Your Search Engine Has Blind Spots (and a Stance)

A data-driven audit of how seven search engines differ — in what they find, how often they fail, and which side of a conspiracy they surface.

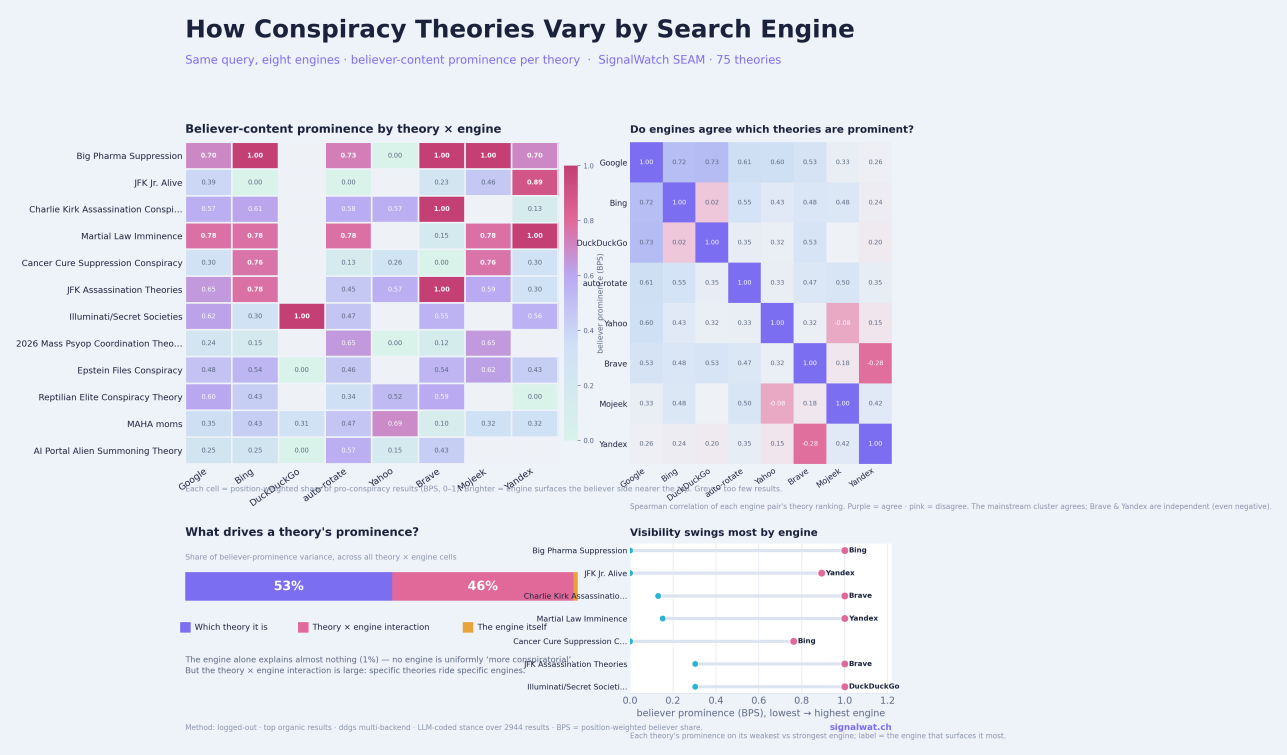
SignalWatch · Search Engine Audit Module (SEAM) · White Paper

The short version

We built a module that runs the same query across seven search engines and measures how they differ. Five findings worth your attention:

- **Our “primary” engine was silently failing.** DuckDuckGo returned 0 results where Bing, Google, Brave and Yandex each returned 10 — the client was quietly rotating away from it and we never knew.
- **No engine is reliable.** Block rates ranged from 10% (Bing) to 100% (Mojeek); even the best top out around 62–75% success.
- **Slowing down barely helps.** A controlled experiment showed request delay is a weak lever — the blocking is IP-level throttling, not rate accumulation.
- **Engines have a stance.** Bing surfaced believer / pro-conspiracy content most prominently; Google surfaced the most debunks — measurable and consistent with the academic literature.
- **It changes what you find.** Converting hits into structured claims, Bing converted at 0.69 and Google at 0.16 — Google’s results are mostly debunks, not the believer content the task needs.

The takeaway for anyone doing OSINT, disinformation research, or content discovery: single-engine search has measurable blind spots *and* a stance bias. Multi-engine, done carefully, is the robust approach.



Engine-variation study: per-theory believer-prominence by engine, cross-engine agreement, and variance decomposition.

Why we looked

SignalWatch tracks conspiracy narratives, which means it leans on web search to discover circulating content. If our search engine systematically omits or down-ranks certain sources, our whole corpus inherits those blind spots — invisibly. So we operationalized the academic *search-engine audit* into something we can run continuously.

