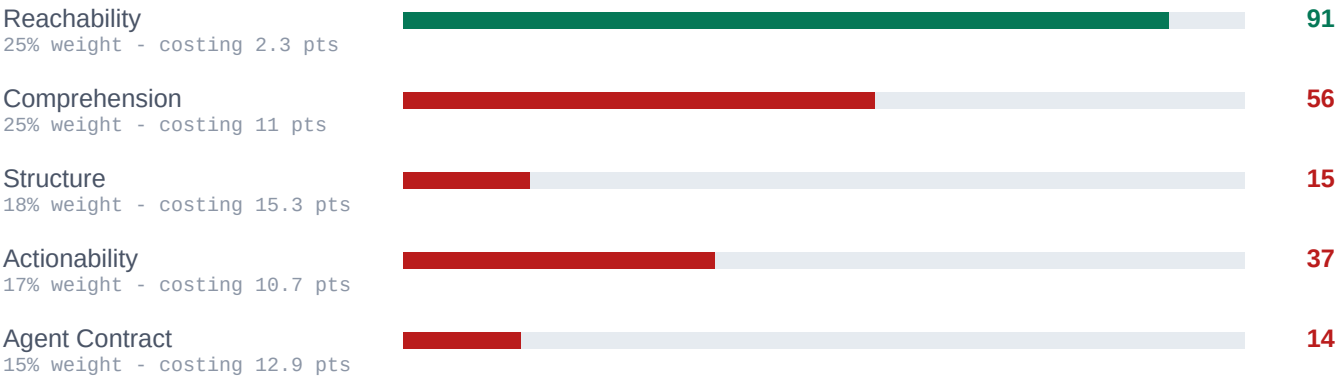
# example.com

https://example.com/

Scanned 2026-09-06 16:55 UTC. Measured with AXRAY against AX specification v1.5.1, published in full at https://axray.online/ax-spec

**48** / 100

**Grade D**




IN PLAIN ENGLISH

## Today an AI assistant would struggle with example.com, and would tell people things about it that are not right.

### Can an AI assistant open this page at all?

YES

Yes. When an assistant goes to look, the page answers it.

This is the thing most sites get wrong first, and this one does not.

### Will it describe this business correctly?

NO

No. There is not enough readable text for it to know what this is, so it writes its own version.

The description people are shown is written by a machine that had nothing to go on, and you will not be told what it said.

### Can it quote the facts - prices, contact details, opening hours?

NO

No. The facts exist only as styled text, so an assistant either skips them or misreads them.

When somebody asks what you charge, a competitor who publishes the number gets quoted and you get "not stated".

### Can it take the next step - follow a link, use a form, get in touch?

NO

No. The paths an agent would take are dead ends.

An agent stops recommending a site at exactly the point where it would have sent you a customer.

### Have the AI companies been told your rules?

NO

No. Every AI company is applying its own default to this site.

You are not choosing between being trained on and being cited: both are being chosen for you.

---

WHAT CHANGES IF YOU ACT

## The difference this makes

**Applying only the 22 configuration-level fixes below takes this page from 48 to 91 out of 100, with no engineering work.**

+ You decide which AI companies may use this content, instead of each of them deciding for you.
+ Prices and contact details become quotable, word for word, instead of guessed at.
+ An agent can carry a visitor from the answer it gave to your form or your phone number.
+ Assistants describe the business in its own words instead of guessing at them.

12 FAILING | 14 WARNINGS

# Fix list, highest value first

FAIL

## Ships structured data

+6.2 pts

No JSON-LD and no microdata anywhere on the page.

Structured data is the only part of your page an agent can consume without interpretation. Everything else it has to infer, and inference is where it invents facts about you.

HOW TO FIX IT - CONFIGURATION ONLY

Add a JSON-LD block describing what this page actually is.

