand correctly listed across the marketplaces relevant to your industry category? Is the listing consistent across various sources including your own website, Google Merchant and other platforms?

Websites hosting Video / User Generated Content (E.g. YouTube) — 11%

AI visibility is not dependent just on textual content. Video are a major source of content generation and consumed today. They contain enormous information that can contribute to how products, brands and topics are represented online.

Videos provide more detailed and contextual information than what a brand's website or product listing can provide. They bring a very different kind of brand signal to the mix. Videos may contain information like product unboxing, usage demonstration, reviews, tutorials, expert commentary, customer experiences and more. This is the experiential information which can provide AI with positive or negative sentiment that is present in the market about the brand.

AI visibility is one thing but the sentiment attached to it is equally important to know whether it will translate to commercial benefit or not. A Brand can have high visibility but what if the sentiment is negative!

Questions a brand marketer must consider - Do we have enough videos in our digital footprint mix? Are our videos getting enough views? Are there any user generated videos about our brand, if yes, are they positive or negative? Do we have product demonstration videos that may make lives of our customers easy? What language should our videos be in?

Social platforms (E.g. LinkedIn, Instagram, Facebook, X) — 7%

Social platforms form an interesting part of the AI visibility. This layer provides signals like presence, freshness, narrative and sentiment.

For B2B companies, LinkedIn can be particularly valuable because it provides brand with credibility through number of comments, credentials of people interacting and whether they have established themselves as subject matter expert or not.

LinkedIn is often considered a strong source of citations in LLM results. Since the user base on LinkedIn is of industry professional, it ranks above all other social platforms.

For B2C or D2C brands, Instagram and Facebook can provide signals like their popularity or sentiment, if a certain topic is discussed in volumes. Here trustworthy sources or should we say, influencers matter.

Social platforms are highly valuable for GEO and AI visibility because they bring the freshness element to the table. These platforms are continuously updated and have virality in their operational nature. What millions of people say can be taken as an authentic sentiment.

Questions a brand marketer must consider - How is AI getting influenced through social media (pun intended)? What recent developments are generating conversations about us? Are competitors generating more conversations around new developments? Is AI reflecting our current social narrative or outdated information?

Discussion Forums (E.g. Reddit) – 6%

Discussion forums like Reddit have been cited as sources more often than not.

Platforms like Reddit have real world user discussions. They give information on what user thinks about a brand. This becomes a good source of information when LLMs come across as recommendation type of queries.

Community discussion forums provide first hand experiences, complaints, recommendations, product or service alternatives, comparisons and unfiltered opinions.

AI systems are effective when they train on real world information, how users interact with each other, how they form sentences, what kind of sentences generate a particular sentiment, they get a clear understanding on what is bias. What better training ground for this than platforms like Reddit. The value that these discussion forums provide has led to companies like OpenAI and Google to partner with them for billions of dollars to get content and train their AI models on it. Best place for a brand to be on - Reddit!

A brand may have an excellent website, strong SEO and great PR but may struggle with recommendation queries if there is no or negative presence on these online communities. This makes reputation management an integral part of GEO.

PR wires (E.g. PRNewswire, Businesswire) – 6%

PR wires tell AI about what new information has an organisation announced. Press releases have been a mechanism for distributing news. News means fresh content while adding to the digital footprint of the brand.

PR wires may contain information about new launches, partnerships, mergers and acquisitions, market expansion and corporate milestones. PR content is often syndicated across multiple websites, brand's own news or blog sections, social media announcements etc. This means a single announcement can generate multiple references to the same entity (organisation) across the web. This makes the organisation discoverable and corroborated across the information eco-system.

The goal of the brand marketers should be to create a meaningful, factual and consistent digital footprint about their brand.

Document sharing websites (E.g. Scribd, SlideShare) — 4%

Document sharing platforms such as Scribd and SlideShare represented approximately 4% in our analysis. While the percentage share may come across as insignificant, it does say something important about AI retrieval of information.

Documents may contain presentations, white papers, reports, case studies, research, product information, industry analysis, training material and more. They provide highly concentrated, to the point and contextual information about an organisation, it's industry or a certain topic it deals with.

Content distribution matters. Brands may publish information on their own website but this limits the distribution and audience to the website visitors only. Document sharing platforms provide for a wider reach. If a larger network of audience interacts with your company's digital footprint, AI agents may take that as a signal of a strong authority and expertise score and is likely to cite your content more than your competitors.

Academic and Research Websites (E.g. arXiv) — 2%

Academic and research sources tell AI what evidence or technical knowledge exists around a brand or a company.

Academic and research sources are not relevant to all industries. They belong to a niche. They can be relevant for industry verticals like AI, technology, space, healthcare, science, engineering, mathematics, advanced computing etc. They provide evidence based authority. While the percentage of the citation will always be low of this category, it will be one of the strong indicators.

The lesson is not that every brand needs to publish academic papers. Different industries have different authority ecosystem. A B2B SaaS, a B2C consumer electronics and a pharmaceutical company will have to have very different GEO strategy in order to succeed. It is important to figure out for technical brands, being referenced in research literature can create strong associations between the brand, technology and underlying concepts.

Conclusion

The objective isn't to make every platform talk about your brand. It is to ensure that the right information about your brand exists in the right places, in the right context, with enough independent corroboration for AI systems to recognize, understand and confidently surface it.

Our citation data points towards broader GEO operating model.

- Don't optimize only the website. Build an ecosystem.
- Don't chase citations blindly. Understand why a source is being cited.
- Don't focus only on rankings. Focus on representation.

- Don't measure only mention rate. Measure source diversity, sentiment and citation quality.
- Don't treat Reddit, YouTube and marketplaces as separate marketing channels. Treat them as AI information sources.
- Don't ask only, "Does AI mention us?" Ask: "What information is AI using to decide whether to mention us or not?"

The future of GEO will not be won by brands that simply publish more content.

It will be won by brands that create a stronger, more consistent and more credible information footprint across the sources AI systems use to construct answers. Our analysis of 38 domains provides a clear signal. AI doesn't learn about brands from one place. It learns from an ecosystem.