



# Q-Ledger

Rapport empirique (baseline + captures B)

- Source : logs Cloudflare (edge), inférence de sessions (gap 30 min), format log-derived.
- Baseline : 2026-01-20 → 2026-02-06 (seq05–seq25, 19 captures taguées; 22 snapshots archivés).
- Captures B (gouvernance) : 2026-01-28 → 2026-02-06, 771 requêtes (scope étendu).
- Captures C (sitemaps) : 2026-01-28 → 2026-02-06, 65 requêtes.

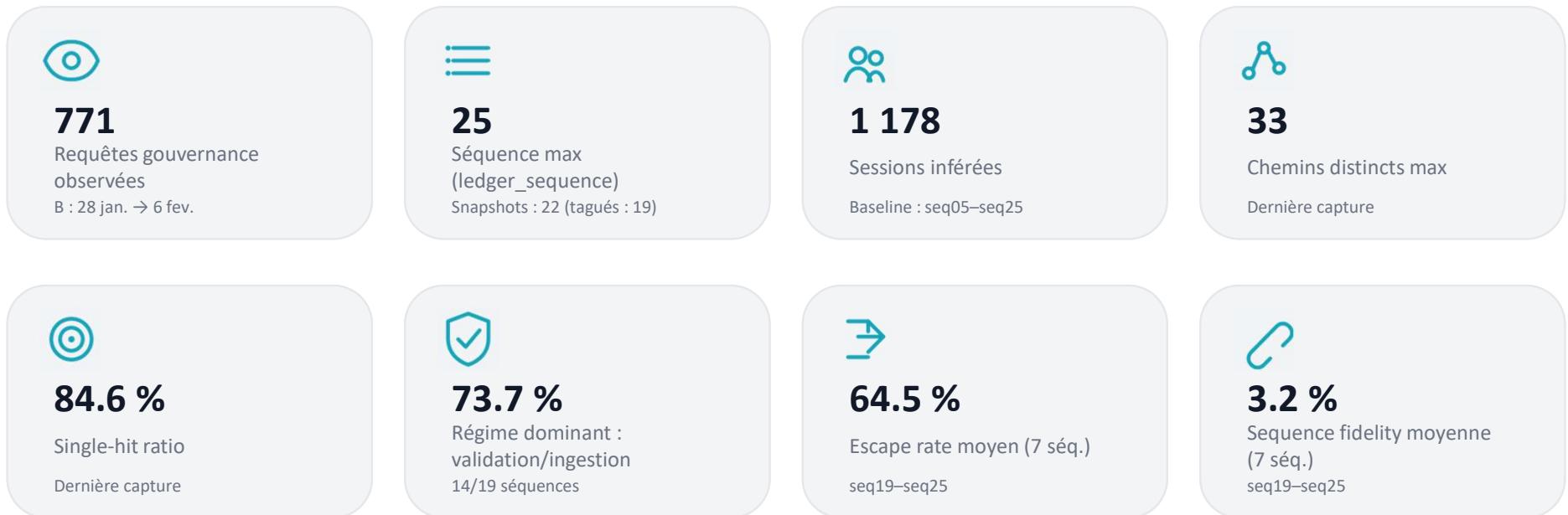
**1 178 sessions inférées (baseline)**

Observation ≠ attestation. Ce rapport décrit des accès observés, pas une conformité ni une ingestion.

Q-Ledger | [gautierdorval.com](http://gautierdorval.com) | artefacts machine-first: `./well-known/q-ledger.*` et `./well-known/q-metrics.*`

# Vue d'ensemble

Indicateurs clés (baseline + captures B)

