

# Markdown Mirrors

Give AI a clean version of every page on your site to read.

Built by [@brycenwood.ai](#) · October 2026

---

OK so this is the second piece of the AI visibility puzzle. The `llms.txt` file you built in System 2 tells AI models *what* your business is. But when AI tries to read individual pages on your site to quote you, it has to wrestle through navigation, popups, scripts, cookie banners, and all the chrome a normal website has for humans. Markdown mirrors solve that.

A markdown mirror is a clean, plain-text version of every page on your website that lives at a predictable URL — just add `/index.md` to any page. Anthropic's docs site does this. Stripe's docs site does this. Almost no small businesses do this yet. That's the opportunity.

---

## OVERVIEW

### What You're Building

One Python script that walks your website folder and, for every HTML page, generates a sibling `index.md` file containing only the actual content. Plus a small Netlify config tweak so those `.md` files render as plain text in browsers (instead of forcing a download), and an update to your `llms.txt` so AI knows the mirrors exist.

Walk Site Folder



Strip Chrome



Convert to Markdown



Save index.md



Serve as text/plain



Tell AI About It

#### WHY THIS STACKS WITH LLMS.TXT

llms.txt is the pitch — the elevator overview for your whole business. Markdown mirrors are the per-page detail. When AI cites you, it can quote from the right page directly. You want both. The two together turn your site into a fully AI-readable surface.

#### BEFORE YOU START

## What You Need

- **Claude Code** installed (the CLI tool, not `claude.ai`)
- **A static website** hosted on Netlify, Vercel, GitHub Pages, Cloudflare Pages, or any host that lets you upload a config file. If you're on Squarespace or Wix, you'll need to ask Claude Code for the workaround for your specific host.
- **Your website code in a folder** you can hand to Claude Code. If your site is in a GitHub repo, that's perfect.
- **Python 3 installed** with the `beautifulsoup4` and `markdownify` packages (Claude Code will install these for you if you ask).

#### TIME + COST

End to end this took me about an hour, including testing. Cost: \$0 ongoing. The script is one Python file you run whenever your site updates. You can wire it into a pre-commit hook later if you want it fully automated.

#### STEP 1

## Build the Generator Script

Open Claude Code in your website project folder. Paste this prompt. It will write a Python script that walks every `index.html` file in your site, strips the chrome, and writes a clean markdown version next to it.

#### PASTE THIS INTO CLAUDE CODE

I want to build markdown mirrors for every page on my website. The goal is to give AI tools (ChatGPT, Claude, Perplexity) a clean version of each page they can read without wrestling with HTML, scripts, or chrome.

Write me a Python script that:

1. Walks my website folder and finds every `index.html` file (skip 404s and any `/thanks/` pages that are `noindex`)
2. For each page, parses the HTML with BeautifulSoup
3. Strips out: `nav`, `footer`, `scripts`, `styles`, `noscripts`, chat widgets, GHL/HubSpot widgets, `iframes`, and any element with a class matching `nav`, `footer`, `cta-split`, or starting with `ghl`
4. Drops empty `div/span` wrappers that have no text content
5. Converts the remaining body HTML to clean markdown using the `markdownify` package
6. Cleans the markdown output: collapses 3+ blank lines, strips standalone "01" "02" step numbers, removes bullet separator characters, removes empty image alt tags
7. Writes the result to `{page_dir}/index.md` with frontmatter at the top:

```
---
title: [page title from <title>]
description: [meta description]
url: [canonical URL]
last_updated: [today's date]
---
```

8. Prints a summary at the end showing how many pages were generated

Save it as `generate_markdown_mirrors.py` in my scripts folder. Make it re-runnable so I can rerun it whenever my site updates. After you write it, run it once and show me the output.

#### WHY FRONTMATTER MATTERS

The frontmatter at the top of each markdown file gives AI extra context — title, description, the canonical URL. When AI quotes the page, it knows where the content came from. Don't skip it.

## STEP 2

# Configure Your Host to Serve .md as Plain Text

By default, most web hosts treat .md files as a download — when you visit one in a browser, it downloads instead of displaying. That's bad for two reasons: humans can't preview them, and some AI fetchers stumble on the content type. You want them to render inline as plain text.

### PASTE THIS INTO CLAUDE CODE

```
I'm hosted on [YOUR HOST – Netlify / Vercel / Cloudflare Pages /
GitHub Pages / Squarespace / etc].
```

```
I just generated markdown mirror files at {page}/index.md across my
site. I need every .md file to be served with Content-Type: text/plain
so they render in the browser instead of triggering a download.
```

```
Show me exactly which config file I need to edit on my host and the
exact lines to add. Then explain how to test it after deploy.
```

```
For Netlify, the answer is the _headers file. For Vercel, it's
vercel.json. For Cloudflare Pages, it's _headers also. Walk me through
whichever one applies to me.
```

### FOR NETLIFY USERS SPECIFICALLY

```
Add this to your _headers file:
```

```
/*.md
```

```
Content-Type: text/plain; charset=utf-8
```

```
Cache-Control: public, max-age=3600
```

```
X-Robots-Tag: index, follow
```

## STEP 3

# Tell AI the Mirrors Exist

You already have an llms.txt file from System 2. Add a section to it that lists every markdown mirror URL so AI fetchers know where to find them.

