

OVERALL SCORE

37 /100 Poor

Executive summary

This site is difficult for AI tools to read right now. Key strengths include AI guidance file and structured data, while homepage access and crawler policy need attention. Recommended next step: remove AI crawler Disallow: / rules or replace them with narrower path-level restrictions for private content only.

Recommended next step

- 1. Remove AI crawler Disallow: / rules or replace them with narrower path-level restrictions for private content only.
- 2. Remove AI crawler Disallow: / rules or add narrower Allow/Disallow rules if AI crawlers should be able to discover public content.
- 3. Add content negotiation for Accept: text/markdown on the homepage and return a markdown representation with Content-Type: text/markdown. Keep the HTML response for regular browser requests.

Signal breakdown

Crawlability Remove AI crawler Disallow: / rules or replace them with narrower path-level restrictions for private content only.	Fail	0/20
Robots.txt Remove AI crawler Disallow: / rules or add narrower Allow/Disallow rules if AI crawlers should be able to discover public content.	Fail	0/15
llms.txt The llms.txt file was found and includes the expected text, length, heading, and URL signals.	Pass	15/15

Sitemap	Fail	0/10
Regenerate sitemap.xml with valid XML and a sitemap <urlset> or <sitemapindex> root element.		
Markdown support	Fail	0/15
Add content negotiation for Accept: text/markdown on the homepage and return a markdown representation with Content-Type: text/markdown. Keep the HTML response for regular browser requests.		
Semantic HTML	Warn	7.1/10
Shorten the meta description to 160 characters or fewer. Add missing semantic elements: article.		
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

Weaveflow - build and run software from any browser

> Create a new app or connect one you already have. Ask Claude to build, fix and publish it from any browser. The agent runs on your own server, so your laptop can stay closed.

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://weaveflow.io/>
- Site type: organization
- Recommended summary: Create a new app or connect one you already have. Ask Claude to build, fix and publish it from any browser. The agent runs on your own server, so your laptop can stay closed.

Core URLs

- [Homepage](<https://weaveflow.io/>): Primary public entry point and canonical site overview.
- [docs](<https://weaveflow.io/docs>): Documentation and implementation guidance for the site.
- [sign in](<https://app.weaveflow.io/login>): Important public page discovered from the homepage navigation.
- [Start a project](<https://app.weaveflow.io/signup>): Important public page

Continued in the full scan report...

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: ClaudeBot

Allow: /

User-agent: Google-Extended

Allow: /

Crawl-rate guidance

Add Crawl-delay under a User-agent block when the site needs gentler crawler pacing.

User-agent: *

Crawl-delay: 10

schema.json

```
JSON
{
  "@context": "https://schema.org",
  "@graph": [
    {
      "@type": "Organization",
      "@id": "https://weaveflow.io/#organization",
      "name": "Weaveflow - build and run software from any browser",
      "description": "Create a new app or connect one you already have. Ask Claude to build, fix and publish it from any browser. The agent runs on your own server, so your laptop can stay closed.",
      "url": "https://weaveflow.io/",
      "logo": "https://weaveflow.io/apple-touch-icon.png"
    },
    {
      "@type": "WebSite",
      "@id": "https://weaveflow.io/#website",
      "name": "Weaveflow - build and run software from any browser",
      "description": "Create a new app or connect one you already have. Ask Claude to build, fix and publish it from any browser. The agent runs on your own server, so your laptop can stay closed.",
      "url": "https://weaveflow.io/",
      "publisher": {
        "@id": "https://weaveflow.io/#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/Sb9QoeWz282cHJajQHWW0>