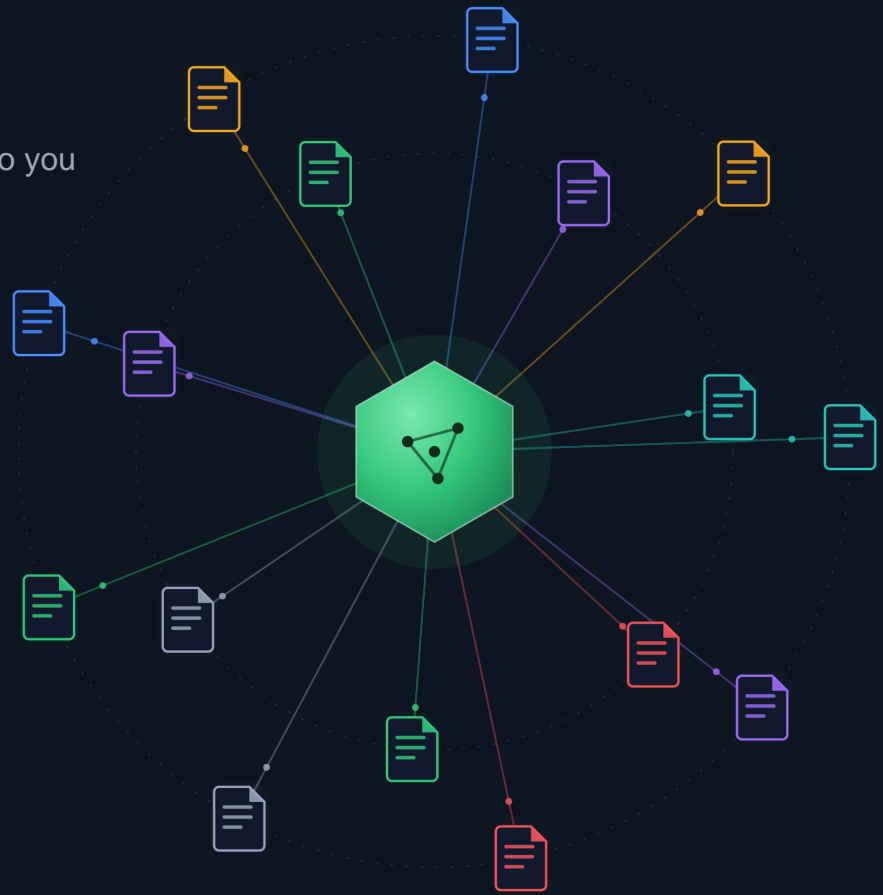


SILVERBACK MARKETING PRESENTS

# AI READINESS GUIDE

A structured kit of files that tells ChatGPT, Claude, Gemini, and Perplexity exactly who you are before they guess wrong.



# What Is This Guide?

The world changed when AI assistants became the first stop for research. Today, when someone wants to find the best product to buy, the right service for their needs, or answers to a specific question, they often ask ChatGPT, Claude, Gemini, or Perplexity before they ever visit a website.

This creates a real problem. AI systems learned what they know from training data collected months or even years ago. If your website was not set up to communicate clearly with AI, the AI might describe your products incorrectly, quote outdated pricing, or direct users to a competitor that was better indexed.

**AI Readiness Files** solve this problem. They are a set of structured files you deploy to your website that tell AI systems, in formats they understand natively, exactly who you are, what you offer, and where to send people.

This guide explains all 17 files in plain English. No technical background required. Each file gets its own page with a plain-English explanation, a real-world analogy, and a direct answer to the question: what does an AI actually do with this file?

