

OVERALL SCORE

65 /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 plain-text page access and homepage access 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. Review AI crawler Disallow rules and keep only the paths that should be excluded from AI crawler access; serve a non-empty HTML homepage with a canonical link tag.
- 3. 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 Warn 10/20

Review AI crawler Disallow rules and keep only the paths that should be excluded from AI crawler access; serve a non-empty HTML homepage with a canonical link tag.

Robots.txt Warn 7.5/15

Fix robots.txt syntax issues so each rule uses Field: value format, directives appear under a User-agent, and Sitemap entries use absolute URLs.

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 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	7.1/10
Add exactly one h1 element that describes the page topic. Add missing semantic elements: main.		
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

Awesome Photo Booth Hire in Johannesburg & Pretoria

> Photo booth hire in Gauteng and surrounds. Get your guests fired up with the timeless appeal of a photo booth! An affordable value-add for any event.

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://selfiecube.co.za/>
- Site type: local business
- Recommended summary: Photo booth hire in Gauteng and surrounds. Get your guests fired up with the timeless appeal of a photo booth! An affordable value-add for any event.

Core URLs

- [Homepage](<https://selfiecube.co.za/>): Primary public entry point and canonical site overview.
 - [Prices](<https://selfiecube.co.za/prices/>): Important public page discovered from the homepage navigation.
 - [Our Booths](<https://selfiecube.co.za/photo-booths/>): Important public page discovered from the homepage navigation.
 - [Selfie AINew](<https://selfiecube.co.za/selfie-ai-photo-booth/>): Important public page discovered from the homepage navigation.
- 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://selfiecube.co.za/#organization",
"name": "Awesome Photo Booth Hire in Johannesburg & Pretoria",
"description": "Photo booth hire in Gauteng and surrounds. Get your guests
fired up with the timeless appeal of a photo booth! An affordable value-add
for any event.",
"url": "https://selfiecube.co.za/",
"logo":
"https://selfiecube.co.za/wp-content/uploads/2021/08/cropped-Favicon-2-300x3
00.png",
"sameAs": [
"https://www.facebook.com/SelfieCube",
"https://www.instagram.com/selfie_cube/"
]
},
{
"@type": "WebSite",
"@id": "https://selfiecube.co.za/#website",
"name": "Awesome Photo Booth Hire in Johannesburg & Pretoria",
"description": "Photo booth hire in Gauteng and surrounds. Get your guests
fired up with the timeless appeal of a photo booth! An affordable value-add
for any event."
}
]
}
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://llmscan.dev/scan/pfwf981_lIamucTuC0Mh7