Where: Inside <head>, one block per page.

```
<script type="application/ld+json">
{
  "@context": "https://schema.org",
  "@type": "Organization",
  "name": "Your company",
  "url": "https://example.com",
  "description": "One sentence an assistant can quote.",
  "sameAs": ["https://x.com/you", "https://github.com/you"]
}
</script>
```

FAIL

## Publishes an llms.txt

+4.2 pts

No /llms.txt or /llms-full.txt at https://example.com.

Without it, an assistant deciding what your site is about crawls whatever it happens to find. llms.txt is the cheapest way to control that first impression, and today most of your competitors do not have one either.

HOW TO FIX IT - CONFIGURATION ONLY

Publish an llms.txt: a title, one paragraph, then curated links grouped by intent.

Where: /llms.txt

```
# Example Inc

> One paragraph an assistant can quote verbatim about what you do and who it is for.

## Docs
- [Quickstart](https://example.com/docs/quickstart): Get running in five minutes.
- [API reference](https://example.com/docs/api): Every endpoint with examples.

## Product
- [Pricing](https://example.com/pricing): Plans, limits and exact prices.
- [Changelog](https://example.com/changelog): What shipped, newest first.

## Policies
- [Terms](https://example.com/terms)
- [Contact](https://example.com/contact): support@example.com
```

FAIL

## Publishes a sitemap

+4.1 pts

No sitemap found at https://example.com/sitemap_index.xml or in robots.txt.

Without a sitemap, discovery depends entirely on crawlable links. Any page not reachable by a plain <a href> is invisible.

HOW TO FIX IT - CONFIGURATION ONLY

Generate a sitemap.xml and reference it from robots.txt.

Where: Append to /robots.txt

```
Sitemap: https://example.com/sitemap.xml
```

FAIL

## Enough substance to be worth citing

+3.3 pts

Only 19 words of readable text reached the agent.

There is not enough here for an assistant to build an answer from, so it will use someone else as the source.

HOW TO FIX IT

Make sure the real content is in the HTML, then make sure there is enough of it.

FAIL

## Open Graph card is complete

+2.8 pts

Missing: og:title, og:description, og:type, og:url, og:image.

When a page is JavaScript-heavy or truncated, Open Graph tags are frequently the only clean summary an agent gets.

HOW TO FIX IT - CONFIGURATION ONLY

Add the four core Open Graph tags plus an image.

```
<meta property="og:title" content="Page title">
<meta property="og:description" content="One sentence.">
<meta property="og:type" content="website">
<meta property="og:url" content="https://example.com/page">
<meta property="og:image" content="https://example.com/og.png">
```

FAIL

## Links onward into the site

+2.7 pts

Only 0 internal links on the page.

A page with no onward links is a cul-de-sac. Everything you publish that is not linked from a crawlable page effectively does not exist.

HOW TO FIX IT

Add a server-rendered navigation and contextual links to related pages.

FAIL

## Has a meta description

+2.6 pts

No <meta name="description"> on the page.

It is the one sentence you get to write yourself. Without it the assistant summarises you from whatever text it happened to read first, which is often your cookie banner.

HOW TO FIX IT - CONFIGURATION ONLY

Add a 120-160 character description that states what the page offers, in plain language.

```
<meta name="description" content="Scan any URL and see exactly what an AI agent can read, understand
  and act on. Get an AX score and a prioritised fix list.">
```

WARN

## Programmatic surface is discoverable

+2.3 pts

No well-known agent manifest, no OpenAPI reference and no developer docs link found.

This is the frontier of AX: sites that publish a machine-callable surface get used by agents directly instead of being scraped approximately.

HOW TO FIX IT - CONFIGURATION ONLY

Publish a minimal agent manifest describing what you offer and where to call it.

Where: /.well-known/mcp.json

```
{
  "name": "Example",
  "description": "What an agent can do with this service.",
  "documentation": "https://example.com/docs",
  "endpoints": [
    { "type": "openapi", "url": "https://example.com/openapi.json" }
  ]
}
```

FAIL

## Reachable by a human, discoverably

+2.2 pts

No mailto, tel, contact link or contactPoint data found.

"How do I get in touch with them" is one of the most common agent-mediated questions about a company. Make the answer machine-readable.

