

# LLM Scan

PUBLIC AI VISIBILITY REPORT

# moxiedocs.com

Scanned Jun 23, 2026, 08:33 UTC

OVERALL SCORE

# 83 /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.

## Signal breakdown

### Crawlability

Warn 10/20

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.txt

Warn 7.5/15

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.txt

Pass 15/15

The llms.txt file was found and includes the expected text, length, heading, and URL signals.

### 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

Pass 10/10

The homepage has a valid title, meta description, heading structure, semantic landmarks, substantial body copy, and descriptive anchor text.

## 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

```
# Moxie Docs | Automated codebase documentation for GitHub
```

```
> Moxie Docs indexes GitHub repositories, generates searchable codebase docs, catches doc drift on merge, and opens Friday Cleanup PRs with MCP AI context.
```

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://moxiedocs.com/>
- Site type: organization
- Recommended summary: Moxie Docs indexes GitHub repositories, generates searchable codebase docs, catches doc drift on merge, and opens Friday Cleanup PRs with MCP AI context.

#### ## Core URLs

- [Homepage](<https://moxiedocs.com/>): Primary public entry point and canonical site overview.
  - [Demo](<https://moxiedocs.com/demo>): Important public page discovered from the homepage navigation.
  - [Workflow](<https://moxiedocs.com/workflow>): Important public page discovered from the homepage navigation.
  - [Solutions](<https://moxiedocs.com/for>): Important public page discovered
- 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://moxiedocs.com/#organization",
      "name": "Moxie Docs | Automated codebase documentation for GitHub",
      "description": "Moxie Docs indexes GitHub repositories, generates searchable codebase docs, catches doc drift on merge, and opens Friday Cleanup PRs with MCP AI context.",
      "url": "https://moxiedocs.com/",
      "logo": "https://moxiedocs.com/apple-touch-icon.png",
      "sameAs": [
        "https://www.linkedin.com/company/moxie-docs",
        "https://x.com/Moxie_Docs"
      ]
    },
    {
      "@type": "WebSite",
      "@id": "https://moxiedocs.com/#website",
      "name": "Moxie Docs | Automated codebase documentation for GitHub",
      "description": "Moxie Docs indexes GitHub repositories, generates searchable codebase docs, catches doc drift on merge, and opens Friday Cleanup PRs with MCP AI context.",
      "url": "https://moxiedocs.com/",
      "publisher": {
        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/\\_2YzuuFqTy4dTGjITUAne](https://www.llmscan.dev/scan/_2YzuuFqTy4dTGjITUAne)