

OVERALL SCORE

54 /100

Needs Work

Executive summary

This site has a useful foundation, but important gaps still limit AI readability. Key strengths include homepage access and crawler policy, while AI guidance file and plain-text page access need attention. Recommended next step: publish an AI guidance file as text or markdown with more than 200 characters, markdown headings, and at least one absolute URL.

Recommended next step

- 1. Publish /llms.txt as text or markdown with more than 200 characters, markdown headings, and at least one absolute URL.
- 2. 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.
- 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 Pass 20/20

The homepage resolves, connects, returns HTTP 200 OK, exposes a canonical URL, and is not blocked by robots.txt.

Robots.txt Pass 15/15

robots.txt allows crawler access and includes Sitemap references.

llms.txt Fail 0/15

Publish /llms.txt as text or markdown with more than 200 characters, markdown headings, and at least one absolute URL.

Sitemap	Pass	10/10
The sitemap index is valid and the checked child sitemap 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	8.6/10
Avoid skipped heading levels so sections progress from h1 to h2 to h3 without gaps.		
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://huggingface.co/#organization",
      "name": "Hugging Face - The AI community building the future.",
      "description": "We’re on a journey to advance and democratize artificial intelligence through open source and open science.",
      "url": "https://huggingface.co/",
      "logo": "https://huggingface.co/front/assets/huggingface_logo-noborder.svg",
      "sameAs": [
        "https://github.com/huggingface",
        "https://twitter.com/huggingface",
        "https://www.linkedin.com/company/huggingface/"
      ]
    },
    {
      "@type": "WebSite",
      "@id": "https://huggingface.co/#website",
      "name": "Hugging Face - The AI community building the future.",
      "description": "We’re on a journey to advance and democratize artificial intelligence through open source and open science.",
      "url": "https://huggingface.co/",
      "publisher": {
        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/NlluPIAEolKTlme_oiTLY