Grounding: Urman, Makhortykh, Ulloa & Kulshrestha (2022), *"Where the earth is flat and 9/11 is an inside job"* — a comparative audit that found every engine *except Google* routinely surfaced conspiracy-promoting pages on page one. We reuse their stance and source-type coding, and add continuous cross-engine overlap and position-weighted believer prominence.

Reliability is engine-specific

Our pipeline called what we believed was DuckDuckGo. In reality the client defaulted to auto-rotation and fell back whenever DuckDuckGo was blocked — which, for our IP, was ~95–100% of the time. The system kept working, so the outage was invisible until we measured per-engine.

Engine	Block rate
Bing	**10%**
Google	16%
Brave	30% (degrades under load)

Yahoo	47%
Yandex	95%
DuckDuckGo	95%
Mojeek	100%

A controlled experiment found engines plateau at ~62–75% and adding delay barely moved the needle (Bing 62%→75% from 0s→20s; Google flat). Recovery after a block was ~60 seconds. The limiter is IP-level throttling plus scrape flakiness — so the answer is **rotation + backoff (and rotating IPs), not waiting.**

Engines have a measurable stance

Using an LLM stance coder (promotes / debunks / mentions), we computed a **Believer Prominence Score (BPS)** — a position-weighted share of pro-conspiracy results. For *flat earth*:

Engine	BPS	% promote	% debunk
Bing	**0.54**	**50%**	**10%**
Brave	0.45	30%	10%
Google	0.32	20%	20%
Yandex	0.20	20%	10%

Google surfaced the most counter-narrative content — consistent with the 2022 finding that Google uniquely suppresses conspiracy material. And the stance propagates downstream: converting each engine's hits into structured believer-claims, conversion rates diverged sharply — **Bing 0.69, Brave 0.32, Yahoo 0.20, Google 0.16.** Google returns *more* results but *fewer usable believer claims*. Which surfaces the data-hygiene point we had to fix: **relevance ≠ stance.** A fact-check is topically relevant but is not believer evidence; ingesting it pollutes the dataset. We now gate ingestion by stance.

How much does multi-engine actually add?

Unique source-domains from **Bing alone** vs the **union of all engines**, across six conspiracy queries (via proxy so every engine could respond, isolating coverage from rate-limiting):

Query	Bing alone	All engines	Lift
flat earth	10	21	×2.1
qanon great awakening	10	23	×2.3
chemtrails geoengineering	10	18	×1.8
great replacement theory	9	19	×2.1
vaccine microchip	9	22	×2.4

stolen election 2020	10	28	×2.8
Total unique domains	**58**	**131**	**×2.26**

Multi-engine surfaces ~2.3× the unique sources a single engine does — and the contribution is lopsided. New domains each engine adds beyond Bing: Google +34, Brave +32, Yandex +9, DuckDuckGo +9, Yahoo +6. Google and Brave do almost all the work; the rest are a thin tail.

The proxy fix, measured

We hypothesized the failures were IP-level throttling, so we routed search calls (only) through a rotating residential proxy. Same five queries, each engine direct vs via-proxy:

Engine	Direct (our IP)	Via residential proxy
Brave	**0%**	**80%**
DuckDuckGo	**0%**	**40%**
Google	80%	100%
Bing	100%	100%
Yahoo	40%	40%
Yandex	40%	40%

IP rotation revived the dead engines (Brave 0→80%, DuckDuckGo 0→40%) — their blocking was purely IP-level. But it's not a cure-all: Yahoo and Yandex stayed flat, because their failures are structural. Knowing *which* kind of failure you have tells you whether to rotate IPs or rotate engines.

What we built in response

- **Reliability / outage tracking** — every engine failure is recorded (never silently treated as "no results"), with per-engine block rates and "currently failing" flags.
- **Cooldown rotation** — a Bing-first engine picker that rests each engine between hits and rotates to the next rested engine on failure, with backoff matched to the measured ~60s recovery.
- **Per-request proxy rotation** — search calls (only) can exit from a rotating IP pool to reset IP-level throttling, without disrupting anything else.
- **Stance-gated ingestion** — only believer-side hits become claims; debunks / mainstream are excluded (and historical pollution was pruned).
- **Tiered deep search** — one reliable engine by default; a Bing + Google + Brave fan-out reserved for high-value queries, capturing most of the 2.3× coverage at a fraction of the cost (~¼¢/search).

Caveats

Logged-out / non-personalized condition; first-page only; scraping-based (a licensed SERP API would harden it); engine algorithms drift, so these are snapshots — which is exactly why the audit runs

continuously rather than once.

References: Urman et al. 2022 (arXiv:2112.01278); de León et al. 2024 (arXiv:2409.10531); Webber/Moffat/Zobel 2010 (Rank-Biased Overlap); Golebiewski & boyd 2019 (Data Voids). All figures from live SEAM runs, 2026-06.