

OVERALL SCORE

41 /100

Needs Work

Executive summary

This site has a useful foundation, but important gaps still limit AI readability. Key strengths include sitemap, while plain-text page access and structured data need attention. Recommended next step: 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.

Recommended next step

1.

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.
2.

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.
3.

Add JSON-LD structured data with Organization or WebSite schema so AI systems can identify the site owner or website entity.

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

Review AI crawler Disallow rules and keep only the paths that should be excluded from AI crawler access.

llms.txtWarn7.5/15

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

Sitemap	Pass	10/10
The sitemap.xml file is valid and contains URL entries.		
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	5.7/10
Add exactly one h1 element that describes the page topic. Add missing semantic elements: main, article, nav, footer. Add visible body copy after script and style removal so the page has at least 200 words.		
Structured data	Fail	0/10
Add JSON-LD structured data with Organization or WebSite schema so AI systems can identify the site owner or website entity.		
Content signals	Fail	0/5
Add the standard directive 'Content-Signal: ai-train=no, search=yes, ai-input=yes' to robots.txt, HTML metadata, or HTTP headers so AI systems can discover content usage preferences.		

Suggested fixes

Fix Structured data

HTML

```
<script type="application/ld+json">
{
  "@context": "https://schema.org",
  "@graph": [
    {
      "@type": "Organization",
      "@id": "https://annotate.foo/#organization",
      "name": "Annotate - Highlight & Annotate the Web",
      "description": "Annotate is a browser extension that lets you highlight
text, add notes, and build a personal knowledge base from anything you read
on the web.",
      "url": "https://annotate.foo/",
      "logo": "https://annotate.foo/favicon.svg"
    },
    {
      "@type": "WebSite",
      "@id": "https://annotate.foo/#website",
      "name": "Annotate - Highlight & Annotate the Web",
      "description": "Annotate is a browser extension that lets you highlight
text, add notes, and build a personal knowledge base from anything you read
on the web.",
      "url": "https://annotate.foo/",
      "publisher": {
        "@id": "https://annotate.foo/#organization"
      }
    },
    {
      "@type": "Text",
      "@id": "https://annotate.foo/#full-scan-report",
      "text": "Continued in the full scan report..."
    }
  ]
}
```

Fix Content signals

HTML

```
<meta http-equiv="Content-Signal" content="ai-train=no, search=yes,
ai-input=yes" />
<meta name="content-signal" content="ai-train=no, search=yes, ai-input=yes"
/>
```

Full report

<https://llmscan.dev/scan/v01OFhI1wOdzwg8e6tODE>