

OVERALL SCORE

50 /100

Needs Work

Executive summary

This site has a useful foundation, but important gaps still limit AI readability. Key strengths include sitemap and structured data, while plain-text page access and page structure need attention. Recommended next step: 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.

Recommended next step

- 1. 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.
- 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. Shorten the meta description to 160 characters or fewer. Add exactly one h1 element that describes the page topic. Add missing semantic elements: main, article, nav, footer. Add visible body copy after script and style removal so the page has at least 200 words.

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.

|                                                                                                                                                                                                                                           |      |        |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|--------|
| <b>llms.txt</b>                                                                                                                                                                                                                           | Warn | 7.5/15 |
| 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   |
| 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>                                                                                                                                                                                                                      | Fail | 0/10   |
| Shorten the meta description to 160 characters or fewer. Add exactly one h1 element that describes the page topic. Add missing semantic elements: main, article, nav, footer. Add visible body copy after script and style removal so the |      |        |
| <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 robots noai/noimageai directive, Content-Signal robots.txt directive so AI systems can consistently discover content usage preferences across robots.txt, HTTP headers, and HTML metadata.                                |      |        |

# Suggested fixes

## llms.txt

### MARKDOWN

# SignalShield: Investor research reports built around how you think

> SignalShield is a Chrome extension and research platform that analyzes financial content as you browse and generates investor reports tailored to how you evaluate stocks.

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://signalshield.io/>
- Site type: organization
- Recommended summary: SignalShield is a Chrome extension and research platform that analyzes financial content as you browse and generates investor reports tailored to how you evaluate stocks.

### ## Core URLs

- [Homepage](<https://signalshield.io/>): 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://signalshield.io/#organization",
"name": "SignalShield: Investor research reports built around how you
think",
"description": "SignalShield is a Chrome extension and research platform
that analyzes financial content as you browse and generates investor reports
tailored to how you evaluate stocks.",
"url": "https://signalshield.io/",
"logo":
"https://signalshield.io/favicon.ico?favicon.1660lyarqwng.ico?dpl=dpl_B8WUE
gfkpjX8qidw9iL3ZYF1Ms6k"
},
{
"@type": "WebSite",
"@id": "https://signalshield.io/#website",
"name": "SignalShield: Investor research reports built around how you
think",
"description": "SignalShield is a Chrome extension and research platform
that analyzes financial content as you browse and generates investor reports
tailored to how you evaluate stocks.",
"url": "https://signalshield.io/",
"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://llmscan.dev/scan/RtLrIHIBpuXhZRuoWOHgk>