

OVERALL SCORE

47 /100

Needs Work

Executive summary

This site has a useful foundation, but important gaps still limit AI readability. Key strengths include AI guidance file and sitemap, 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 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	Warn	7.1/10
Shorten the meta description to 160 characters or fewer. 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 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

llms.txt

MARKDOWN

Casa Editrice Toni | Libri e Software per Patenti e Autoscuole

> Casa Editrice Toni - Da tre generazioni forniamo materiale didattico di qualita per patenti AM, A, B, C, D, E, ADR e software innovativi per autoscuole in tutta Italia.

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://editricetoni.it/>
- Site type: organization
- Recommended summary: Casa Editrice Toni - Da tre generazioni forniamo materiale didattico di qualita per patenti AM, A, B, C, D, E, ADR e software innovativi per autoscuole in tutta Italia.

Core URLs

- [Homepage](<https://editricetoni.it/>): Primary public entry point and canonical site overview.
- [](<https://editricetoni.it/cdn-cgi/l/email-protection>): Important public page discovered from the homepage navigation.
- [Patentino Ciclomotore](<https://editricetoni.it/patente-am>): Important public page discovered from the homepage navigation.
- [Il ciclomotore e la
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 blocked AI crawlers; remove or narrow
conflicting Disallow rules if your crawler target requires precedence.
User-agent: GPTBot
Allow: /

User-agent: ChatGPT-User
Allow: /

User-agent: ClaudeBot
Allow: /

User-agent: Claude-Web
Allow: /

User-agent: PerplexityBot
Allow: /

User-agent: Google-Extended
Allow: /

# Crawl-rate guidance
Continued in the full scan report...
```

schema.json

JSON

```
{
"@context": "https://schema.org",
"@graph": [
{
"@type": "Organization",
"@id": "https://editricetoni.it/#organization",
"name": "Casa Editrice Toni | Libri e Software per Patenti e Autoscuole",
"description": "Casa Editrice Toni - Da tre generazioni forniamo materiale
didattico di qualita per patenti AM, A, B, C, D, E, ADR e software
innovativi per autoscuole in tutta Italia.",
"url": "https://editricetoni.it/",
"logo": "https://editricetoni-it-nextjs.pages.dev/images/LogoToni.png"
},
{
"@type": "WebSite",
"@id": "https://editricetoni.it/#website",
"name": "Casa Editrice Toni | Libri e Software per Patenti e Autoscuole",
"description": "Casa Editrice Toni - Da tre generazioni forniamo materiale
didattico di qualita per patenti AM, A, B, C, D, E, ADR e software
innovativi per autoscuole in tutta Italia.",
"url": "https://editricetoni.it/",
"publisher": {
"@id": "https://editricetoni.it/#organization"
},
"inLanguage": "it"
}
]
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/9kf_3T7R5V12f4kGnMZq-