

CRAWLER POLICY

AI Crawler Governance Playbook

A decision framework for OAI-SearchBot, GPTBot, Googlebot, Google-Extended, robots.txt, sitemaps and llms.txt.

BUILT TO BE USED

Decision frameworks

Working registers

Implementation worksheets

2026 FIELD EDITION / 12 PAGES

HOW TO USE THIS MANUAL

A working document, not a one-page checklist.

Use the explanation on each page to make a decision, then complete the adjacent register or worksheet. Assign owners and dates. Revisit the same pages after a launch, campaign or platform-policy change so the record becomes useful over time.

01

Separate discovery, snippets, training and orientation.

OUTCOME MAP

02

Make an explicit OAI-SearchBot and GPTBot decision.

OPENAI AGENTS

03

Separate Googlebot Search controls from Google-Extended.

GOOGLE AGENTS

04

Turn robots.txt into an accountable policy record.

ROBOTS REGISTER

05

Verify the response each crawler actually receives.

LIVE TEST PROTOCOL

06

Keep sitemap, links, schema and llms.txt in their lanes.

DISCOVERY STACK

07

Publish a useful guide without overselling it.

LLMS.TXT

08

Implement policy changes with rollback criteria.

RELEASE PLAN

09

Review agents, rules and outcomes as a system.

MONTHLY GOVERNANCE

10

Keep the claim attached to its source and its limit.

SOURCE NOTES + LIMITS

OUTCOME MAP

Separate discovery, snippets, training and orientation.

Before writing a rule, name the outcome. Search discovery, page indexing, snippet presentation, model training and llms.txt orientation are separate layers with different controls and owners.

DECISION PRINCIPLES

- robots.txt controls crawl access for compliant agents.
- Page directives may control indexing or snippets where supported.
- Training agents can be distinct from search agents.
- llms.txt is an optional guide, not a permission layer.

ACTION CHECK

- ☐ Name desired outcome.
- ☐ Identify supporting control.
- ☐ Assign policy owner.
- ☐ Record implementation owner.
- ☐ Set review date.

WORKING REGISTER

Capture the evidence, owner and next decision. Use one row per page, prompt, source or issue.

OUTCOME	CONTROL	OWNER

DECISION / NOTES / NEXT REVIEW

Treat the completed page as a dated observation, not a permanent truth.

OPENAI AGENTS

Make an explicit OAI-SearchBot and GPTBot decision.

OpenAI distinguishes OAI-SearchBot for search discovery from GPTBot for potential training. Decide each policy separately and test robots, CDN and firewall behavior in production.

DECISION PRINCIPLES

- Allowing search does not require the same training policy.
- A blocked crawler may still leave a title-only discovery path under documented conditions.
- noindex must be readable to supporting crawlers.
- Referral traffic can be tracked separately.

ACTION CHECK

- ☐ Decide search access.
- ☐ Decide training access.
- ☐ Write rules.
- ☐ Test edge response.
- ☐ Document approval.

WORKING REGISTER

Capture the evidence, owner and next decision. Use one row per page, prompt, source or issue.

AGENT	DECISION	EVIDENCE

DECISION / NOTES / NEXT REVIEW

Treat the completed page as a dated observation, not a permanent truth.

GOOGLE AGENTS

Separate Googlebot Search controls from Google-Extended.

Googlebot governs crawling for Google Search, including its AI features. Google-Extended controls certain Gemini training and grounding uses and does not control Search inclusion or ranking.

DECISION PRINCIPLES

- Google Search AI features use existing Search controls.
- Snippet controls are separate from crawl permission.
- Google-Extended is a product token, not an indexing directive.
- Policy should follow the business objective.

ACTION CHECK

- ☐ Review Googlebot rules.
- ☐ Review snippet controls.
- ☐ Decide Google-Extended.
- ☐ Test production.
- ☐ Record rationale.

WORKING REGISTER

Capture the evidence, owner and next decision. Use one row per page, prompt, source or issue.

TOKEN	OUTCOME	RULE

DECISION / NOTES / NEXT REVIEW

Treat the completed page as a dated observation, not a permanent truth.

ROBOTS REGISTER

Turn robots.txt into an accountable policy record.

For every user agent, record the paths affected, desired outcome, rationale, approver, implementation owner and next review. Preserve previous versions and annotate changes.

DECISION PRINCIPLES

- A copied snippet is not governance.
- Wildcards and path matching need careful testing.
- Robots is public and should not expose secret paths as a security tactic.
- Migrations require a full policy review.

ACTION CHECK

- ☐ Inventory rules.
- ☐ Map paths.
- ☐ Name approver.
- ☐ Version changes.
- ☐ Schedule review.

WORKING REGISTER

Capture the evidence, owner and next decision. Use one row per page, prompt, source or issue.

AGENT / PATH	RATIONALE	REVIEW

DECISION / NOTES / NEXT REVIEW

Treat the completed page as a dated observation, not a permanent truth.

LIVE TEST PROTOCOL

Verify the response each crawler actually receives.

Test DNS, redirects, status, robots response, page HTML, headers and edge bot controls. Record timestamp and network location. Server logs provide the strongest operational confirmation when available.

