

OVERALL SCORE

34 /100

Poor

Executive summary

This site is difficult for AI tools to read right now. Key strengths include sitemap and structured data, while homepage access and crawler policy need attention. Recommended next step: remove AI crawler Disallow: / rules or replace them with narrower path-level restrictions for private content only.

Recommended next step

- 1. Remove AI crawler Disallow: / rules or replace them with narrower path-level restrictions for private content only.
- 2. Remove AI crawler Disallow: / rules or add narrower Allow/Disallow rules if AI crawlers should be able to discover public content.
- 3. Publish /llms.txt as text or markdown with more than 200 characters, markdown headings, and at least one absolute URL.

Signal breakdown

Crawlability

Fail 0/20

Remove AI crawler Disallow: / rules or replace them with narrower path-level restrictions for private content only.

Robots.txt

Fail 0/15

Remove AI crawler Disallow: / rules or add narrower Allow/Disallow rules if AI crawlers should be able to discover public content.

llms.txt

Fail 0/15

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

|                                                                                                                                                                                                                               |      |        |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|--------|
| <b>Sitemap</b>                                                                                                                                                                                                                | Pass | 10/10  |
| The sitemap.xml file is valid and contains URL entries.                                                                                                                                                                       |      |        |
| <b>Markdown support</b>                                                                                                                                                                                                       | 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.                                      |      |        |
| <b>Semantic HTML</b>                                                                                                                                                                                                          | Warn | 8.6/10 |
| Shorten the meta description to 160 characters or fewer.                                                                                                                                                                      |      |        |
| <b>Structured data</b>                                                                                                                                                                                                        | Pass | 10/10  |
| Valid JSON-LD structured data was found with core Organization or WebSite schema.org types.                                                                                                                                   |      |        |
| <b>Content signals</b>                                                                                                                                                                                                        | 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

# BookBoy - Turn PDFs into Audiobooks on Mac

> Turn the text PDFs you own into audiobooks on your Mac. Choose a voice, listen page by page without waiting for a whole book, and keep your generated audio on your Mac.

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://www.bookboy.xyz/>
- Site type: organization
- Recommended summary: Turn the text PDFs you own into audiobooks on your Mac. Choose a voice, listen page by page without waiting for a whole book, and keep your generated audio on your Mac.

### ## Core URLs

- [Homepage](<https://www.bookboy.xyz/>): Primary public entry point and canonical site overview.
- [Guides](<https://www.bookboy.xyz/guides>): Documentation and implementation guidance for the site.
- [Compare](<https://www.bookboy.xyz/compare>): Important public page discovered from the homepage navigation.
- [Feedback](<https://www.bookboy.xyz/feedback>): Important public page

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://www.bookboy.xyz/#organization",
"name": "BookBoy - Turn PDFs into Audiobooks on Mac",
"description": "Turn the text PDFs you own into audiobooks on your Mac.
Choose a voice, listen page by page without waiting for a whole book, and
keep your generated audio on your Mac.",
"url": "https://www.bookboy.xyz/",
"logo": "https://www.bookboy.xyz/bookboy.png",
"sameAs": [
"https://x.com/0xAdilHusain"
]
},
{
"@type": "WebSite",
"@id": "https://www.bookboy.xyz/#website",
"name": "BookBoy - Turn PDFs into Audiobooks on Mac",
"description": "Turn the text PDFs you own into audiobooks on your Mac.
Choose a voice, listen page by page without waiting for a whole book, and
keep your generated audio on your Mac.",
"url": "https://www.bookboy.xyz/",
"publisher": {
"@id": "https://www.bookboy.xyz/#organization"
}
}
]
}
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/Vx2SntASNOaG2f7ozvxDa>