WHITEPAPER · AUGUST 2026

What makes a trusted source

We asked three AI systems to describe two real people, then checked every source they cited. A page ranking sixth in Google was cited zero times. A government register was cited twice. This paper sets out what we measured, what it means, and where we were wrong.

ASCRIBED PRIMARY RESEARCH 15 AUGUST 2026 APPROX. 4,200 WORDS

01

Why the question changed

For twenty years the question was which page ranks. That question is being replaced by a different one, because the interface changed. A ranked list hands the reader ten options and lets them choose. An answer chooses for them, and to choose it must decide which sources it believes.

That decision is now the whole game, and it is not the same decision as ranking. A page can rank well and never be cited. A page nobody visits can be cited constantly. The two systems are optimising for different things: one for what a person might want to read, the other for what a machine can safely repeat.

We wanted to know what the second system actually trusts. Not in theory, and not from vendor documentation, but by asking it and reading the receipts.

02

What we did

We took two real subjects, both of whom consented, and asked three AI systems to describe them. Then we recorded every source each system cited.

Separately, we took two real press campaigns and checked all 29 published articles by hand, confirming whether each one actually named the subject rather than merely covering the story. And we watched a knowledge panel appear from nothing over eleven days, with a dated capture on each side of it.

3

AI systems queried

29

Articles verified individually

39

Citations recorded and classified

Everything below is drawn from those measurements. Where a measurement failed, or where we could not verify something, we say so rather than leaving it out. Section ten sets out the limitations, of which there are several.

A declared interest. We sell the thing this research is about. That is a reason to read our numbers sceptically, and a reason we have published the method in enough detail for anyone to repeat it and disagree.

03

Finding one: models cite registers, not rankings

Subject A has an open wiki entry that ranks **sixth in Google** for his own name. It is well written, it carries his photograph, and it is almost certainly feeding his knowledge panel: the date of birth Google displays appears in machine-readable form nowhere else.

When we asked the models about him, that page was cited **zero times**.

Citations recorded for subject A. Google returned no AI Overview for the query at all, so there was nothing to cite from.

What Perplexity did cite, for subject A: LinkedIn twice, a journalist database twice, an author archive, his own website, and trade press. What ChatGPT produced was a fluent, confident paragraph with **no sources whatsoever** and no checkable specifics.

Subject B is the contrast that makes the point. He has a structured record and a company on a public register. Perplexity cited **Companies House twice**, an encyclopaedia, his own site and LinkedIn, and correctly named his role and company. ChatGPT, which produced nothing citable for subject A, cited a regulatory news distributor for subject B and got his position right.

> The difference between the two subjects was not popularity, effort or spend. It was whether a structured, third-party record existed that stated plainly who they were.

This is the single most useful thing we measured. Ranking and being cited are different outcomes with different inputs. An open wiki can rank well because it is a well-optimised page about a person. It is not cited, we think, because it is user-editable, and a system that must stand behind an answer has a reason to prefer a register that cannot be casually rewritten.

04

Finding two: a byline is not a citation

Subject A has written for national outlets for over a decade. His name returns 28 results on the first three pages of Google. We classified all of them.

WHAT THE RESULT IS	COUNT	BUILDS THE ENTITY
Profiles he owns or controls	11	Weakly. It corroborates identity, it does not evidence it
Author archives, things he wrote	6	No. Authorship, not attestation

Public registers	2	Yes, strongly
Open wiki	1	Ranks, but is not cited by models
Data brokers and aggregators	6	No
Independent articles about him	2	Yes. This is the scarce material

He is indexed as an **author**, not as a **subject**. Almost everything findable under his name is either something he owns or something he wrote. That is the occupational trap of anyone who has spent a career publishing on behalf of others.

The distinction matters because the two do different work. A byline demonstrates that a publisher was willing to print your words. A citation demonstrates that somebody independent thought you were worth writing about. Entity confidence is built from the second and only lightly assisted by the first.

The practical test

Ask of any result: **did somebody else decide to write this about me?** If the answer is no, it is corroboration at best. Both belong on a record. Only one of them answers the question a machine is asking.

05

Finding three: coverage counts are inflated by syndication

We took two real campaigns and verified every article individually, visiting each live page and confirming the subject was named in the body rather than trusting the coverage report.

CAMPAIGN	LISTED	NAMED THE SUBJECT	INDEPENDENT ORIGINALS

Subject A, June 2026	17	12	8
Subject C, July 2026	12	6	6
Combined	29	18	14

Two separate shrinkages happen here, and they compound.

First, **not every article about a story names the person behind it**. For subject C, six of twelve listed items did not mention him in the body at all. A coverage report that counts outlets is answering a different question from one that counts citations. Both are legitimate. Only one of them is about the person.

Second, **syndication multiplies domains without multiplying sources**. In subject A's campaign, four South African regional titles carried one identical article and two Australian titles carried another. Six mastheads, two pieces of journalism. The domains are genuinely different and the editorial decisions were genuinely separate, but the text is the same, and a system looking for independent agreement can see that as easily as we can.

