LLM Scan

PUBLIC AI VISIBILITY REPORT

scoutcrm.io

Scanned Jul 31, 2026, 11:01 UTC

OVERALL SCORE

49 /100

Needs Work

Executive summary

This site has a useful foundation, but important gaps still limit AI readability. Key strengths include AI guidance file and sitemap, 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

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

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

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.

Semantic HTML

Warn 8.6/10

Add missing semantic elements: article.

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

```
# Scout The Hub that runs itself

> An agentic workspace for B2B teams. CRM, docs, calendar, tasks, and
integrations, all operated within a single interface

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://scoutcrm.io/
- Site type: organization
- Recommended summary: An agentic workspace for B2B teams. CRM, docs,
calendar, tasks, and integrations, all operated within a single interface

## Core URLs

- [Homepage](https://scoutcrm.io/): Primary public entry point and canonical
site overview.
- [Changelog](https://scoutcrm.io/changelog/): Editorial updates,
announcements, and current context.
- [Log in](https://scoutcrm.io/login/): Important public page discovered
from the homepage navigation.
- [Try Scout](https://scoutcrm.io/waitlist/): Important public page
discovered from the homepage navigation.
```

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://scoutcrm.io/#organization",
"name": "Scout The Hub that runs itself",
"description": "An agentic workspace for B2B teams. CRM, docs, calendar,
tasks, and integrations, all operated within a single interface",
"url": "https://scoutcrm.io/",
"logo": "https://scoutcrm.io/icon.svg"
},
{
"@type": "WebSite",
"@id": "https://scoutcrm.io/#website",
"name": "Scout The Hub that runs itself",
"description": "An agentic workspace for B2B teams. CRM, docs, calendar,
tasks, and integrations, all operated within a single interface",
"url": "https://scoutcrm.io/",
"publisher": {
"@id": "https://scoutcrm.io/#organization"
},
"inLanguage": "en"
}
]
}
```

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/P9i8hIQPre-B11RSWzK1V