



# Méthodologie

↓ Télécharger  
PDF

English



Dernière mise à jour : 2026-05-29

sen-ai.fr mesure la manière dont les systèmes d'IA générative (ChatGPT, Gemini, Claude, Mistral) citent les marques lorsqu'ils répondent à de vraies questions d'acheteurs. Cette page documente comment nous procédons : les sources de données, les fournisseurs d'IA, le paradigme statistique et les garde-fous qui rendent les chiffres honnêtes. Si vous êtes acheteur, DPO ou auditeur conformité, c'est la page à joindre à votre DPA. Si vous êtes marketeur, c'est la page qui explique pourquoi un scan unique ne suffit pas.

## 1. Le problème que nous mesurons

Quand un internaute demande à ChatGPT "quelle est une bonne crème hydratante pour peau sensible", la réponse cite certaines marques et pas d'autres. Ce classement est le nouveau SEO. Contrairement aux résultats Google, les réponses des IA varient d'un run à l'autre - la même question sur le même modèle peut produire des mix de marques différents 30 secondes plus tard. Quiconque mesure la visibilité IA à partir d'un seul scan lit du bruit.

Le rôle de sen-ai.fr est de stabiliser ce signal à grande échelle, puis de faire ressortir les écarts actionnables : quelles questions vous citent, lesquelles citent vos concurrents, quelles sources les IA consultent et où sur votre site les IA atterrissent.

## 2. Le paradigme N-runs

Chaque scan répète chaque question **N fois par fournisseur** au lieu de la poser une fois. Par défaut sur les offres payantes : N=10. Les équipes peuvent configurer plus bas pour le coût ou plus haut pour la confiance. Le résultat est ensuite moyenné avec un

intervalle de variance publié, pour que l'utilisateur sache si un taux de citation de 22 %  
sonair "solide à ±3pts" ou "bruité à ±18pts".



Pourquoi N=10 précisément : la variance empirique pour les fournisseurs que nous utilisons est de ~3-5 % pour OpenAI en mode grounded, ~10-15 % pour Gemini non-grounded, ~3-5 % pour Claude grounded. À N=10 l'erreur standard descend à ±1-2pts sur le mix - assez serré pour détecter des évolutions trimestrielles.

Source pour la caractérisation de la variance : étude de mesure Fishkin / O'Donnell 2026, plus nos propres benchmarks internes sur 50+ scans dans les verticales cosmétique, pharmaceutique, automobile et services aux particuliers en France.

### 3. Fournisseurs d'IA & modèles

sen-ai.fr interroge des systèmes d'IA généraliste tiers en **mode lecture seule**. Nous n'entraînons, ne fine-tunons et ne ré-injectons aucune réponse dans les modèles. Mix actuel :

- **OpenAI ChatGPT** - GPT-5.4-mini pour les runs équilibrés en coût, GPT-5.4 pour les runs premium. Mode grounded (recherche web) quand disponible.
- **Google Gemini** - Gemini 2.5 Flash pour le coût, Gemini 2.5 Pro pour le premium. Hébergement UE via Google Ireland Ltd.
- **Anthropic Claude** - Haiku 4.5 pour l'extraction JSON structurée (parseur de mentions de marque, juge de sentiment) ; Sonnet 4.6 pour les runs premium.
- **Mistral Le Chat** - Mistral Large. Hébergement UE. Critique pour le marché français (41 % du trafic Mistral = France).

Le mix évolue : les fournisseurs sont ajoutés lorsque leur adoption dépasse le seuil de 5 % de part de recherche par pays. Nous publions le mix actuel sur le rapport per-scan pour que les comparaisons historiques restent reproductibles.

### 4. Personas & questions

Pour chaque marque suivie nous générons une taxonomie de topics (à partir des mots-clés de positionnement du site sur Google), puis un jeu de personas (archétypes d'utilisateurs synthétiques) ancrés sur chaque topic, puis 5 types de questions par persona (informationnel, commercial, transactionnel, etc.).

Les questions sont **brand-agnostic** volontairement - elles décrivent une situation  chat, pas la marque. Cela évite de biaiser l'IA et reflète la manière dont les vrais utilisateurs formuleraient leurs requêtes dans ChatGPT / Perplexity.



Les générateurs de personas + questions sont eux-mêmes des appels Claude Haiku contraints par un schéma JSON strict. Pas d'humain dans la boucle pour la v1 ; le propriétaire du workspace peut éditer chaque persona ou question avant de lancer le scan.

