

OVERALL SCORE

39

/100

Poor

### Executive summary

This site is difficult for AI tools to read right now. Key strengths include AI guidance file 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. 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.

### Signal breakdown

|                                                                                                                                                                     |      |       |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|-------|
| <div>Crawlability</div> <div>Remove AI crawler Disallow: / rules or replace them with narrower path-level restrictions for private content only.</div>              | Fail | 0/20  |
| <div>Robots.txt</div> <div>Remove AI crawler Disallow: / rules or add narrower Allow/Disallow rules if AI crawlers should be able to discover public content.</div> | Fail | 0/15  |
| <div>llms.txt</div> <div>The llms.txt file was found and includes the expected text, length, heading, and URL signals.</div>                                        | Pass | 15/15 |

|                                                                                                                                                                                                                               |      |        |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|--------|
| <b>Sitemap</b>                                                                                                                                                                                                                | 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.                                                                                          |      |        |
| <b>Markdown support</b>                                                                                                                                                                                                       | 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.                                 |      |        |
| <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

# Yaps - The local AI notes app you talk to

> Hold a key and speak. Clean text lands in any app, and every word becomes a searchable Markdown note. Ask your notes and hear them back. 100% on-device, private, free to start.

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.yaps.ai/>
- Site type: organization
- Recommended summary: Hold a key and speak. Clean text lands in any app, and every word becomes a searchable Markdown note. Ask your notes and hear them back. 100% on-device, private, free to start.

### ## Core URLs

- [Homepage](<https://www.yaps.ai/>): Primary public entry point and canonical site overview.
- [Dictation](<https://www.yaps.ai/features/dictation>): Important public page discovered from the homepage navigation.
- [Text to speech](<https://www.yaps.ai/features/text-to-speech>): Important public page discovered from the homepage navigation.
- [Voice notes](<https://www.yaps.ai/features/voice-notes>): Important public

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.yaps.ai/#organization",
"name": "Yaps - The local AI notes app you talk to",
"description": "Hold a key and speak. Clean text lands in any app, and every
word becomes a searchable Markdown note. Ask your notes and hear them back.
100% on-device, private, free to start.",
"url": "https://www.yaps.ai/",
"logo": "https://www.yaps.ai/apple-touch-icon.png",
"sameAs": [
"https://github.com/richawo/yaps-releases/releases/latest"
]
},
{
"@type": "WebSite",
"@id": "https://www.yaps.ai/#website",
"name": "Yaps - The local AI notes app you talk to",
"description": "Hold a key and speak. Clean text lands in any app, and every
word becomes a searchable Markdown note. Ask your notes and hear them back.
100% on-device, private, free to start.",
"url": "https://www.yaps.ai/",
"publisher": {
"@id": "https://www.yaps.ai/#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/QtJ6ITIPenKI01a1G9QXr>