



# Methodology

↓ Download  
PDF

Français  
⇒

Last updated : 2026-05-29

sen-ai.fr measures how generative AI systems (ChatGPT, Gemini, Claude, Mistral) reference brands when answering real buyer questions. This page documents how we do it - the data sources, the AI providers, the statistical paradigm, and the safeguards we wire in so the numbers are honest. If you're a procurement / DPO / compliance reviewer, this is the page to attach to your DPA. If you're a marketer, this is the page that explains why a single scan isn't enough.

## 1. The problem we measure

When a user asks ChatGPT "what's a good moisturiser for sensitive skin", the answer surfaces certain brands and not others. That ranking is the new SEO. Unlike Google's search results, AI answers vary across runs - the same question on the same model can produce different brand mixes 30 seconds apart. Anyone measuring AI visibility from one scan is reading noise.

sen-ai.fr's job is to stabilise that signal at scale, then surface the actionable deltas : which questions cite you, which cite competitors, which sources do the AI systems consult, and where on your site do the AI systems land.

## 2. The N-runs paradigm

Every scan repeats each question **N times per provider** instead of asking once. Default for paid plans is N=10 ; teams can configure

lower for cost or higher for confidence. The output is then averaged with a published variance band so the user knows whether a 22% citation rate is "solid at  $\pm 3$ pts" or "noisy at  $\pm 10$ pts".

Why N=10 specifically : empirical variance for the providers we use is ~3-5% for OpenAI grounded mode, ~10-15% for Gemini ungrounded, ~3-5% for Claude grounded. At N=10 the standard error drops to  $\pm 1-2$ pts across the mix - tight enough to detect quarterly trend changes.

Source for the variance characterisation : Fishkin / O'Donnell 2026 measurement study plus our own internal benchmarks across 50+ scans on French cosmetics, pharma, automotive, and consumer-services verticals.

### 3. AI providers & models

sen-ai.fr queries third-party general-purpose AI systems in **read-only mode**. We don't train, fine-tune, or feed responses back into any model. Current mix :

- **OpenAI ChatGPT** - GPT-5.4-mini for cost-balanced runs, GPT-5.4 for premium runs. Grounded mode (web search) when available.
- **Google Gemini** - Gemini 2.5 Flash for cost, Gemini 2.5 Pro for premium. EU-hosted via Google Ireland Ltd.
- **Anthropic Claude** - Haiku 4.5 for structured-JSON extraction (brand mention parser, sentiment judge) ; Sonnet 4.6 for premium runs.
- **Mistral Le Chat** - Mistral Large. EU-hosted. Critical for the French market (41% of Mistral traffic = France).

The mix evolves : providers get added when their adoption crosses the 5% search-share threshold per country. We publish the current mix on the per-scan report so historical comparisons stay reproducible.

## 4. Personas & questions



For each tracked brand we generate a topic taxonomy (from the site's ranking keywords on Google), then a set of personas (synthetic user archetypes) anchored on each topic, then 5 question types per persona (informational, commercial, transactional, etc.).

The questions are **brand-agnostic** on purpose - they describe a buyer situation, not the brand. This avoids leading the AI and matches how real users phrase things in ChatGPT / Perplexity.

The persona + question generators are themselves Claude Haiku calls bound by a strict JSON schema. No human-in-the-loop required for v1 ; the workspace owner can edit any persona or question before launching the scan.

## 5. Brand classification & mention extraction

Every LLM response is parsed by an "analyser" pass (Claude Haiku, strict JSON output) that extracts every brand-like entity, classifies its sentiment (positive / negative / neutral), and tags whether it's the focus brand or a competitor.

Brand classification is workspace-controlled. The system surfaces every entity it finds ; the workspace owner promotes them to *my\_brand* / *competitor* / *ignored* via the Brands tab. Auto-classification is conservative - unknown entities default to *discovered* and stay invisible from the metrics until the owner curates them. No hardcoded vertical / region / brand lists.

A Haiku-as-judge layer (Sprint candidate) will start overturning obvious sentiment false positives (e.g. "not suitable for X" misclassified as negative when the underlying statement is factual). Until then, the radar tools downplay isolated negative signals via conservative severity buckets.

## 6. Citation extraction



When an AI provider exposes citation metadata (OpenAI grounded mode, Gemini Vertex AI search grounding, etc.), we persist the URLs verbatim. When it doesn't, we parse the response text for explicit URL patterns. Each citation is annotated with the publishing domain, the URL, and a 200-character context snippet.

Citations feed several downstream features : Page Audit (which of YOUR pages get cited), Competitors (which pages of RIVALS get cited), PR / Media (which press domains cite the brand vs competitors), YouTube (which video creators surface), Reddit (which threads the AI mines).

## 7. Variance handling & statistical honesty

Single-run metrics are flagged as "low confidence" in the UI. Multi-run metrics surface with their variance band visible (planned in the Variance UI sprint - shipping alongside the N=10 ramp).


Cross-scan comparisons require a minimum of 7 days between scans to avoid measuring intra-week LLM update noise. Trend metrics use rolling averages with a 30-day window.

## 8. Privacy & AI Act posture

sen-ai.fr is classified as **limited-risk AI** under the EU AI Act (Regulation (EU) 2024/1689), applicable 2 August 2026. We are a *downstream deployer* of third-party AI systems, not a provider of general-purpose AI. Transparency obligations apply (Article 50) ; no conformity assessment or CE marking.

Personal data exposure is minimal :

- We store only the email + name of users who access a workspace.

- AI provider prompts contain brand / domain / topic data only -  no end-user PII.
- All data is hosted in the EU (Hetzner, Falkenstein, Germany).
- US-based AI providers (OpenAI, Anthropic) operate under EU-US Data Privacy Framework + SCC.



Full disclosure in our [Privacy Policy](#). Logged-in customers can download a per-scan transparency report and an org-level DPIA template from the in-app Compliance section.

## 8.1 Sub-processors changelog

History of sub-processor additions, sunsets, and scope changes since the initial AI Act disclosure. Updated in lock-step with the in-app compliance pack.

DATE	CHANGE	SUB-PROCESSOR	NOTES
2026-05-29	INITIAL DISCLOSURE	(all)	Initial AI Act compliance pack published. Sub-processors registry frozen at 6 entries (Hetzner, OpenAI, Google Ireland, Anthropic, Stripe Europe, Babbar).

## 9. What sen-ai.fr does NOT do

- We don't train, fine-tune, or contribute to any AI model.
- We don't scrape competitor websites at scale beyond what an LLM already cited.
- We don't impersonate user agents (no Googlebot spoofing, no fake browser strings).
- We don't use prompt injection or any technique that exploits provider terms of service.
- We don't engage in any practice prohibited by Article 5 of the EU AI Act (social scoring, real-time biometric ID, exploitative manipulation, etc.).

## 10. Contact

Compliance / methodology questions : [contact@sen-ai.fr](mailto:contact@sen-ai.fr).

DPO not yet mandated (sen-ai.fr is below the threshold). The  
 above contact handles all data-subject-rights requests and  
replies within 30 days as required by GDPR.



*from data to bloom*

[contact@sen-ai.fr](mailto:contact@sen-ai.fr)

[Privacy](#) [Terms](#)

[Mentions légales](#)

© 2026 sen-ai.fr. Tous droits réservés.