HOW TO FIX IT - CONFIGURATION ONLY

Add a contactPoint to your Organization schema and link a real contact page.

```
{"@type":"Organization","contactPoint":[{"@type":"ContactPoint","contactType":"customer
  support","email":"support@example.com","url":"https://example.com/contact"}]}
```

FAIL

## Declares what an agent can do here

+2.2 pts

The page declares no potentialAction, so every affordance on it - search, order, book, subscribe, get in touch - has to be inferred from the HTML.

Inference is where an agent gets it wrong or gives up. Declaring the action, its URL template and the parameters it takes is the difference between an agent that can complete a task on your site and one that reads about it and leaves.

HOW TO FIX IT - CONFIGURATION ONLY

Declare the actions this page supports with schema.org potentialAction.

Where: Any page. The site-wide SearchAction belongs on the home page; an OrderAction or ReserveAction belongs on the page that offers it.

```
{
  "@context": "https://schema.org",
  "@type": "WebSite",
  "url": "https://example.com/",
  "potentialAction": {
    "@type": "SearchAction",
    "target": {
      "@type": "EntryPoint",
      "urlTemplate": "https://example.com/search?q={query}"
    },
    "query-input": "required name=query"
  }
}
```

FAIL

### Main content is marked as such

+2.0 pts

No <main> element. Landmarks present: none.

Without a main landmark an extractor has to guess your content boundary heuristically, and on marketing pages it usually guesses the header.

HOW TO FIX IT - CONFIGURATION ONLY

Wrap the primary content in <main>, and each self-contained item in <article>.

```
<main>
  <article>
    <h1>...</h1>
    ...
  </article>
</main>
```

WARN

### Has a descriptive <title>

+2.0 pts

The title is only 14 characters: "Example Domain".

Very short titles force the assistant to guess at context and it will guess from the URL.

HOW TO FIX IT - CONFIGURATION ONLY

Expand the title to name both the page subject and the brand.

FAIL

### Exposes an agent-callable manifest

+1.8 pts

No /.well-known/mcp.json and no ai-plugin.json.

This is the highest-leverage unclaimed slot on the web right now. The sites that publish a callable surface get used directly instead of guessed at.

HOW TO FIX IT

Publish a manifest naming what an agent can do with you and where to call it.

Where: /.well-known/mcp.json

```
{
  "schema_version": "2025-06-18",
  "name": "example",
  "description": "Search the catalogue and check availability.",
  "documentation": "https://example.com/docs",
  "servers": [
    { "type": "http", "url": "https://mcp.example.com" }
  ],
  "contact": "support@example.com"
}
```

WARN

### Declares a canonical URL

+1.5 pts

No <link rel="canonical"> on the page.

Agents dedupe sources by canonical URL. Without one, the same content behind different query strings competes with itself and dilutes every citation.

HOW TO FIX IT - CONFIGURATION ONLY

Add a self-referencing canonical link to every page.

```
<link rel="canonical" href="https://example.com/">
```

FAIL

### Identity is machine-verifiable

+1.5 pts

No Organization structured data on the page.

Entity resolution is how assistants decide two mentions are the same company. Unresolved entities get merged with, or mistaken for, someone else.

HOW TO FIX IT - CONFIGURATION ONLY

Publish Organization data with sameAs links to your verified profiles.

```
{"@context":"https://schema.org","@type":"Organization","name":"Example
  Inc","url":"https://example.com","logo":"https://example.com/logo.png","sameAs":["https://x.com/exam
  ple","https://github.com/example","https://www.linkedin.com/company/example"]}
```

WARN

### Content exists without running JavaScript

+1.5 pts

Only 19 words arrived in the raw HTML. Some of the page is probably assembled in the browser.

Partial server rendering means agents see a fragment of your argument and may summarise you from the fragment.

HOW TO FIX IT

Server-render the primary content of the page, not just the shell and navigation.