---

## The 7 Categories

The 17 files fall into 7 natural categories, each playing a distinct role in helping AI understand and accurately represent your website:

<b>Identity &amp; Permissions</b>	The files that introduce your brand and control which AI systems get access
<b>Content Files</b>	The files that give AI the full story about your site in readable text
<b>Map &amp; Navigation</b>	The files that help AI navigate your site structure efficiently
<b>Intelligence Files</b>	The files that give AI deep knowledge of your entities, products, and user intents
<b>Research Files</b>	The files that power AI research and retrieval pipelines
<b>Policy Files</b>	The files that set the rules for how AI can use your content
<b>Operations Files</b>	The files that help your team deploy and maintain everything

# Contents

## — Identity & Permissions

- `robots.txt` — The Doorman
- `ai.txt` — Your AI Business Card

## — Content Files

- `llms.txt` — The Cheat Sheet
- `llms-full.txt` — The Deep Dive

## — Map & Navigation

- `ai-sitemap.xml` — The GPS
- `sitemap.md` — The Human-Readable Map

## — Intelligence Files

- `ai-entities.json` — The Encyclopedia
- `ai-intent.json` — The Traffic Director
- `ai-schema.json` — The Identity Card
- `.well-known/ai-plugin.json` — The Plugin Badge

## — Research Files

- `rag-index.json` — The Research Database
- `rag-index.jsonl` — The Streamlined Database

## — Policy Files

- `ai-disclosure.txt` — The Transparency Report
- `training-data-policy.txt` — The License Agreement

## — Operations Files

- `structured-data-guide.md` — The Developer Handbook
- `manifest.json` — The Master Inventory
- `deployment-checklist.md` — The Launch Plan

# robots.txt

## ◆ THE DOORMAN

Controls who gets in and where they can go

PRIORITY	WHO READS IT	UPDATE WHEN
<b>Critical</b>	<b>All web crawlers and AI bots</b>	<b>When site structure changes</b>

**robots.txt** is one of the oldest files on the web, first introduced in 1994. Every website has one, or at least should have one. It sits at the front door of your website and tells automated visitors — search engines, AI bots, data scrapers — which sections of the site they are welcome to access.

Think of your website as a hotel. The robots.txt file is the **hotel doorman**. It greets every bot that arrives and hands them a list with certain areas marked as staff only. A bot that follows the rules will never enter your checkout pages, private account sections, or backend admin tools.

For AI readiness, robots.txt gets a meaningful upgrade. It now includes specific instructions for each major AI crawler, including ChatGPT's GPTBot, Anthropic's ClaudeBot, Google's extended AI crawler, and others. You can tell them exactly which files to read, like your llms.txt and ai-sitemap.xml, and which pages to leave alone entirely.

**Why does this matter?** Without clear rules, AI systems can index your checkout flow, login page, or outdated promotional content. They then repeat that incorrect information confidently to real users. With a properly configured robots.txt, you are giving AI a guided tour of only the content that represents you well.

### THINK OF IT LIKE...

*Imagine a library where some shelves are open to visitors and others are for staff only. robots.txt is the sign at the entrance that tells each visitor which sections they are free to browse.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- Checks it before crawling any page on your site
- Follows Allow and Disallow rules to determine which pages to index
- Discovers your AI-specific files through Sitemap references listed inside
- Learns where to find your llms.txt, ai-sitemap.xml, and other readiness files

# ai.txt

## ✦ YOUR AI BUSINESS CARD

Introduces your brand, products, and ground rules to every AI system

PRIORITY	WHO READS IT	UPDATE WHEN
<b>Critical</b>	<b>AI crawlers, LLM training pipelines</b>	<b>Quarterly, or when products change</b>

**ai.txt** is the newest member of the rules-for-robots family, and it is purpose-built for the age of AI. Where robots.txt simply says which pages are accessible, ai.txt goes much further. It is your brand's complete introduction to every AI system on the internet.

Picture it as the **ultimate business card**. Instead of your name and phone number, it contains your entire company identity: what you sell, what you are known for, which topics you speak authoritatively on, and what AI systems are and are not permitted to do with your content.

ai.txt might tell an AI system something like this: We are a cooperative selling outdoor gear. Our main products are camping equipment, hiking gear, and bicycles. Our authoritative topics include tent selection, backpacking safety, and trail skills. You may use our content for RAG search indexing, but commercial model training requires a license agreement.

**Why does this matter?** Without ai.txt, an AI system trying to learn about your company has to guess. It might describe you inaccurately, attribute products you do not carry, or miss your biggest differentiators entirely. This file gives AI systems a single authoritative source of truth, written by you.

