

OVERALL SCORE

80 /100

AI-Readable

Executive summary

This site is mostly readable by AI tools, with a few improvements still recommended. Key strengths include AI guidance file and plain-text page access, while homepage access and crawler policy need attention. Recommended next step: review AI crawler Disallow rules and keep only the paths that should be excluded from AI crawler access; serve a non-empty HTML homepage with a canonical link tag.

Recommended next step

1. Review AI crawler Disallow rules and keep only the paths that should be excluded from AI crawler access; serve a non-empty HTML homepage with a canonical link tag.

2. Fix robots.txt syntax issues so each rule uses Field: value format, directives appear under a User-agent, and Sitemap entries use absolute URLs.

3. Use exactly one h1 element and move secondary section titles to h2-h6. Add missing semantic elements: main.

Signal breakdown

Crawlability Review AI crawler Disallow rules and keep only the paths that should be excluded from AI crawler access; serve a non-empty HTML homepage with a canonical link tag.	Warn 10/20
Robots.txt Fix robots.txt syntax issues so each rule uses Field: value format, directives appear under a User-agent, and Sitemap entries use absolute URLs.	Warn 7.5/15
llms.txt The llms.txt file was found and includes the expected text, length, heading, and URL signals.	Pass 15/15

Sitemap	Pass	10/10
The sitemap.xml file is valid and contains URL entries.		
Markdown support	Pass	15/15
The homepage returns markdown content when requested with Accept: text/markdown, giving AI systems a cleaner text representation with less navigation chrome to parse.		
Semantic HTML	Warn	7.1/10
Use exactly one h1 element and move secondary section titles to h2-h6. Add missing semantic elements: main.		
Structured data	Pass	10/10
Valid JSON-LD structured data was found with core Organization or WebSite schema.org types.		
Content signals	Pass	5/5
Consider adding Content-Signal HTTP header, AI-specific head meta tags, robots noai/noimageai directive so AI systems can consistently discover content usage preferences across robots.txt, HTTP headers, and HTML metadata.		

Suggested fixes

llms.txt

MARKDOWN

getclaw: Your AI workforce, live in 5 minutes

> Hire AI workers with chat, inbox, browser, memory, files, logs, and redeploy controls. Launch in 5 minutes from \$20/month with up to 60% cheaper LLM routing.

This llms.txt file summarizes the public, canonical resources that AI assistants and crawlers should use to understand this site.

Site Overview

- Canonical URL: <https://getclaw.sh/>
- Site type: organization
- Recommended summary: Hire AI workers with chat, inbox, browser, memory, files, logs, and redeploy controls. Launch in 5 minutes from \$20/month with up to 60% cheaper LLM routing.

Core URLs

- [Homepage](<https://getclaw.sh/>): Primary public entry point and canonical site overview.
- [by deployed.engineering](<https://deployed.engineering/>): Important public page discovered from the homepage navigation.
- [BlogBlog](<https://getclaw.sh/blog>): Editorial updates, announcements, and current context.
- [ChangelogChangelog](<https://getclaw.sh/changelog>): Editorial updates, announcements, and current context.

robots.txt additions

TXT

```
# robots.txt additions
# Copy these blocks into the existing robots.txt file. Keep current rules
unless a note calls out a conflicting Disallow.

# AI crawler access
# Add explicit Allow rules for blocked AI crawlers; remove or narrow
conflicting Disallow rules if your crawler target requires precedence.
User-agent: GPTBot
Allow: /

User-agent: ChatGPT-User
Allow: /

User-agent: ClaudeBot
Allow: /

User-agent: Claude-Web
Allow: /

User-agent: PerplexityBot
Allow: /

User-agent: Google-Extended
Allow: /

# Crawl-rate guidance
Continued in the full scan report...
```

schema.json

JSON

```
{
  "@context": "https://schema.org",
  "@graph": [
    {
      "@type": "Organization",
      "@id": "https://getclaw.sh/#organization",
      "name": "getclaw: Your AI workforce, live in 5 minutes",
      "description": "Hire AI workers with chat, inbox, browser, memory, files,
logs, and redeploy controls. Launch in 5 minutes from $20/month with up to
60% cheaper LLM routing.",
      "url": "https://getclaw.sh/",
      "logo": "https://getclaw.sh/_next/image?url=%2Flogo.png&w=96&q=75"
    },
    {
      "@type": "WebSite",
      "@id": "https://getclaw.sh/#website",
      "name": "getclaw: Your AI workforce, live in 5 minutes",
      "description": "Hire AI workers with chat, inbox, browser, memory, files,
logs, and redeploy controls. Launch in 5 minutes from $20/month with up to
60% cheaper LLM routing.",
      "url": "https://getclaw.sh/",
      "publisher": {
        "@id": "https://getclaw.sh/#organization"
      },
      "inLanguage": "en"
    }
  ]
}
Continued in the full scan report...
```

Content-Signal

TXT

Content-Signal recommendations

Use these directives to make AI-use preferences explicit for compliant crawlers and AI systems. They are advisory signals, so keep them aligned with robots.txt, terms, and access controls.

Recommended values

- ai-train=no: AI model training, fine-tuning, and dataset creation.
- search=yes: AI search indexing, snippets, and discovery.
- ai-input=yes: AI answer grounding, retrieval, and generated-response context.

HTTP response header

```
'''http
Content-Signal: ai-train=no, search=yes, ai-input=yes
Content-Usage: train-ai=n
'''
```

Best for site-wide or route-specific policies because the signal travels with every response, including pages that AI systems fetch directly.

HTML meta tag alternatives

Add these inside the document '`<head>`' when server headers are not

continued in the full scan report...

Full report

<https://llmscan.dev/scan/cJJP1yJ7xJcPk7kcyZQK2>