## 5. Classification des marques & extraction des mentions

Chaque réponse LLM est analysée par une passe "analyseur" (Claude Haiku, sortie JSON stricte) qui extrait chaque entité de type marque, classe son sentiment (positif / négatif / neutre) et identifie s'il s'agit de la marque cible ou d'un concurrent.

La classification est pilotée par le workspace. Le système remonte chaque entité qu'il trouve ; le propriétaire du workspace les promeut en *my\_brand* / *competitor* / *ignored* via l'onglet Marques. L'auto-classification reste conservatrice - les entités inconnues restent en *discovered* et sont invisibles dans les métriques tant que le propriétaire ne les a pas curées. Aucune liste de verticale / région / marque codée en dur.

Une couche Haiku-as-judge (sprint dédié) renverse les faux positifs de sentiment évidents (par exemple "non adapté pour X" mal classé en négatif alors que la formulation est factuelle). En attendant, les outils radar minimisent les signaux négatifs isolés via des buckets de sévérité conservateurs.

## 6. Extraction des citations

Quand un fournisseur d'IA expose des métadonnées de citation (OpenAI grounded mode, Gemini Vertex AI search grounding, etc.), nous persistons les URLs telles quelles. Quand ce n'est pas le cas, nous parsons le texte de la réponse à la recherche de motifs d'URL explicites. Chaque citation est annotée avec le domaine éditeur, l'URL et un extrait de contexte de 200 caractères.

Les citations alimentent plusieurs features en aval : Audit de pages (lesquelles de VOS pages sont citées), Concurrents (lesquelles des pages des RIVAUX sont citées), PR / Médias (quels domaines presse citent la marque versus les concurrents), YouTube (quels créateurs vidéo remontent), Reddit (quels threads l'IA mine).

## 7. Variance & honnêteté statistique



Les métriques en single-run sont signalées comme "confiance faible" dans l'UI. Les métriques multi-runs apparaissent avec leur intervalle de variance visible (prévu dans le sprint Variance UI - livré avec le ramp N=10).

Les comparaisons cross-scan requièrent un minimum de 7 jours entre deux scans pour éviter de mesurer le bruit des updates IA intra-semaine. Les métriques de tendance utilisent des moyennes glissantes sur une fenêtre de 30 jours.

## 8. Vie privée & posture EU AI Act

sen-ai.fr est classé **IA à risque limité** au sens du règlement européen sur l'IA (Règlement (UE) 2024/1689), applicable au 2 août 2026. Nous sommes *déployeur aval* de systèmes d'IA tiers, pas fournisseur d'IA généraliste. Les obligations de transparence s'appliquent (article 50) ; pas d'évaluation de conformité ni de marquage CE.

L'exposition aux données personnelles est minimale :

- Nous ne stockons que l'email + nom des utilisateurs ayant accès à un workspace.
- Les prompts envoyés aux IA contiennent uniquement des données de marque / domaine / topic - pas de PII des utilisateurs finaux.
- Toutes les données sont hébergées dans l'UE (Hetzner, Falkenstein, Allemagne).
- Les fournisseurs d'IA basés aux US (OpenAI, Anthropic) opèrent sous le cadre EU-US Data Privacy Framework + clauses contractuelles types (SCC).

Divulgateur complète dans notre [Politique de confidentialité](#). Les clients connectés peuvent télécharger un rapport de transparence per-scan et un modèle de DPIA au niveau de l'organisation depuis la section Conformité de l'application.

### 8.1 Changelog des sous-traitants

Historique des ajouts, retraits et changements de portée des sous-traitants depuis la publication initiale de la documentation EU AI Act. Mis à jour en parallèle du pack de conformité in-app.

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

## 9. Ce que sen-ai.fr NE fait PAS



- Nous n'entraînons, ne fine-tunons et ne contribuons à aucun modèle d'IA.
- Nous ne scrappons pas les sites concurrents à grande échelle au-delà de ce qu'un LLM a déjà cité.
- Nous n'usurpons pas d'User-Agent (pas de Googlebot spoofing, pas de fausse signature de navigateur).
- Nous n'utilisons pas de prompt injection ni aucune technique exploitant les conditions d'utilisation des fournisseurs.
- Nous n'exerçons aucune pratique interdite par l'article 5 du règlement IA européen (scoring social, identification biométrique temps réel, manipulation, etc.).

## 10. Contact

Questions conformité / méthodologie : [contact@sen-ai.fr](mailto:contact@sen-ai.fr).

Pas de DPO obligatoire à ce stade (sen-ai.fr est sous le seuil). Le contact ci-dessus traite toutes les demandes de droits des personnes concernées et répond sous 30 jours comme l'exige le RGPD.



*from data to bloom*

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

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

[Mentions légales](#)

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