### THINK OF IT LIKE...

*Think of ai.txt like the briefing document you would hand a journalist before a media interview. It covers everything you want them to know, your name, your story, your products, and your ground rules, so they can represent you accurately.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- Reads it to learn your brand identity, product categories, and authoritative topics
- Uses it to decide how to represent your company in AI-generated answers
- Follows your stated training data policy for what it can and cannot do with your content
- Uses your own brand language and product names rather than making educated guesses

# llms.txt

## ◆ THE CHEAT SHEET

A quick-read summary of your entire website, written for AI

PRIORITY	WHO READS IT	UPDATE WHEN
<b>Critical</b>	<b>LLMs including ChatGPT, Claude, Gemini, and Perplexity</b>	<b>Monthly, or when site structure changes</b>

**llms.txt** is a recently established standard from llmstxt.org, designed specifically for large language models — the AI brains powering ChatGPT, Claude, Gemini, and Perplexity. It is a structured text file that gives AI a complete map of your website in a format it can read and understand in seconds.

Imagine you need to brief a new assistant on your entire company before their first customer call. You would not hand them a stack of 500 web pages. You would give them a **cheat sheet** — organized, scannable, covering all the highlights. That is exactly what llms.txt does.

The file follows a specific format. Your company name sits at the top, followed by a short description paragraph, then organized sections for each part of your website including products, services, blog, and support, each with direct links. At the bottom, you list the most common questions people ask about your company so AI knows which queries your site is best equipped to answer.

**Why does this matter?** When someone asks ChatGPT about your products, the AI might generate an answer from training data that is months or years out of date. When it can instead read your current llms.txt, it gives an accurate, up-to-date answer and links directly to your real pages.

### THINK OF IT LIKE...

*llms.txt is like the table of contents and executive summary of your website combined into one short document that an AI can read in seconds, rather than crawling thousands of individual pages.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- Reads it to get an accurate, structured map of your entire website
- Uses your section headers to understand how your site content is organized
- Follows your links to find the right pages to recommend to users
- Uses your priority queries to understand which questions your site answers best

# llms-full.txt

## ◆ THE DEEP DIVE

The extended version, rich context for AI systems that need the full story

PRIORITY	WHO READS IT	UPDATE WHEN
<b>Critical</b>	<b>RAG pipelines, AI knowledge bases, LLMs</b>	<b>Quarterly, or when major product changes occur</b>

**llms-full.txt** is the extended edition of llms.txt. While the standard version is a quick cheat sheet, llms-full.txt is the complete briefing document. It includes rich descriptions, detailed Q and A sections, full category explanations, and everything an AI system needs to speak knowledgeably about your organization.

Put simply: llms.txt is the back-of-book summary. llms-full.txt is the **full book**.

Where llms.txt might say Camping and Hiking with a link, llms-full.txt goes into genuine depth: covering what gear categories exist, how they are organized, what questions buyers commonly have, and authoritative answers to each one. This richness is especially valuable for AI-powered search engines like Perplexity and for RAG pipelines that pull context from multiple sources before generating an answer.

**Why does this matter?** An AI that can only answer shallow questions about your company is far less useful than one that can answer nuanced questions with confidence. llms-full.txt is what makes that depth of knowledge possible.

### THINK OF IT LIKE...

*If llms.txt is a Wikipedia article summary, llms-full.txt is the complete Wikipedia article with all sections, details, and references included.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- Reads it to generate rich, accurate answers about your products and services
- Uses your Q and A sections to respond to common user questions correctly
- Feeds it into RAG pipelines to create high-quality context for AI assistants
- Uses your detailed descriptions to match the right products to the right users

# ai-sitemap.xml

## ◆ THE GPS

A map of your website with AI-specific directions built right in

PRIORITY	WHO READS IT	UPDATE WHEN
High	AI crawlers, semantic search indexers	Monthly, or when new pages are added

**ai-sitemap.xml** is an upgraded version of the traditional XML sitemap that search engines have relied on for years. Every website should have a sitemap — it is a structured list of every page with timestamps and priority scores. The ai-sitemap takes that foundation and extends it specifically for AI systems.

A regular sitemap tells search engines which pages exist and when they were updated. The ai-sitemap tells AI systems the same thing, but also describes what each page is about, what type of content it contains, and provides a plain-English summary of what users will find there.