WARN

### States an explicit AI policy

+1.3 pts

robots.txt has no per-agent AI groups, and there is no TDM reservation or noai directive anywhere.

With no explicit policy, every operator applies its own default to you and you have no documented position if you later object.

HOW TO FIX IT - CONFIGURATION ONLY

Say what you want, per crawler. Allow the ones that send traffic; decide separately about the training ones.

Where: /robots.txt

```
# Answer engines: allowed, they cite and link us.
User-agent: OAI-SearchBot
User-agent: PerplexityBot
User-agent: Claude-SearchBot
User-agent: DuckAssistBot
Allow: /

# User-triggered fetches: allowed, a person is waiting.
User-agent: ChatGPT-User
User-agent: Claude-User
User-agent: Perplexity-User
Allow: /

# Training crawlers: your call. This example allows them.
User-agent: GPTBot
User-agent: ClaudeBot
User-agent: Google-Extended
Allow: /
```

WARN

### Site search is machine-usable

+1.3 pts

No crawlable GET search form and no SearchAction.

Without a GET-addressable search, an agent looking for one specific page on your site has to fetch dozens of pages to find it, and usually gives up first.

HOW TO FIX IT - CONFIGURATION ONLY

Expose search at a GET URL like /search?q=... and declare it as a SearchAction.

WARN

### Declares AI usage preferences the standard way

+1.2 pts

No Content-Usage header and no Content-Usage rule in robots.txt, so this site states no preference in the form the standards work is converging on.

robots.txt can only say whether a crawler may fetch you. It has never been able to say what may be done with the bytes afterwards. The IETF vocabulary is the first machine-readable way to draw that line, and stating your position now costs one header and settles a question you would otherwise argue about later.

HOW TO FIX IT - CONFIGURATION ONLY

Send a Content-Usage header, or add the same directive to robots.txt, saying whether you permit training and search use.

Where: Draft standard: draft-ietf-aipref-vocab and draft-ietf-aipref-attach. It expresses a preference rather than enforcing one, and it sits alongside your existing robots.txt rules rather than replacing them.

```
# As an HTTP response header on your pages:
Content-Usage: train-ai=n

# Or, for the whole site, in /robots.txt:
User-agent: *
Content-Usage: train-ai=n, search=y
Allow: /
```

WARN

## Headings form a sane outline

+1.2 pts

Only 1 heading on the page.

Agents chunk long pages by heading. A page with no outline is chunked arbitrarily and quoted out of context.

HOW TO FIX IT - CONFIGURATION ONLY

Break the content into sections with <h2> headings that read as answers to real questions.

---

WARN

## Declares content licensing

+1.2 pts

No rel="license" link, no license or copyright fields in structured data, and no RSL licence declared.

If you care how your content is reused, say so where machines can read it. A footer copyright line is not machine-readable.

HOW TO FIX IT - CONFIGURATION ONLY

Declare the licence on the page and in your structured data, or publish an RSL licence if you intend to charge for reuse.

```
<!-- The simple version: -->
<link rel="license" href="https://example.com/terms">

<!-- Or Really Simple Licensing, which can state terms and a price: -->
<link rel="license" type="application/rsl+xml" href="https://example.com/license.xml">

<!-- and, in /robots.txt: -->
<!-- License: https://example.com/license.xml -->
```

---

WARN

## Publishes security.txt

+1.0 pts

No /.well-known/security.txt.

It is the standard machine-readable place to say who to contact. Agents and researchers both look there first.

HOW TO FIX IT - CONFIGURATION ONLY

Add a security.txt with a contact address and an expiry date.

Where: /.well-known/security.txt

```
Contact: mailto:security@example.com
Expires: 2027-01-01T00:00:00.000Z
Preferred-Languages: en
```

---

WARN

## Says when the content changed

+0.8 pts

Only one freshness signal present.

Agents deprioritise content they cannot date, and re-fetch undated pages more often, which costs you bandwidth for nothing.