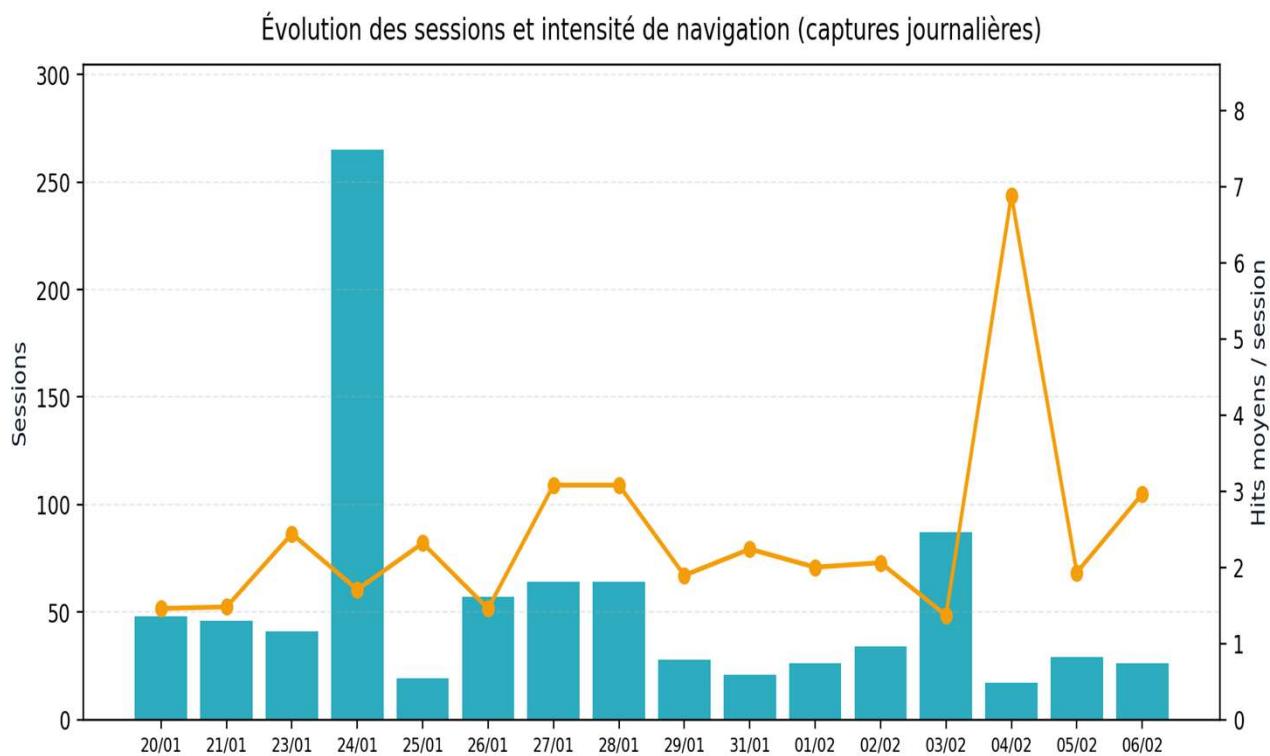


Note : les captures B/C (gouvernance + sitemaps) sont un sous-ensemble focalisé (fenêtre 28 jan.–6 fev.). La baseline couvre la continuité (seq05–seq25).

Q-Ledger | Lecture : </doctrine/q-layer/q-ledger> et </doctrine/q-layer/q-metrics>

# Évolution temporelle des sessions

Sessions (barres) et intensité de navigation (ligne)



## Classification des régimes

- 20–23/01 : validation/ingestion
- 24/01 : exploration/navigation (pic)
- 25–28/01 : validation/ingestion
- 29/01 : mixte (multi-captures)
- 31/01–03/02 : validation/ingestion
- 04/02 : exploration/navigation
- 05–06/02 : validation/ingestion

## Synthèse (séquences taguées)

- validation/ingestion : 14
- exploration/navigation : 3
- mixte : 2

Lecture : le régime est une heuristique (patrons de navigation), pas une classification d'intention.

# Agents IA et crawlers observés

Distribution des user agents (B : gouvernance)

Répartition des agents observés (B, gouvernance)



■ SemrushBot	379	(49.2%)
■ Navigateurs	211	(27.4%)
■ YandexBot	54	(7.0%)
■ MJ12bot	30	(3.9%)
■ Barkrowler	24	(3.1%)
■ Python httpx	18	(2.3%)
■ DataForSeo	12	(1.6%)
■ AhrefsBot	13	(1.7%)
■ GPTBot	5	(0.6%)
■ PerplexityBot	5	(0.6%)
■ Autres	20	(2.6%)

