

AI & LLM VISIBILITY

---

# AI Crawler Whitelist: Yang Wajib Diperbolehkan

robots.txt konfigurasi yang mempersilakan AI crawler penting akses konten, sambil memblokir yang problematik. Daftar 2026 current.

---

FORMAT

Reference

ESTIMASI BACA

5 menit

UNTUK

Technical SEO, Developer

Block semua AI crawler = block visibility di ChatGPT, Gemini, Perplexity. Tapi allow semua = data leakage ke training yang mungkin tidak kamu mau. Ini tabel current crawler dan rekomendasi policy untuk perusahaan B2B.

## Tabel AI crawler 2026

CRAWLER	OPERATOR	PURPOSE	REKOMENDASI
GPTBot	OpenAI	Training + search (ChatGPT)	Allow (untuk visibility)
ChatGPT-User	OpenAI	Real-time fetch saat user query	Allow (critical untuk citation)
OAI-SearchBot	OpenAI	Search product	Allow
ClaudeBot	Anthropic	Training + search	Allow
Claude-Web	Anthropic	Real-time fetch	Allow
anthropic-ai	Anthropic	Legacy training	Allow atau block
Google-Extended	Google Gemini	Training data control	Allow (untuk Gemini)
Googlebot	Google	Regular + AI Overviews	JANGAN PERNAH block
PerplexityBot	Perplexity	Search + citation	Allow
cohere-ai	Cohere	Training	Optional
YouBot	You.com	Search	Allow
CCBot	Common Crawl	Open dataset (trained oleh banyak)	Allow
Meta-ExternalAgent	Meta	Training	Block kalau khawatir
Bytespider	ByteDance	TikTok AI training	Block kalau tidak relevan
Amazonbot	Amazon	Alexa + shopping	Allow
Applebot-Extended	Apple	Siri/Spotlight AI	Allow

## Template robots.txt

```
# Allow all major AI crawlers
User-agent: GPTBot
Allow: /

User-agent: ChatGPT-User
Allow: /

User-agent: OAI-SearchBot
Allow: /

User-agent: ClaudeBot
Allow: /

User-agent: Claude-Web
Allow: /

User-agent: Google-Extended
Allow: /

User-agent: PerplexityBot
Allow: /

User-agent: Applebot-Extended
Allow: /

# Block crawler yang tidak kamu mau
User-agent: Bytespider
Disallow: /

User-agent: Meta-ExternalAgent
Disallow: /

# Standard rules
User-agent: *
Allow: /
Disallow: /admin/
Disallow: /private/

Sitemap: https://example.com/sitemap.xml
```

## Monitoring crawler

- Log access untuk User-Agent AI crawler.**  
Pastikan mereka benar-benar crawl. Kalau 0 hit selama 30 hari = crawler mungkin tidak honor robots.txt.
- Track citation di LLM setelah allow.**  
Test prompt bulanan. Measure frekuensi kemunculan.
- Review new AI crawler setiap kuartal.**  
Landscape berubah. Crawler baru muncul, yang lama deprecated.
- Coordinate dengan legal/compliance.**

BUTUH BANTUAN MENJALANKAN INI?

# Kami bisa pasang seluruh entity infrastructure untuk bisnismu.

Hibranwar Labs adalah agensi entity SEO untuk perusahaan B2B, manufaktur, dan penerbit yang butuh muncul di Google Knowledge Panel, AI Overviews, dan citation LLM (ChatGPT, Gemini, Perplexity). Proyek dimulai dari audit entity gratis.

LAYANAN YANG RELEVAN UNTUK TOPIK INI

**LLM Priming Pack**

llms.txt, llms-full.txt, citation-ready content, fact-density audit. Brand siap dikutip ChatGPT, Gemini, Perplexity.



## Audit entity gratis — 48 jam

Kirim satu URL brand atau nama perusahaan. Kami kembalikan laporan berisi: skor entity, gap Knowledge Graph, skema yang hilang, sumber citation LLM yang sudah/belum konsisten, dan 3 tindakan prioritas.

**Request audit gratis >**

[services.hibranwar.com/audit/](https://services.hibranwar.com/audit/)

TENTANG PENULIS

**Ibrahim Anwar** — Direktur di PT Arsindo Integrasi Pompa, PT Witanabe Integrasi Indonesia, dan PT Hibrkraft Kreasi Indonesia. Penulis 6 buku di Google Play Books. Komite Tetap KADIN Kota Bogor bidang Ekonomi Kreatif.

Lead magnet ini adalah No. 36 dari 100 dokumen gratis yang diterbitkan Hibranwar Labs untuk publik. Boleh dibagikan, dicetak, dan dikutip dengan atribusi.

KONTAK & KANAL

**Web:** [services.hibranwar.com](https://services.hibranwar.com) · [hibranwar.com](https://hibranwar.com)

**Audit gratis:** [services.hibranwar.com/audit/](https://services.hibranwar.com/audit/)

**Email:** [hibr.anwar@gmail.com](mailto:hibr.anwar@gmail.com)

**LinkedIn:** [linkedin.com/in/hibranwar](https://linkedin.com/in/hibranwar)