HOW TO FIX IT - CONFIGURATION ONLY

Send Last-Modified and ETag, and publish dateModified in your structured data.

```
{"@type":"Article","datePublished":"2026-01-15","dateModified":"2026-08-01"}
```

---

WARN

### Declares a link preview card

+0.8 pts

No twitter:* card tags.

Several agents fall back to card metadata when they cannot extract the body reliably.

HOW TO FIX IT - CONFIGURATION ONLY

Add a summary_large_image card.

```
<meta name="twitter:card" content="summary_large_image">
<meta name="twitter:title" content="Page title">
<meta name="twitter:description" content="One sentence.">
```

WARN

### robots.txt is present and parseable

+0.4 pts

No robots.txt served (HTTP 404).

Without a robots.txt every crawler applies its own defaults and you have no way to express a preference. It is also the file most agents read first.

HOW TO FIX IT - CONFIGURATION ONLY

Publish a robots.txt that allows answer engines and points at your sitemap.

Where: /robots.txt

```
User-agent: *
Allow: /

Sitemap: https://example.com/sitemap.xml
```

WARN

### Declares a correct content type

+0.3 pts

Content-Type: text/html - charset is missing.

Without a declared charset, accented and non-Latin text can be mangled in AI answers that quote you.

HOW TO FIX IT - CONFIGURATION ONLY

Append `; charset=utf-8` to the Content-Type header.

21 PASSING

## What this page already gets right

PASS Page responds to a plain HTTP request

PASS No bot wall in front of the content

PASS The status code matches what the page says

PASS Answer engines are allowed to crawl

PASS User-triggered fetchers are allowed

PASS Autonomous agents are allowed

PASS Retrieval providers are allowed

PASS Classic search crawlers are allowed

PASS No directive suppressing AI use of this page

PASS Short redirect chain

PASS Responds fast enough for an agent budget

PASS HTML payload stays inside an agent context budget

PASS HTML is compressed on the wire

PASS Re-fetching is cheap for a crawler

PASS Exactly one <h1> naming the page

PASS Passages make sense on their own

PASS Signal-to-markup ratio

PASS Declares a document language

PASS Text arrived intact, not mangled

PASS URL describes the content

PASS Navigation uses real links

23 checks did not apply to this page and were excluded from the score entirely: Coding agents can read the documentation; Real agent user-agents are served normally; The page describes itself consistently; Answers questions in a form that can be quoted; Images carry alt text; Uses semantic elements, not div soup; Provides a no-JavaScript fallback; Structured data parses cleanly; Structured data describes the right thing; Structured data agrees with the page; Publishes a feed of its content; Declares its language alternates; Prices are machine-readable; Tabular data is marked up as a table; Dates are machine-readable; Declares its place in the site; Inline microdata annotations; Link text describes the destination; Form fields are labelled; Form fields have stable names; Fields declare autocomplete tokens; Controls are real buttons; llms.txt is actually useful.

CRAWLER ACCESS

# Who can actually reach this page

Resolved from robots.txt for this exact path, one crawler at a time. Blocking a training crawler is a legitimate business decision. Blocking an answer engine removes you from the results that would have linked to you.

## Answer engine

Indexes you so an assistant can cite and link you. Blocking it removes you from AI answers.

| | | |
|---|---|---|
| OAI-SearchBot | OpenAI | No rule |
| Claude-SearchBot | Anthropic | No rule |
| PerplexityBot | Perplexity | No rule |
| Applebot | Apple | No rule |
| meta-webindexer | Meta | No rule |
| Amazonbot | Amazon | No rule |
| DuckAssistBot | DuckDuckGo | No rule |
| Bravebot | Brave | No rule |
| YouBot | You.com | No rule |
| TongyiBot | Alibaba | No rule |
| PetalBot | Huawei | No rule |
| YandexAdditional | Yandex | No rule |

## User-triggered

Fetches your page because a human asked an assistant to open it right now. Blocking it breaks a live request.

