

OVERALL SCORE

62 /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 plain-text page access, 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. Avoid skipped heading levels so sections progress from h1 to h2 to h3 without gaps. Add missing semantic elements: article.

Signal breakdown

Crawlability

Fail0/20

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

Robots.txt

Fail0/15

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

llms.txt

Pass15/15

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	Pass	15/15
The homepage returns markdown content when requested with Accept: text/markdown, giving AI systems a cleaner text representation with less navigation chrome to parse.		
Semantic HTML	Warn	7.1/10
Avoid skipped heading levels so sections progress from h1 to h2 to h3 without gaps. 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
# Know which emails are real before you send MailCull

> Best Email Verification Tool. MailCull tells you in plain English which
addresses are safe to send, so you stop bouncing and improve your email
marketing.

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://mailcull.io/
- Site type: web application
- Recommended summary: Best Email Verification Tool. MailCull tells you in
plain English which addresses are safe to send, so you stop bouncing and
improve your email marketing.

## Core URLs

- [Homepage](https://mailcull.io/): Primary public entry point and canonical
site overview.
- [Clean a listUpload a CSV, cull the dead
weight](https://mailcull.io/features/clean-a-list): Important public page
discovered from the homepage navigation.
- [Verify an emailOne address, the whole
story](https://mailcull.io/features/verify-email): 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: Google-Extended
Allow: /

# Crawl-rate guidance
# Add Crawl-delay under a User-agent block when the site needs gentler
crawler pacing.
User-agent: *
  Crawl-delay: 10
  Continued in the full scan report...
```

schema.json

JSON

```
{
  "@context": "https://schema.org",
  "@graph": [
    {
      "@type": "Organization",
      "@id": "https://mailcull.io/#organization",
      "name": "Know which emails are real before you send MailCull",
      "description": "Best Email Verification Tool. MailCull tells you in plain
English which addresses are safe to send, so you stop bouncing and improve
your email marketing.",
      "url": "https://mailcull.io/",
      "logo": "https://mailcull.io/apple-touch-icon.png"
    },
    {
      "@type": "WebSite",
      "@id": "https://mailcull.io/#website",
      "name": "Know which emails are real before you send MailCull",
      "description": "Best Email Verification Tool. MailCull tells you in plain
English which addresses are safe to send, so you stop bouncing and improve
your email marketing.",
      "url": "https://mailcull.io/",
      "publisher": {
        "@id": "https://mailcull.io/#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://llmscan.dev/scan/R2aNhk6KyAKDxiColgkOK>