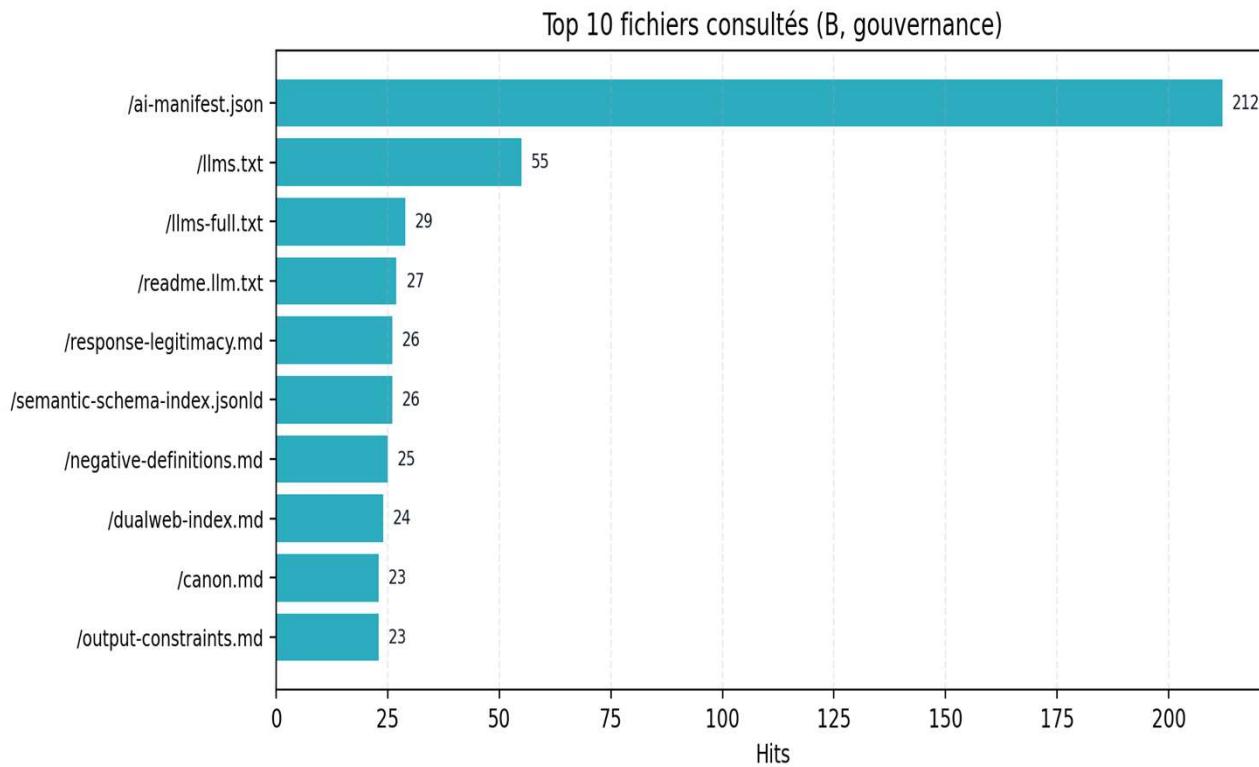
## Observations clés

- SemrushBot domine (~ 49 % des hits).
- GPTBot : 5 hits; PerplexityBot : 5 hits.
- UA de navigateurs génériques (humain ou headless).
- YandexBot et MJ12bot restent actifs (crawl SEO).
- httpx/Python observé (18 hits) : souvent instrumentation ou scripts.

Important : un user agent n'est pas une preuve d'identité (spoofing possible).  
Q-Ledger observe des accès, il n'atteste pas l'ingestion, la compréhension ou la conformité.

# Fichiers de gouvernance les plus consultés

Top fichiers (hits) et répartition par couches (B)



## Architecture par couches (hits)

- Entrypoints : 223
- Routage/découverte : 129
- Contraintes : 104
- Ontologie (incl. fallback) : 122
- Identité & canon : 45
- Q-Layer : 47
- Observabilité : 33
- Autres (index, policy, reporting, traceabilité) : 68