| | | |
|---|---|---|
| ChatGPT-User | OpenAI | No rule |
| Claude-User | Anthropic | No rule |
| Perplexity-User | Perplexity | No rule |
| MistralAI-User | Mistral | No rule |
| Google-NotebookLM | Google | No rule |
| GoogleAgent-URLContext | Google | No rule |
| kagi-fetcher | Kagi | No rule |
| meta-externalfetcher | Meta | No rule |
| Kimi-User | Moonshot AI | No rule |
| Amzn-User | Amazon | No rule |
| cohere-ai | Cohere | No rule |

## Autonomous agent

Browses, compares and buys on a person's behalf. Blocking it removes you from the shortlist before anyone sees it.

| | | |
|---|---|---|
| ChatGPT Agent | OpenAI | No rule |
| Operator | OpenAI | No rule |
| GoogleAgent-Mariner | Google | No rule |
| Gemini-Deep-Research | Google | No rule |
| AmazonBuyForMe | Amazon | No rule |
| NovaAct | Amazon | No rule |
| Manus-User | Butterfly Effect | No rule |

## Retrieval provider

Sells the index that other companies' AI products search. Blocking it removes you from all of them at once.

| | | |
|---|---|---|
| ExaSearchBot | Exa | No rule |
| TavilyBot | Tavily | No rule |

| | | |
|---|---|---|
| `FirecrawlAgent` | Firecrawl | **No rule** |
| `ShapBot` | Parallel | **No rule** |
| `Diffbot` | Diffbot | **No rule** |
| `Cloudflare-AutoRAG` | Cloudflare | **No rule** |
| `bedrockbot` | Amazon | **No rule** |

### Classic search

Traditional search indexing, increasingly the substrate AI answers are built on.

| | | |
|---|---|---|
| `Googlebot` | Google | **No rule** |
| `Bingbot` | Microsoft | **No rule** |

### Coding agent

Reads documentation while a developer is building against you. Blocking it is how your API gets used wrong.

| | | |
|---|---|---|
| `Claude-Code` | Anthropic | **No rule** |
| `Cursor` | Cursor | **No rule** |
| `Google-Gemini-CLI` | Google | **No rule** |
| `Devin` | Cognition | **No rule** |
| `opencode` | opencode | **No rule** |
| `Trae` | ByteDance | **No rule** |

### Training crawler

Collects content for model training. Blocking it is a legitimate business choice with no traffic cost.

| | | |
|---|---|---|
| `GPTBot` | OpenAI | **No rule** |
| `ClaudeBot` | Anthropic | **No rule** |
| `Google-Extended` | Google | **No rule** |
| `Applebot-Extended` | Apple | **No rule** |
| `meta-externalagent` | Meta | **No rule** |
| `CCBot` | Common Crawl | **No rule** |
| `Bytespider` | ByteDance | **No rule** |
| `DeepSeekBot` | DeepSeek | **No rule** |
| `AI2Bot` | Ai2 | **No rule** |

---

WHAT SURVIVES BEING CUT UP

## The pieces an assistant is actually handed

No assistant reads a whole page. A retrieval system cuts it into passages of a few hundred tokens and hands back one of them, and the answer is written from that passage alone. This page cuts into 1 passages of about 300 tokens; 100% of its own content sits in a passage that can be read on its own, and 0% of what would be retrieved is navigation rather than content.

---

EVIDENCE

## What we actually received

Every number in this report comes from one plain HTTP response. No browser, no JavaScript, no retries.

| | |
|---|---|
| Final URL | `https://example.com/` |
| HTTP status | `200` |

| | |
|---|---|
| Response time | 95 ms |
| HTML size | 559 B |
| Inline script | 0 B (0% of payload) |
| Readable words | 19 |
| Text-to-markup | 22.7% |
| Title | Example Domain |
| Language | en |
| Canonical | (none) |
| Detected stack | (none detected) |
| robots.txt | not found |
| sitemap.xml | not found |
| llms.txt | not found |

---

Prepared by Northwind Digital. Measured with AXRAY against the AX specification, which is published in full with every weight at https://axray.online/ax-spec - anybody can re-run the same scan, which is what makes the number worth quoting.