> Eighteen outlets is a true statement and a misleading number. Six independent originals is the honest one, and it is still a good campaign.

This is the specific mechanism by which mass distribution disappoints. The same release on four hundred sites is one source wearing four hundred mastheads. It is not fraudulent and it is not worthless, but it does not do what buyers are usually told it does.

06

Finding four: a contested name suppresses the entity

Subject A shares his name with at least five other findable people. One of them outranks him for his own name. We confirmed on a public register that the higher-

ranking man is a different person, born **sixteen years later**. A professional network offered "20+ profiles" under the same name.

This is not a cosmetic problem. Every signal that arrives under a shared name has to be allocated to somebody, and where allocation is ambiguous, the safe behaviour for a machine is to hedge. That is exactly what we observed: a panel that exists and stays thin, and a model that produces fluent description with no citations attached.

Why identifiers beat adjectives

A description says what somebody is like. An identifier says which one they are. A company number, a register entry, a stable profile URL repeated identically everywhere: these are boring, and they are what resolves the ambiguity that otherwise caps everything else you do.

07

Finding five: structure triggers, coverage sustains

Subject B had three decades of public record and no knowledge panel. We built the structured layer underneath his existing footprint: a referenced record on a public knowledge base, one block of structured data on his own site describing six linked entities, and a single consistent job title rolled out everywhere sources disagreed.

INPUT	QUANTITY
	11
	1

	6
	0
	0

The attribution is unusually clean, and it is the panel itself that supplies it. The panel displays a month of birth and an education history. **Both facts exist in machine-readable form in exactly one place on the open web: the record we created.** They are not in his press coverage, because his press coverage does not contain them.

We are not claiming this generalises to everyone. It generalises to people who already have a real public record that nothing has ever tied together, which in our experience is most senior people. Where no genuine record exists, none of this applies, and the honest answer is to say so and stop.

Note what this does **not** say. It does not say coverage is unnecessary. The panel that appeared is thin, and it is thin because there is little independent writing to build a description from. Structure is what makes an entity resolve. Coverage is what gives it something to say. Buying one and calling it the other is the most common mistake in this market.

08

Finding six: you cannot measure a panel with a scraper

We include this because we got it wrong, publicly, during the research.

A commercial search-results API reported that a claimed knowledge panel had no description and linked to no official website. We wrote that up as fact. The panel's verified owner could see both, on his screen, at the same moment.

The explanation is mundane and important. A scraper sees an anonymous, non-personalised view from a datacentre. Panel elements load client-side and frequently never appear in a static capture. A claimed panel shows its verified owner more than it shows an anonymous request. **Absence in a scrape is not absence in the panel.**

> **The rule we now work to.** Monitoring must report three states, not two: present, absent, and *not observed, with the reason*. Anything a tool cannot see is recorded as unobserved and never as missing. Where a commercial promise depends on the measurement, it is settled against the platform's own structured interface or by a human looking, not by a scrape.

We think this is underappreciated across the industry. A great deal of reporting in search and reputation is generated by tools that cannot see what they claim to be measuring, and reported to clients as fact. The failure is quiet, because a false negative looks exactly like work still to do.

09

A hierarchy of trusted sources

Pulling the findings together, this is the order we now work in. It is a working hypothesis built on a small sample, not a law.

TIER	EXAMPLES	WHY IT CARRIES WEIGHT
1. Public registers	Company registries, regulatory filings	Legally maintained, not editable by the subject, unambiguous identifiers. Cited directly by models in our tests
2. Structured knowledge bases	Referenced, machine-readable entity records	Purpose-built for machines. Demonstrably fed a panel within days in our measurement

3. Independent editorial	An article somebody else wrote about you	Scarce, and the only thing that gives an entity something to say
4. The entity's own marked-up site	schema.org on a domain you control	The anchor everything else is tied back to. Self-published, so it corroborates rather than evidences
5. Owned profiles	Professional networks, journalist databases	Cited heavily in our tests, but easy to create, so they establish consistency more than credibility
6. Syndicated copies	The same article on many domains	Reach without independence. Counts once
7. Open wikis	User-editable encyclopaedias outside the major projects	Can rank well. Cited zero times in our tests, and rejected as a reference by the projects that matter
8. Author archives	Your own byline pages	Evidence you write, not evidence you matter

The uncomfortable implication for anyone selling in this market is that tiers one and two are cheap and unglamorous, and tiers six and eight are what most packages are actually made of.

10

Registers are national. Identifiers are not.

The hierarchy above puts public registers at the top, and our measurement showed a UK register being cited directly by a model. That finding comes with an obvious limit: there is no global Companies House. Anyone outside the United Kingdom needs the local equivalent, and they are not equivalent in quality.