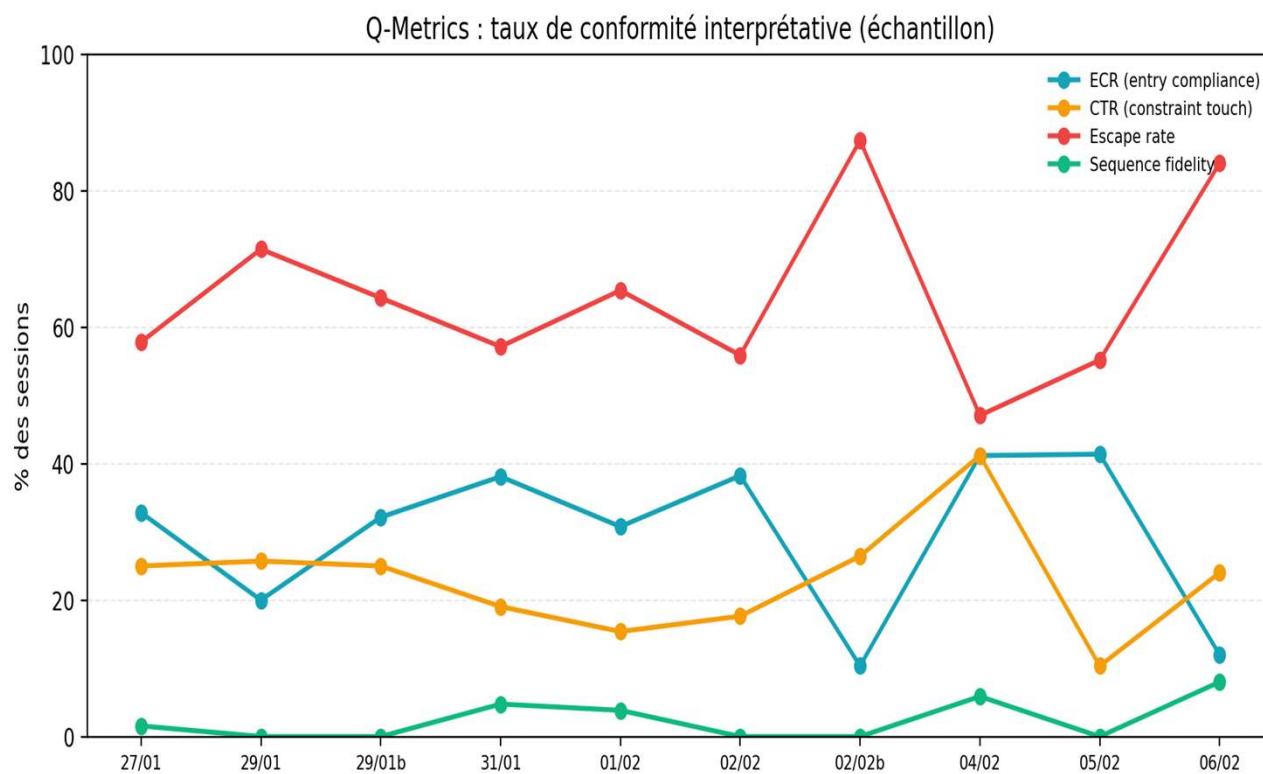
### Lecture utile :

- /ai-manifest.json = 212 hits (95 % des hits d'entrypoints)
- Déviation fréquente : entrée → sortie (escape) plutôt que parcours complet.

B : scope étendu (gouvernance). Les hits incluent SEO crawlers + navigateurs + agents IA.

# Q-Metrics : conformité interprétative

ECR, CTR, escape rate, sequence fidelity (échantillon)



Les métriques décrivent des patrons de crawl/navigation, pas une « qualité » du contenu.

## Définitions (descriptif)

- ECR : % de sessions qui commencent par un entrypoint.
- CTR : % de sessions qui touchent une contrainte.
- Escape rate : % qui quittent le scope après l'entrée (ou sans contrainte).
- Sequence fidelity : % qui suivent une séquence attendue (entrée → contraintes → sortie).