Think of it as the difference between a paper road map and a **GPS with points of interest**. The road map shows the streets. The GPS tells you what is at each destination before you arrive. Each URL in the ai-sitemap carries additional tags: a content-type label, a topic description, and a one-sentence summary that removes all ambiguity about what the page contains.

**Why does this matter?** Without this context, an AI crawler reading your products page might not know if you are selling items, reviewing them, or publishing editorial content about them. The ai-sitemap removes all of that uncertainty.

### THINK OF IT LIKE...

*The ai-sitemap.xml is like a museum floor guide that does not just show you which rooms exist, but describes every exhibit inside each one so visitors know exactly what they will find before they walk in.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- Uses it to efficiently discover and index all major pages on your site
- Reads content-type and topic tags to correctly categorize each page
- Uses summaries to understand pages without needing to fetch and parse each one
- Prioritizes pages based on your importance scores, starting with the most valuable

# sitemap.md

## ◆ THE HUMAN-READABLE MAP

Your sitemap in plain text, readable by both people and AI equally

PRIORITY	WHO READS IT	UPDATE WHEN
<b>Medium</b>	<b>LLMs, developers, content teams</b>	<b>Monthly, or when site structure changes</b>

**sitemap.md** is the most approachable file in the AI readiness set, and that simplicity is intentional. It is a plain-text, markdown-formatted overview of your website's structure. No XML tags. No JSON syntax. Just organized sections, links, and descriptions that anyone or any AI can read and understand immediately.

While ai-sitemap.xml is built for machines, sitemap.md is built for **humans and AI equally**. A developer can open it and immediately understand how the site is structured. An AI system can read it and instantly grasp how different sections relate to each other.

The file is organized into sections: your main product or service categories, resources and blog content, company pages, and your AI readiness files. Each section lists the key pages with their URLs and brief descriptions, written in plain conversational language.

**Why does this matter?** For AI systems that work by reading text rather than parsing XML or JSON, sitemap.md is the most accessible starting point for understanding your site. It is also the first document a new developer or content team member reaches for when they need a quick mental model of how everything fits together.

### THINK OF IT LIKE...

*If your website were a city, sitemap.md would be the welcome brochure handed out at the visitor center — organized, friendly, and giving you just enough information to navigate with confidence.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- Reads it to build an immediate mental model of your website structure
- Uses section organization to understand how your content is grouped
- References it when generating navigation suggestions for users
- Shares it with content and developer teams as a living site reference document

# ai-entities.json

## ✦ THE ENCYCLOPEDIA

A structured catalog of everything your website contains

PRIORITY	WHO READS IT	UPDATE WHEN
High	AI systems, knowledge graph builders	Quarterly, or when categories change

**ai-entities.json** is a structured catalog of every significant element on your website — every product category, subcategory, service, brand, and key concept — organized in a format that AI systems can read and reason about with precision.

In AI terminology, entities are the key nouns that matter in your world. For a sporting goods retailer, entities include Camping and Hiking, Sleeping Bags, brand names like The North Face, Co-op Membership, and Expert Advice articles. For a university, they might be Computer Science, Bachelor's Degree programs, Fall Semester, and Financial Aid.

Think of **ai-entities.json** as your company's **formal encyclopedia**. It is not written in prose. It is structured data: each entity carries a name, type, description, related URLs, and connections to related entities. This structure powers the knowledge graph that AI systems build internally to understand your world.

**Why does this matter?** Without entity data, AI systems make their best guesses about how things relate. With it, they can say with confidence that sleeping bags are a subcategory of camping gear, which your company sells at a specific URL — because you told them exactly that.

### THINK OF IT LIKE...

*ai-entities.json is like the index of a comprehensive encyclopedia. Instead of page numbers, each entry links to a URL, includes a description, and connects to related topics in a way any search system can understand.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- Builds a knowledge graph of your products, services, and categories from it
- Correctly classifies what your site sells and how everything is organized
- Understands relationships between categories, subcategories, and related items
- Provides accurate product and service names rather than paraphrased approximations

# ai-intent.json

## ✦ THE TRAFFIC DIRECTOR

Maps every question someone might ask to the right page on your site

PRIORITY	WHO READS IT	UPDATE WHEN
High	AI assistants, conversational AI, AI search engines	Quarterly, or when site structure changes

**ai-intent.json** is one of the most immediately useful files in the entire AI readiness set. It is a lookup table that maps real user questions — the actual things people type into AI assistants — to the best page on your website to answer each one.

