

OVERALL SCORE

56 /100

Needs Work

Executive summary

This site has a useful foundation, but important gaps still limit AI readability. Key strengths include plain-text page access and structured data, while AI guidance file and sitemap need attention. Recommended next step: publish an AI guidance file as text or markdown with more than 200 characters, markdown headings, and at least one absolute URL.

Recommended next step

- Publish /llms.txt as text or markdown with more than 200 characters, markdown headings, and at least one absolute URL.
- 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.
- Publish a valid XML sitemap at /sitemap.xml and reference it from robots.txt so crawlers and AI systems can discover important URLs.

Signal breakdown

CrawlabilityWarn10/20

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.

Robots.txtWarn7.5/15

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

llms.txtFail0/15

Publish /llms.txt as text or markdown with more than 200 characters, markdown headings, and at least one absolute URL.

Sitemap	Fail	0/10
Publish a valid XML sitemap at /sitemap.xml and reference it from robots.txt so crawlers and AI systems can discover important URLs.		
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	8.6/10
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

Affonso - Affiliate Software for SaaS

> Launch your SaaS affiliate program in under 15 minutes. Starting at 15/mo
- no transaction fees, fraud detection, managed payouts in 190+ currencies.

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://affonso.io/>
- Site type: web site
- Recommended summary: Launch your SaaS affiliate program in under 15 minutes. Starting at 15/mo - no transaction fees, fraud detection, managed payouts in 190+ currencies.

Core URLs

- [Homepage](<https://affonso.io/>): Primary public entry point and canonical site overview.
 - [Pricing](<https://affonso.io/pricing>): Pricing, plans, and packaging information.
 - [For Affiliates](<https://marketplace.affonso.io/>): Important public page discovered from the homepage navigation.
 - [Login](<https://affonso.io/auth/login>): Important public page discovered from the homepage navigation.
- 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: 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://affonso.io/#organization",
      "name": "Affonso - Affiliate Software for SaaS",
      "description": "Launch your SaaS affiliate program in under 15 minutes.
Starting at 15/mo - no transaction fees, fraud detection, managed payouts in
190+ currencies.",
      "url": "https://affonso.io/",
      "logo": "https://files.affonso.io/img/supstarter_logo.svg",
      "sameAs": [
        "https://x.com/useAffonso",
        "https://www.linkedin.com/company/affiliate-software/"
      ]
    },
    {
      "@type": "WebSite",
      "@id": "https://affonso.io/#website",
      "name": "Affonso - Affiliate Software for SaaS",
      "description": "Launch your SaaS affiliate program in under 15 minutes.
Starting at 15/mo - no transaction fees, fraud detection, managed payouts in
190+ currencies.",
      "url": "https://affonso.io/",
      "publisher": {
        "name": "Affonso - Affiliate Software for SaaS",
        "url": "https://affonso.io/"
      }
    }
  ]
}
```

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://www.llmscan.dev/scan/NhNv0WPSOafVlhx8cjL3U>