

OVERALL SCORE

38 /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. 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	Warn	7.5/15
Publish /llms.txt as text or markdown with more than 200 characters, markdown headings, and at least one absolute URL.		

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	5.7/10
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.		
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

ReceiptsMRR - The database of verified Shopify revenues

> Connect your Shopify. We pull the orders. The leaderboard you can't fake. Real ecom revenue from real operators - ranked, verified, and live.

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://receiptsmrr.com/>
- Site type: organization
- Recommended summary: Connect your Shopify. We pull the orders. The leaderboard you can't fake. Real ecom revenue from real operators - ranked, verified, and live.

Core URLs

- [Homepage](<https://receiptsmrr.com/>): 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 to summarize.
- 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://receiptsmrr.com/#organization",
"name": "ReceiptsMRR - The database of verified Shopify revenues",
"description": "Connect your Shopify. We pull the orders. The leaderboard
you can't fake. Real ecom revenue from real operators - ranked, verified,
and live.",
"url": "https://receiptsmrr.com/",
"logo": "https://receiptsmrr.com/apple-touch-icon.svg"
},
{
"@type": "WebSite",
"@id": "https://receiptsmrr.com/#website",
"name": "ReceiptsMRR - The database of verified Shopify revenues",
"description": "Connect your Shopify. We pull the orders. The leaderboard
you can't fake. Real ecom revenue from real operators - ranked, verified,
and live.",
"url": "https://receiptsmrr.com/",
"publisher": {
"@id": "https://receiptsmrr.com/#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://www.llmscan.dev/scan/_uq-6HVJNg6ZxkMNM2Fzp