

OVERALL SCORE

28 /100 Poor

Executive summary

This site is difficult for AI tools to read right now. Key strengths include sitemap and content signals, while homepage access and crawler policy need attention. Recommended next step: remove AI crawler Disallow: / rules or replace them with narrower path-level restrictions for private content only.

Recommended next step

- 1. Remove AI crawler Disallow: / rules or replace them with narrower path-level restrictions for private content only.
- 2. Remove AI crawler Disallow: / rules or add narrower Allow/Disallow rules if AI crawlers should be able to discover public content.
- 3. 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.

Signal breakdown

Crawlability Remove AI crawler Disallow: / rules or replace them with narrower path-level restrictions for private content only.	Fail	0/20
Robots.txt Remove AI crawler Disallow: / rules or add narrower Allow/Disallow rules if AI crawlers should be able to discover public content.	Fail	0/15
llms.txt Publish /llms.txt as text or markdown with more than 200 characters, markdown headings, and at least one absolute URL.	Warn	7.5/15

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	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	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

Fix Structured data

HTML

```
<script type="application/ld+json">
{
  "@context": "https://schema.org",
  "@graph": [
    {
      "@type": "Organization",
      "@id": "https://app.instdub.com/#organization",
      "name": "InstDub - Your Language. Every Video.",
      "description": "InstDub - Your Language. Every Video. Instantly turn any
online video into natural speech in your language with real-time AI
translation and dubbing.",
      "url": "https://app.instdub.com/",
      "logo": "https://app.instdub.com/icon32.png"
    },
    {
      "@type": "WebSite",
      "@id": "https://app.instdub.com/#website",
      "name": "InstDub - Your Language. Every Video.",
      "description": "InstDub - Your Language. Every Video. Instantly turn any
online video into natural speech in your language with real-time AI
translation and dubbing.",
      "url": "https://app.instdub.com/",
      "publisher": {
        "@id": "https://app.instdub.com/#organization"
      },
      "inLanguage": "en"
    }
  ]
}
```

Continued in the full scan report...

Full report

https://llmscan.dev/scan/JdLBCE_3Orx9FFMXvJA29