#### PASTE THIS INTO CLAUDE CODE

Open my llms.txt file. Add a new section called "Markdown Mirrors" right above the "What Makes Us Different" section. Inside that new section, list every index.md URL on my site. The format should be:

```
## Markdown Mirrors (Clean AI-Readable Versions)
```

Every page on this site has a plain markdown mirror. Add /index.md to any URL to get the clean content without navigation, scripts, or layout chrome.

```
- https://yourdomain.com/index.md
- https://yourdomain.com/about/index.md
- https://yourdomain.com/services/index.md
(etc, one per page)
```

Generate the full list automatically by reading my website folder. Skip any noindex pages (thanks pages, 404s).

#### STEP 4

## Test It

After you push the changes and your site redeploys, run through this test:

1. **Open one of your mirror URLs in a browser** — for example `yourdomain.com/business-wraps/index.md`. It should render as clean plain text, not download.
2. **Curl the URL from terminal** — `curl -I https://yourdomain.com/business-wraps/index.md`. The Content-Type header should say `text/plain`.
3. **Paste a mirror URL into ChatGPT** — ask "what does this business do?" ChatGPT will fetch the URL and return a clean summary. If it nails it, your mirrors work.
4. **Paste the same URL into Claude or Perplexity** — same test. Verify all three platforms can read it cleanly.

#### IF THE TEST FAILS

If the .md still downloads instead of displaying, the Content-Type header didn't take. Check that `_headers` is in your site's root and the format is exact (the indentation matters on Netlify). Redeploy and test again.

## STEP 5

# Keep It Updated

Whenever you update a page on your site, you'll want to regenerate that page's markdown mirror so AI is reading the latest version. Two ways to handle this:

## Option A — Manual rerun

Just run `python3 generate_markdown_mirrors.py` whenever you update content. Takes about 2 seconds. Commit + push. Done.

## Option B — Automated via pre-commit hook

### PASTE THIS INTO CLAUDE CODE

I want my markdown mirrors to regenerate automatically whenever I commit changes to my website. Set up a pre-commit hook that:

1. Runs `generate_markdown_mirrors.py` every time I git commit
2. Adds any newly generated/updated `.md` files to the commit automatically
3. Doesn't fail the commit if the script has any non-fatal warnings

Walk me through the `.git/hooks/pre-commit` setup for my repo. Make it so I never have to think about it again.

## MY RESULTS

# What Happened After I Shipped This

- **27 markdown mirrors** generated across my site in about 2 seconds
- **Total time to build + ship:** ~1 hour from "what's a markdown mirror" to live in production
- **Cost:** \$0 (just the Claude Code subscription I already had)
- **ChatGPT, Claude, and Perplexity** can now read each page in one clean fetch with no HTML noise
- **The video about it** hit my biggest reach yet within hours of posting

### REAL TALK

This is one of those "almost nobody is doing this yet" plays. In 12 months it'll be standard. Right now you can ship it in an hour and be early. That's the whole game with AI visibility — be early, be specific, give the machines something clean to quote.

### TIPS

## Things I Learned Building This

1. **Keep the markdown clean.** If your generator leaves visual chrome like "01" "02" or bullet separator characters in the output, AI will get tripped up. Strip aggressively.
2. **Frontmatter helps AI.** The title/description/URL block at the top of each .md file gives the AI context. Don't skip it.
3. **Don't change the URL pattern later.** Once you decide on /page/index.md or /page.md, stick with it. Changing breaks any AI fetcher that already learned where to look.
4. **Update llms.txt every time you add a new page.** Otherwise AI won't know the new mirror exists. Or wire that into your generator script.
5. **Test with all three: ChatGPT, Claude, Perplexity.** They have different fetchers and sometimes one will work while another won't. You want all three reading you.
6. **Don't gate the .md files with login or paywall.** The whole point is for AI to read them freely. If you need gating, gate the HTML version, not the markdown.

# That is the whole system.

Zero ad spend. Zero coding experience. The other playbooks  
live at [brycenwood.com/playbook](https://brycenwood.com/playbook), one download each.

---

Built by Brycen Wood · [@brycenwood.ai](https://twitter.com/brycenwood.ai) · [brycenwood.com](https://brycenwood.com)