

# LLM Scan

PUBLIC AI VISIBILITY REPORT

# www.besourceable.com

Scanned Jul 6, 2026, 11:02 UTC

OVERALL SCORE

# 68 /100

Needs Work

## Executive summary

This site has a useful foundation, but important gaps still limit AI readability. Key strengths include crawler policy and plain-text page access, while sitemap and homepage access need attention. Recommended next step: regenerate sitemap.xml with valid XML and a sitemap `<urlset>` or `<sitemapindex>` root element.

## Recommended next step

1. Serve a non-empty HTML homepage with a canonical link tag that points to the preferred public URL.
2. Regenerate sitemap.xml with valid XML and a sitemap `<urlset>` or `<sitemapindex>` root element.
3. Publish `/llms.txt` as text or markdown with more than 200 characters, markdown headings, and at least one absolute URL.

## Signal breakdown

### Crawlability

Warn 10/20

Serve a non-empty HTML homepage with a canonical link tag that points to the preferred public URL.

### Robots.txt

Pass 15/15

Add Sitemap directives to robots.txt so crawlers can find canonical sitemap files.

### 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

Fail 0/10

Regenerate sitemap.xml with valid XML and a sitemap <urlset> or <sitemapindex> root element.

## 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

Shorten the meta description to 160 characters or fewer. 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 AI-specific head meta tags, robots noai/noimageai directive, Content-Signal robots.txt directive so AI systems can consistently discover content usage preferences across robots.txt, HTTP headers, and HTML metadata.

# Suggested fixes

## llms.txt

### MARKDOWN

# Sourceable | AI Search Visibility & AEO Platform

> Track brand mentions, monitor your share of voice, and rank higher on ChatGPT, Claude, Gemini, and Perplexity with our Answer Engine Optimization (AEO) platform.

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.besourceable.com/>
- Site type: organization
- Recommended summary: Track brand mentions, monitor your share of voice, and rank higher on ChatGPT, Claude, Gemini, and Perplexity with our Answer Engine Optimization (AEO) platform.

### ## Core URLs

- [Homepage](<https://www.besourceable.com/>): Primary public entry point and canonical site overview.
- [Blog](<https://www.besourceable.com/blog>): Editorial updates, announcements, and current context.
- [AI Robots.txt Checker](<https://www.besourceable.com/tools/robots-checker>): 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.

# Sitemap discovery
# Add this top-level directive so search and AI crawlers can discover the
canonical sitemap.
Sitemap: https://www.besourceable.com/sitemap.xml

# AI crawler access
# Add explicit Allow rules for AI crawlers so public content permissions are
clear.
User-agent: GPTBot
Allow: /

User-agent: ChatGPT-User
Allow: /

User-agent: ClaudeBot
Allow: /

User-agent: Claude-Web
Allow: /

User-agent: PerplexityBot
Allow: /
Continued in the full scan report...
```

## schema.json

### JSON

```
{
  "@context": "https://schema.org",
  "@graph": [
    {
      "@type": "Organization",
      "@id": "https://www.besourceable.com/#organization",
      "name": "Sourceable | AI Search Visibility & AEO Platform",
      "description": "Track brand mentions, monitor your share of voice, and rank
higher on ChatGPT, Claude, Gemini, and Perplexity with our Answer Engine
Optimization (AEO) platform.",
      "url": "https://www.besourceable.com/",
      "logo": "https://www.besourceable.com/logo-icon.svg?v=2",
      "sameAs": [
        "https://twitter.com/BeSourceable",
        "https://www.linkedin.com/company/besourceable",
        "https://x.com/BeSourceable"
      ]
    },
    {
      "@type": "WebSite",
      "@id": "https://www.besourceable.com/#website",
      "name": "Sourceable | AI Search Visibility & AEO Platform",
      "description": "Track brand mentions, monitor your share of voice, and rank
higher on ChatGPT, Claude, Gemini, and Perplexity with our Answer Engine
Optimization (AEO) platform.",
      "url": "https://www.besourceable.com/",
      "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/\\_SHWpftjUnV26uHHv4dgw](https://www.llmscan.dev/scan/_SHWpftjUnV26uHHv4dgw)