

Log file analysis worksheet

Use this alongside the guide. Fill it in for one investigation, one question at a time. Every recommendation names the rung of the Evidence Ladder it rests on.

1. The question

Question to resolve (one only)	
Site section / hostname	
Why it matters to the business	
Owner and date	

2. The Evidence Ladder

Rung	What proves it	Where to look	You cannot claim yet
1 Requested	Log line with verified bot identity	Access or edge logs	That the page was understood or kept
2 Served	Status, bytes, time, correct live content	Logs plus a live HTTP check	That the page is indexable
3 Indexed	URL Inspection, indexing reports	Search Console	That it ranks or earns traffic
4 Cited	Saved answers: prompt, date, platform, cited URL	Your prompt log, referrals	That a fetch caused the citation
5 Converted	Analytics events, CRM records	Analytics, CRM	That a log change caused revenue

3. Dataset checks (before any finding)

Check	Minimum evidence	Done
Log reliability	Hostname, layer (CDN or origin), date range, timezone, gaps, sampling	
Crawler identity	Source IP, claimed user agent, verification method, variant	
URL relevance	HTML vs resource, section, indexability, canonical and sitemap status	
Technical issue	HTTP code, affected URLs, repeatability, linked ticket	
Outcome	Comparable follow-up requests, indexing state, search and business metrics, kept separate	

4. Bot verification record

Run a reverse lookup on the log IP, check the hostname ends in googlebot.com, google.com or googleusercontent.com, then run a forward lookup and confirm it returns the original IP.

IP from log	Claimed bot	Reverse hostname	Forward IP	Verified? (Y/N)

5. AI crawler roles

Provider	Search	Training	User-triggered
OpenAI	OAI-SearchBot	GPTBot	ChatGPT-User
Anthropic	Claude-SearchBot	ClaudeBot	Claude-User
Perplexity	PerplexityBot	None documented	Perplexity-User
Google	Googlebot	Google-Extended (robots.txt token, never in logs)	Google user-triggered fetchers

A verified AI bot request shows a fetch. It does not show the prompt, training use or a citation. Record intended access policy per role before editing robots.txt or firewall rules.

6. URL evidence record

URL or pattern	Evidence	Likely cause	Change	Owner	Validate by

7. Report line

__ of __ priority URLs (__%) were observed in verified Googlebot HTML requests during this window, and __ were not observed in this dataset. __ of those have no discoverable internal link in the companion crawl; the other __ still need diagnosis. __ of __ verified requests (__%) returned 429 or 5xx, and __% went to URLs reviewed as low value. No change in indexing, rankings or conversions is claimed from this data.