

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 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 Fix robots.txt syntax issues so each rule uses Field: value format, directives appear under a User-agent, and Sitemap entries use absolute URLs.	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 The sitemap.xml file is valid and contains URL entries.	Pass 10/10

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 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
INJECT.md - prompt-injection audit for LLM apps
> Your AI reads the web, your email, your docs - and anything it reads can tell it what to do. Run 46 real injection payloads and see what gets through.
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://inject.md/ - Site type: organization - Recommended summary: Your AI reads the web, your email, your docs - and anything it reads can tell it what to do. Run 46 real injection payloads and see what gets through.
Core URLs
- [Homepage](https://inject.md/): Primary public entry point and canonical site overview.
Content Recommendations
- Add one-sentence descriptions for product, service, documentation, pricing, support, blog, and policy pages that should be easy for AI systems
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://inject.md/#organization",
      "name": "INJECT.md - prompt-injection audit for LLM apps",
      "description": "Your AI reads the web, your email, your docs - and anything
it reads can tell it what to do. Run 46 real injection payloads and see what
gets through.",
      "url": "https://inject.md/"
    },
    {
      "@type": "WebSite",
      "@id": "https://inject.md/#website",
      "name": "INJECT.md - prompt-injection audit for LLM apps",
      "description": "Your AI reads the web, your email, your docs - and anything
it reads can tell it what to do. Run 46 real injection payloads and see what
gets through.",
      "url": "https://inject.md/",
      "publisher": {
        "@id": "https://inject.md/#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/Qdp3PEM5A7IXCQTpyrzP0>