

OVERALL SCORE

78 /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 homepage access and crawler policy need attention. Recommended next step: serve a non-empty HTML homepage with a canonical link tag that points to the preferred public URL.

Recommended next step

- 1. Serve a non-empty HTML homepage with a canonical link tag that points to the preferred public URL.
- 2. Fix robots.txt syntax issues so each rule uses Field: value format, directives appear under a User-agent, and Sitemap entries use absolute URLs.
- 3. Use exactly one h1 element and move secondary section titles to h2-h6. Avoid skipped heading levels so sections progress from h1 to h2 to h3 without gaps. Add missing semantic elements: article.

Signal breakdown

Crawlability Serve a non-empty HTML homepage with a canonical link tag that points to the preferred public URL.	Warn 10/20
Robots.txt Fix robots.txt syntax issues so each rule uses Field: value format, directives appear under a User-agent, and Sitemap entries use absolute URLs.	Warn 7.5/15
llms.txt The llms.txt file was found and includes the expected text, length, heading, and URL signals.	Pass 15/15

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	5.7/10
Use exactly one h1 element and move secondary section titles to h2-h6. 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

```
# Cloudflare: Build for the agent era

> Welcome to Cloudflare - Powering the next generation of applications

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://www.cloudflare.com/
- Site type: organization
- Recommended summary: Welcome to Cloudflare - Powering the next generation
of applications

## Core URLs

- [Homepage](https://www.cloudflare.com/): Primary public entry point and
canonical site overview.
- [Products](https://www.cloudflare.com/products/): Important public page
discovered from the homepage navigation.
- [Solutions](https://www.cloudflare.com/solutions/): Important public page
discovered from the homepage navigation.
- [Resources](https://www.cloudflare.com/resource-hub/): Important public
page discovered from the homepage navigation.
- [Pricing](https://www.cloudflare.com/plans/): Pricing, plans, and
packaging information.
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 AI crawlers so public content permissions are
clear.
User-agent: ClaudeBot
Allow: /

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

schema.json

```
JSON
{
  "@context": "https://schema.org",
  "@graph": [
    {
      "@type": "Organization",
      "@id": "https://www.cloudflare.com/#organization",
      "name": "Cloudflare: Build for the agent era",
      "description": "Welcome to Cloudflare - Powering the next generation of applications",
      "url": "https://www.cloudflare.com/",
      "logo": "https://www.cloudflare.com/favicon.ico"
    },
    {
      "@type": "WebSite",
      "@id": "https://www.cloudflare.com/#website",
      "name": "Cloudflare: Build for the agent era",
      "description": "Welcome to Cloudflare - Powering the next generation of applications",
      "url": "https://www.cloudflare.com/",
      "publisher": {
        "@id": "https://www.cloudflare.com/#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://www.llmscan.dev/scan/YxU53Nv0kY3A1wbt8HCBR>