JURISDICTION	THE REGISTER	WHAT IT GIVES YOU

United Kingdom	Companies House	Company number, officer ID, month and year of birth, appointment history. Free, complete, and the strongest of the lot
Ireland	Companies Registration Office	Company number, directors
United States	SEC EDGAR, plus state registries	No single national company register. EDGAR covers filers and gives a CIK; otherwise it is state by state
Germany	Handelsregister	HRB number, managing directors
France	Registre National des Entreprises	SIREN and SIRET numbers
Australia	ASIC, Australian Business Register	ACN and ABN
Canada	Corporations Canada, plus provincial	Corporation number
Singapore	ACRA	UEN, officers
Hong Kong	Companies Registry	Company number
India	Ministry of Corporate Affairs	CIN, and a Director Identification Number per person
UAE	DIFC and ADGM public registers; mainland licensing sits with each emirate	Free-zone registers are searchable. Mainland is fragmented, which is a real problem for Gulf-based subjects

Two things follow from that table. The first is that a client's jurisdiction materially changes how much tier-one evidence is available to them, and a UK director is unusually well served. The second is that fragmentation is exactly the problem cross-border identifiers were invented to solve.

The portable layer

Identifiers travel where registers do not. These are the ones worth attaching to a record, and most have a dedicated property on the major structured knowledge bases, which is the practical reason they matter.

IDENTIFIER	COVERS	WHY IT HELPS
LEI	Legal entities, globally	ISO 17442, issued under a single global system. The closest thing to a universal company identifier
OpenCorporates ID	Companies across many jurisdictions	Aggregates national registers into one addressable record, useful where the register itself is hard to link
ISIN	Listed securities	Ties a person's company to a traded instrument
ORCID	Researchers and authors	A persistent identifier for a *person*, which is rarer and more valuable than a company one
ISNI and VIAF	Authors, creators, public figures	Maintained by libraries. Slow, conservative, and precisely the kind of source that is hard to game

The rule that survives the jurisdiction

Ask of any source: **can the subject edit it, and can a machine address it unambiguously?** A register scores well on both because it is maintained by somebody else and issues a number. That test works in every country, which the register list does not.

What we did not measure. Our citation data covers one register, in one country, for one subject. We have no evidence about how models treat EDGAR, ASIC, ACRA or the free-zone registers, and it would be wrong to

assume they are weighted like Companies House. Treat the table above as where to look, not as a ranking.

11

Limitations, and what would change our mind

This is a small study and we would rather say so than have it said for us.

- **Two subjects.** Two people are an illustration, not a sample. The register-versus-wiki contrast is striking, but it rests on a single comparison.
- **One point in time.** Every measurement was taken in a few days in August 2026. These systems change weekly.
- **Google's AI Overview never triggered.** It did not appear for either person-name query, so our third system contributed nothing except its own absence.
- **Gemini went untested.** The interface rejected our model identifiers. It is one of the four surfaces we care about and we have no data on it.
- **The models were reached through a commercial proxy**, not directly. We cannot fully separate the proxy's behaviour from the model's.
- **Citation lists are not reasoning.** A cited source is one the system chose to show. What it weighed internally is not visible to us or anyone else outside.
- **We sell this.** Stated again because it belongs in the limitations, not just the introduction.

What would change our mind: a larger subject set showing open wikis being cited regularly; a panel appearing with no structured record behind it; or evidence that syndicated copies are weighted independently rather than collapsed. We would publish any of those.

12

Method

Citation testing

Each model was asked a single neutral question of the form "Who is [name], the British [occupation]?", with web search enabled, through a commercial API. Every URL in the returned citation structure was recorded and classified by hand. No prompt engineering was used, and no attempt was made to steer the answer.

Article verification

Every article in each campaign's coverage list was loaded in a real browser. An article counted as naming the subject only if the surname appeared in the rendered body text. Items behind a paywall were recorded as *unverified*, never as failures, because we could not read them. Articles sharing an identical headline were collapsed into one original with its syndications listed separately.

Panel measurement

Baseline and result were captured as dated full-page screenshots of the same query on consecutive days. Panel contents were additionally pulled through a search API, and, after the error described in section eight, cross-checked with the panel's verified owner.

Reproducibility

The verification tooling is ours and unremarkable: a headless browser, a name check against rendered text, and a grouping step. Anybody with a search API and an afternoon can repeat this on their own subjects, and we would be glad to see somebody try and get a different answer.

On the anonymity of subjects

Subject B is named in a separate case study published with his written permission. Subjects A and C are described but not named here, because consent to be a worked example in one document is not consent to appear

in every document. Where a finding depends on a specific verifiable fact, the fact is given without the name attached.

Conclusion

The short version

Being findable and being citable are different problems. The first is solved by content and ranking. The second is solved by being unambiguously identifiable in places that are hard to edit and easy to check.

Most reputation spending goes on tier six and tier eight of the hierarchy above, because that is what is easy to sell and satisfying to receive. The things that moved the needle in our measurements were a register entry, a structured record and a handful of articles somebody else chose to write.

None of that is expensive. It is just unglamorous, and it has to be true.

See how we do it Read the case study

[ascribed]

hello@ascribed.ai · ascribed.ai

Published 15 August 2026 by Ascribed, Dubai. All measurements were taken between 13 and 15 August 2026. The named case study is published with the subject's written permission; other subjects consented to the research and are not named. Screenshots and raw citation lists are held on file and available on request. We sell entity and reputation infrastructure, which is a commercial interest in these findings and a reason to check them rather than take them on trust.