

OVERALL SCORE

73 /100

AI-Readable

Executive summary

This site is mostly readable by AI tools, with a few improvements still recommended. Key strengths include AI guidance file and plain-text page access, while page structure and homepage access need attention. Recommended next step: 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.

Recommended next step

1. 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.
2. 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.
3. Fix robots.txt syntax issues so each rule uses Field: value format, directives appear under a User-agent, and Sitemap entries use absolute URLs.

Signal breakdown

CrawlabilityWarn10/20

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.txtWarn7.5/15

Fix robots.txt syntax issues so each rule uses Field: value format, directives appear under a User-agent, and Sitemap entries use absolute URLs.

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	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	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		
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

Your Newsletter, Your Voice. Published in 5 Minutes | HeyNews

> HeyNews learns your writing style from past newsletters, curates stories from RSS, Reddit, social media, and blogs, then generates publish-ready issues in minutes. 14-day free trial. Free Growth Audit (\$99 value) when yo

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://heynews.co/>
- Site type: organization
- Recommended summary: HeyNews learns your writing style from past newsletters, curates stories from RSS, Reddit, social media, and blogs, then generates publish-ready issues in minutes. 14-day free trial. Free Growth Audit (\$99 value) when yo

Core URLs

- [Homepage](<https://heynews.co/>): Primary public entry point and canonical site overview.

Content Recommendations

~ Add one-sentence descriptions for product, service, documentation, 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: /
```

schema.json

JSON

```
{
  "@context": "https://schema.org",
  "@graph": [
    {
      "@type": "Organization",
      "@id": "https://heynews.co/#organization",
      "name": "Your Newsletter, Your Voice. Published in 5 Minutes | HeyNews",
      "description": "HeyNews learns your writing style from past newsletters,
      curates stories from RSS, Reddit, social media, and blogs, then generates
      publish-ready issues in minutes. 14-day free trial. Free Growth Audit ($99
      value) when you sign up.",
      "url": "https://heynews.co/",
      "logo": "https://heynews.co/assets/icons/apple-touch-icon.png"
    },
    {
      "@type": "WebSite",
      "@id": "https://heynews.co/#website",
      "name": "Your Newsletter, Your Voice. Published in 5 Minutes | HeyNews",
      "description": "HeyNews learns your writing style from past newsletters,
      curates stories from RSS, Reddit, social media, and blogs, then generates
      publish-ready issues in minutes. 14-day free trial. Free Growth Audit ($99
      value) when you sign up.",
      "url": "https://heynews.co/",
      "publisher": {
        "@id": "https://heynews.co/#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/ruwkbLM9DFcv8FKSx_p2Q