

# LLM Scan

PUBLIC AI VISIBILITY REPORT

# pdftomarkdown.ai

Scanned Jul 7, 2026, 11:01 UTC

OVERALL SCORE

**56** /100

Needs Work

## Executive summary

This site has a useful foundation, but important gaps still limit AI readability. Key strengths include crawler policy and AI guidance file, while plain-text page access and structured data need attention. Recommended next step: 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.

## Recommended next step

1. 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.
2. Serve a non-empty HTML homepage with a canonical link tag that points to the preferred public URL.
3. Add JSON-LD structured data with Organization or WebSite schema so AI systems can identify the site owner or website entity.

## 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

robots.txt allows crawler access and includes Sitemap references.

### 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

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

Fail 0/10

Add JSON-LD structured data with Organization or WebSite schema so AI systems can identify the site owner or website entity.

## Content signals

Fail 0/5

Add the standard directive 'Content-Signal: ai-train=no, search=yes, ai-input=yes' to robots.txt, HTML metadata, or HTTP headers so AI systems can discover content usage preferences.

# Suggested fixes

## Fix Structured data

### HTML

```
<script type="application/ld+json">
{
  "@context": "https://schema.org",
  "@graph": [
    {
      "@type": "Organization",
      "@id": "https://pdftomarkdown.ai/#organization",
      "name": "PDF to Markdown Converter | PDFtoMarkdown.ai",
      "description": "Convert any PDF to clean Markdown instantly. No code, no
      setup, no subscription. Works with ChatGPT, Claude, and Obsidian. Free for
      10 pages.",
      "url": "https://pdftomarkdown.ai/",
      "logo": "https://pdftomarkdown.ai/apple-touch-icon.png"
    },
    {
      "@type": "WebSite",
      "@id": "https://pdftomarkdown.ai/#website",
      "name": "PDF to Markdown Converter | PDFtoMarkdown.ai",
      "description": "Convert any PDF to clean Markdown instantly. No code, no
      setup, no subscription. Works with ChatGPT, Claude, and Obsidian. Free for
      10 pages.",
      "url": "https://pdftomarkdown.ai/",
      "publisher": {
        "@id": "https://pdftomarkdown.ai/#organization"
      },
      "inLanguage": "en"
    }
  ]
}
```

Continued in the full scan report...

## Fix Content signals

### HTML

```
<meta http-equiv="Content-Signal" content="ai-train=no, search=yes,
ai-input=yes" />
<meta name="content-signal" content="ai-train=no, search=yes, ai-input=yes"
/>
```

## Full report

<https://www.llmscan.dev/scan/LzkVRoSFrkZsRQO6lcpnO>