

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 homepage access and AI guidance file, while plain-text page access and crawler policy 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. 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

Crawlability The homepage resolves, connects, returns HTTP 200 OK, exposes a canonical URL, and is not blocked by robots.txt.	Pass 20/20
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	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	Pass	10/10
The homepage has a valid title, meta description, heading structure, semantic landmarks, substantial body copy, and descriptive anchor text.		
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 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

```
# AI WALLAH | Automate Repetitive Work for Indian SMEs

> Custom AI systems for Indian SMEs. Automate up to 80% of repetitive work -
data entry, emails, handoffs. Workflow-first.

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://aiwallah.in/
- Site type: organization
- Recommended summary: Custom AI systems for Indian SMEs. Automate up to 80%
of repetitive work - data entry, emails, handoffs. Workflow-first.

## Core URLs

- [Homepage](https://aiwallah.in/): Primary public entry point and canonical
site overview.
- [About](https://aiwallah.in/about): Background information about the
organization or site owner.
- [Projects](https://aiwallah.in/projects): Important public page discovered
from the homepage navigation.
- [Solutions](https://aiwallah.in/solutions): Important public page
discovered from the homepage navigation.
- [Founder](https://aiwallah.in/resume): Important public page discovered
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: 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
```

schema.json

```
JSON
{
  "@context": "https://schema.org",
  "@graph": [
    {
      "@type": "Organization",
      "@id": "https://aiwallah.in/#organization",
      "name": "AI WALLAH | Automate Repetitive Work for Indian SMEs",
      "description": "Custom AI systems for Indian SMEs. Automate up to 80% of repetitive work - data entry, emails, handoffs. Workflow-first.",
      "url": "https://aiwallah.in/",
      "logo": "https://aiwallah.in/favicon.png"
    },
    {
      "@type": "WebSite",
      "@id": "https://aiwallah.in/#website",
      "name": "AI WALLAH | Automate Repetitive Work for Indian SMEs",
      "description": "Custom AI systems for Indian SMEs. Automate up to 80% of repetitive work - data entry, emails, handoffs. Workflow-first.",
      "url": "https://aiwallah.in/",
      "publisher": {
        "@id": "https://aiwallah.in/#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://llmscan.dev/scan/jVPIfc5uttDzOIlc0OEcZ>