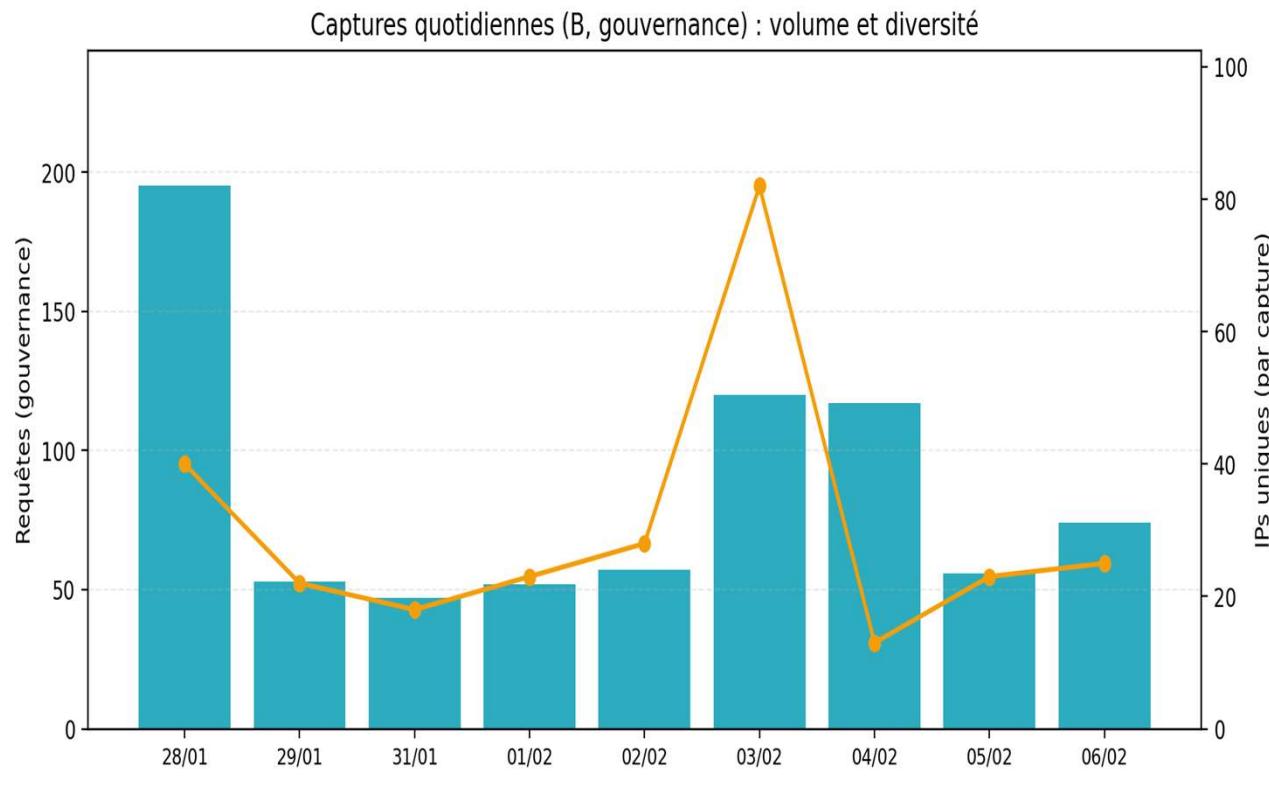
### Lecture rapide :

- Escape rate élevé (moyenne 7 séq. : 64.5 %).
- Sequence fidelity faible (moyenne 7 séq. : 3.2 %).

Avertissement : métriques non normatives (instrumentation).

# Captures quotidiennes de gouvernance

Volumes consolidés et pics (B : gouvernance, C : sitemaps)



