

OVERALL SCORE

38 /100

Poor

Executive summary

This site is difficult for AI tools to read right now. Key strengths include AI guidance file, while plain-text page access and sitemap need attention. Recommended next step: make the homepage respond successfully to requests with Accept: text/markdown, either by serving markdown directly or by adding a markdown variant URL with correct content negotiation.

Recommended next step

1. Make the homepage respond successfully to requests with Accept: text/markdown, either by serving markdown directly or by adding a markdown variant URL with correct content negotiation.
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. 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

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 Review AI crawler Disallow rules and keep only the paths that should be excluded from AI crawler access.	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	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	Fail	0/15
Make the homepage respond successfully to requests with Accept: text/markdown, either by serving markdown directly or by adding a markdown variant URL with correct content negotiation.		
Semantic HTML	Warn	5.7/10
Shorten the meta description to 160 characters or fewer. Use exactly one h1 element and move secondary section titles to h2-h6. Add missing semantic elements: article.		
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://github.com/#organization",
      "name": "GitHub Change is constant. GitHub keeps you ahead. GitHub",
      "description": "Join the world’s most widely adopted, AI-powered developer platform where millions of developers, businesses, and the largest open source community build software that advances humanity.",
      "url": "https://github.com/",
      "logo": "https://github.githubassets.com/favicons/favicon.png",
      "sameAs": [
        "https://github.com/#start-of-content",
        "https://github.com/",
        "https://github.com/login",
        "https://github.com/features/copilot",
        "https://github.com/features/spark",
        "https://github.com/features/models",
        "https://github.com/mcp",
        "https://github.com/features/actions",
        "https://github.com/features/codespaces",
        "https://github.com/features/issues",
        "https://github.com/features/code-review",
        "https://github.com/security/advanced-security",
        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://www.llmscan.dev/scan/_3Mts9ICD1ROyjNr5rjLt