DECISION PRINCIPLES

- Source control does not prove production behavior.
- Caches can delay a corrected policy.
- Bot-management products may challenge or block requests.
- Test representative URLs, not only the homepage.

ACTION CHECK

- ☐ Fetch robots.txt.
- ☐ Fetch test URLs.
- ☐ Inspect headers.
- ☐ Review edge logs.
- ☐ Retest after cache expiry.

WORKING REGISTER

Capture the evidence, owner and next decision. Use one row per page, prompt, source or issue.

TEST	EXPECTED	ACTUAL

DECISION / NOTES / NEXT REVIEW

Treat the completed page as a dated observation, not a permanent truth.

DISCOVERY STACK

Keep sitemap, links, schema and llms.txt in their lanes.

The sitemap surfaces canonical URLs, internal links establish navigable context, visible content carries the evidence, schema describes supported facts and llms.txt provides a concise optional map. None overrides a blocked page.

DECISION PRINCIPLES

- Sitemaps should contain canonical public URLs.
- Schema must match visible content.
- llms.txt should be concise and maintained.
- Internal links remain essential for people and crawlers.

ACTION CHECK

- ☐ Validate sitemap.
- ☐ Review internal links.
- ☐ Validate schema.
- ☐ Draft llms.txt.
- ☐ Assign maintainer.

WORKING REGISTER

Capture the evidence, owner and next decision. Use one row per page, prompt, source or issue.

LAYER	JOB	OWNER

DECISION / NOTES / NEXT REVIEW

Treat the completed page as a dated observation, not a permanent truth.

LLMS.TXT

Publish a useful guide without overselling it.

Use a root-level markdown file with a short site summary, important links and optional sections. Label it as an orientation guide and keep claims consistent with the visible site. Do not call it a ranking factor or crawler permission standard.

DECISION PRINCIPLES

- The current convention is a proposal.
- Link to authoritative pages instead of copying the site.
- Do not place confidential information in a public file.
- Update it when core routes or claims change.

ACTION CHECK

- ☐ Write one-paragraph summary.
- ☐ Choose core links.
- ☐ Add research links.
- ☐ State editorial limits.
- ☐ Set review owner.

WORKING REGISTER

Capture the evidence, owner and next decision. Use one row per page, prompt, source or issue.

SECTION	LINK	WHY

DECISION / NOTES / NEXT REVIEW

Treat the completed page as a dated observation, not a permanent truth.

RELEASE PLAN

Implement policy changes with rollback criteria.

Review with product, legal, security and search owners as appropriate. Capture the current state, approve the target policy, deploy, test production and monitor crawl plus referral signals.

DECISION PRINCIPLES

- Policy decisions may have legal and commercial implications.
- One owner should coordinate the release.
- Define rollback before deployment.
- Allow for crawler revisit delays.

ACTION CHECK

- ☐ Capture current state.
- ☐ Approve target policy.
- ☐ Deploy change.
- ☐ Run live tests.
- ☐ Monitor and review.

WORKING REGISTER

Capture the evidence, owner and next decision. Use one row per page, prompt, source or issue.

STEP	OWNER	DATE

DECISION / NOTES / NEXT REVIEW

Treat the completed page as a dated observation, not a permanent truth.

MONTHLY GOVERNANCE

Review agents, rules and outcomes as a system.

Each month, compare the policy register with production responses, crawler logs, sitemap health, index state and AI referral observations. Recheck official documentation before changing rules.

DECISION PRINCIPLES

- New agents and product tokens can appear.
- A stale rule can survive multiple redesigns.
- Access signals and visibility signals belong in separate reports.
- Record decisions even when no change is made.

ACTION CHECK

☐ Review official docs.

☐ Compare live robots.

☐ Inspect logs.

☐ Check discovery files.

☐ Approve next review.

WORKING REGISTER

Capture the evidence, owner and next decision. Use one row per page, prompt, source or issue.

CHECK	FINDING	DECISION

DECISION / NOTES / NEXT REVIEW

Treat the completed page as a dated observation, not a permanent truth.

SOURCE NOTES + LIMITS

Keep the claim attached to its source and its limit.

A field manual should make the work repeatable without turning research into certainty. Record the original source, publication date, exact claim, population or platform studied, and the boundary between a documented finding and your team's interpretation.

DECISION PRINCIPLES

- Link to original documentation or research whenever it is available.
- Separate a platform requirement from a community convention or agency recommendation.
- Do not convert correlation, a case example or an illustrative score into a guarantee.
- Recheck crawler, schema and search guidance before implementation because platforms change.

ACTION CHECK

- ☐ Record the source URL and access date.
- ☐ Copy the exact claim in your own words.
- ☐ Note the methodology or platform scope.
- ☐ Write the practical implication separately.
- ☐ Set a date to revalidate the guidance.

EVIDENCE LEDGER

Use one row per claim. The evidence column should name the original document, not a secondary summary.

CLAIM	SOURCE	LIMIT

DECISION / NOTES / NEXT REVIEW

Treat the completed page as a dated observation, not a permanent truth.