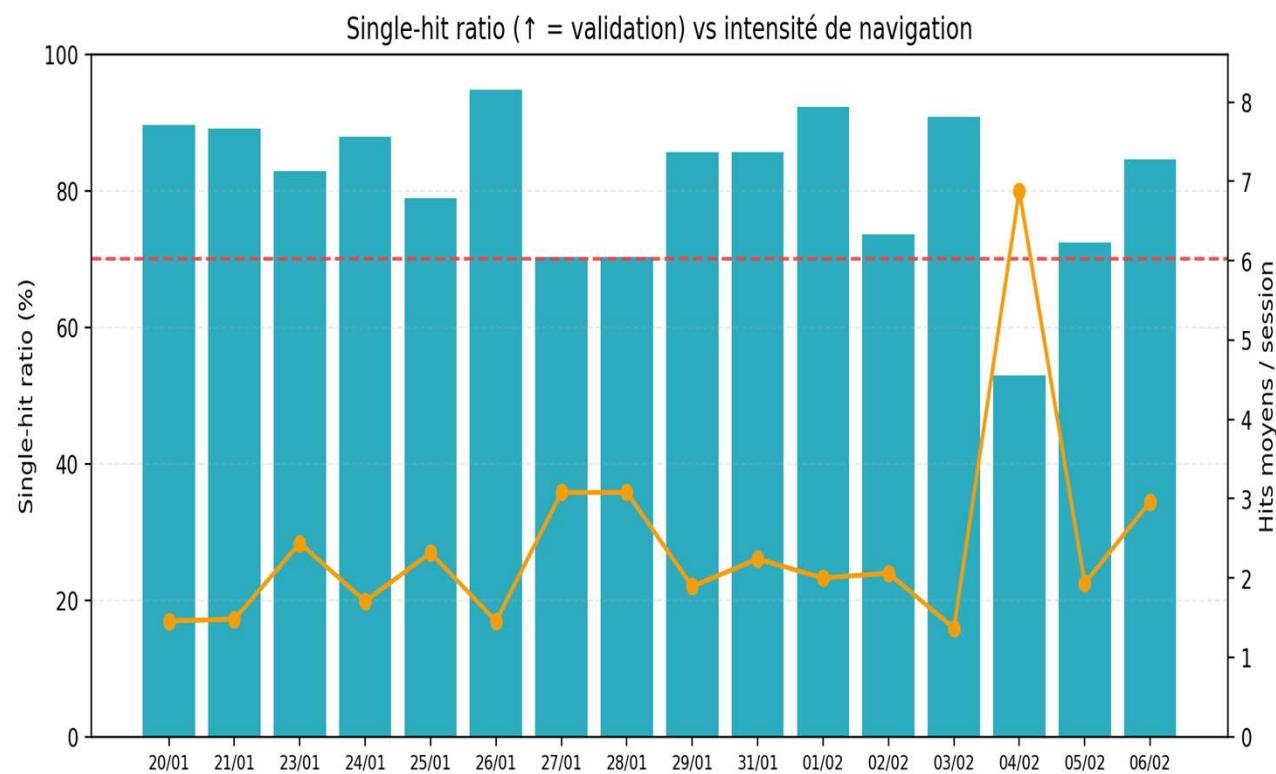
## Volumes consolidés

- Requêtes (gouvernance) : 771
- Requêtes (sitemaps) : 65
- Pic requêtes : 195 (28 jan.)
- Pic IPs uniques (par capture) : 82 (3 fev.)
- IPs distinctes (période, B) : 225
- Ratio requêtes / IP distincte : 3.4

Note : la somme des IPs uniques quotidiennes = 274 (non dédupliqué).  
Les IPs peuvent changer (CDN, NAT, rotation).

# Comportement des crawlers : single-hit vs exploration

Barres : single-hit ratio | Ligne : hits moyens / session



## Interprétation

- Single-hit élevé : validation/ingestion (entrée unique, faible profondeur).
- Single-hit bas : exploration/navigation (plusieurs hits).
- Seuil visuel : 70 % (ligne rouge).
- Exemple : 04/02 = 52.9 % et 6.88 hits/session (exploration).
- Exemple : 06/02 = 84.6 % et 2.96 hits/session (validation).

Implication : la découvrabilité ne suffit pas.  
Pour réduire l'escape, il faut rendre les séquences « naturelles » pour les crawlers (liens, conventions, sitemaps, hints).

# Constats et prochaines étapes

Ce que les données montrent (et ne montrent pas)

## Constats

- Les accès « endpoint-only » dominent : beaucoup valident, peu parcourent.
- Endpoint leader : /ai-manifest.json (95 % des hits d'entrypoints en B).
- Escape rate élevé : la gouvernance est détectée, mais rarement suivie jusqu'aux contraintes.
- Sequence fidelity très basse : les séquences attendues ne sont pas le comportement par défaut.
- Les UA « IA » sont visibles, mais non attribuables sans attestation (spoofing).

## Prochaines étapes

- Phase expérimentale A : découvrabilité passive (maintenir /.well-known/\* + monitoring).
- Durcir l'opposabilité : introduire Q-Attest (signature, ancrage, chaîne de confiance).
- Optimiser les séquences : liens internes, sitemaps ciblés, index « machine-first ».
- Améliorer l'observabilité : distinguer 200/304, cache, bots, UA, et enrichir les tags.
- Publier une page d'observations : baseline, biais (cache, NAT, robots), limites.