Every day, millions of people ask AI assistants questions like how does this company's membership work, can I rent gear near me, or which product is best for beginners. Without ai-intent.json, the AI guesses which page of your site to recommend. With it, you have given the AI a direct, authoritative answer: here is the exact URL.

Think of it as a **traffic director** standing at the crossroads between user questions and your website. When a question arrives, the traffic director points immediately to the correct destination. No detours, no guessing, no sending someone to your homepage when they needed the pricing page.

**Why does this matter?** AI systems that know precisely where to send users are dramatically more useful than ones that only describe your products in general terms. This file transforms AI from an information source into a genuine navigation tool for your site.

### THINK OF IT LIKE...

*ai-intent.json is like a knowledgeable store employee who knows the answer to every common customer question and knows exactly which aisle, shelf, or department to point that customer toward.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- Matches incoming user queries to the correct page on your site
- Routes how-to questions to guides and educational content
- Routes purchase-intent questions to the right product category pages
- Sends users to the most relevant page rather than defaulting to your homepage

# ai-schema.json

## ✦ THE IDENTITY CARD

A structured identity file that search engines and AI both understand perfectly

PRIORITY	WHO READS IT	UPDATE WHEN
High	Search engines, AI systems, developers	Annually, or when company information changes

**ai-schema.json** uses the Schema.org standard — a vocabulary developed jointly by Google, Microsoft, Yahoo, and Yandex — to describe your organization in a way that both search engines and AI systems understand with complete precision.

Where ai.txt describes your company in plain English, ai-schema.json describes it in a highly structured, machine-readable format. It defines exactly what type of organization you are, your founding date, your address, your social media profiles, your phone number, and even your site search functionality. All of this follows a format that has been standardized across the entire web for over a decade.

Think of it as your company's **official identity card** with a machine-readable code on the back. When a system scans it, it instantly knows exactly who you are and retrieves all your key details with zero ambiguity. This same structured data, known as JSON-LD, is also embedded directly in your website's code to power Google's rich results in search.

**Why does this matter?** When an AI is trying to determine whether two different mentions of your company name across the web refer to the same organization, ai-schema.json provides the definitive answer. It is your source-of-truth identity file.

### THINK OF IT LIKE...

*ai-schema.json is like your company's official government registration filing — a formal, standardized record of who you are, what you do, and where to find you, recognized and understood by every system that reads it.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- Uses it to confidently identify your organization across different web properties
- Feeds the structured data into knowledge graphs and entity resolution systems
- Powers enhanced search results with star ratings, hours, and contact details
- Links your website, social profiles, and other online presence as one unified entity

# .well-known/ai-plugin.json

## ✦ THE PLUGIN BADGE

Registers your site with AI assistant platforms as an official resource

PRIORITY	WHO READS IT	UPDATE WHEN
Medium	AI assistants, AI plugin registries, developer platforms	When contact information or product categories change

**ai-plugin.json** lives in a special folder on your website called **.well-known**. This folder is a web standard for machine-readable identity files. It is where browsers, security tools, and now AI systems look for official declarations about a site and what it is capable of.

The ai-plugin.json file draws from the plugin format that became widely recognized through AI assistant platforms. It declares that your website is available as an AI resource, meaning an AI assistant can officially use your site to answer user queries about your products, services, or content.

Think of it as your **official membership badge** in the AI ecosystem. It registers your site's name, description, contact information, product categories, and key URLs with AI platforms in a standardized format they all recognize. When an AI system is deciding which sources to use for a query, having a valid ai-plugin.json is like raising your hand with a verified credential.

**Why does this matter?** As AI assistant platforms continue evolving toward official integrations with websites, ai-plugin.json positions your site to be recognized and used as a first-class official source rather than just another scraped result from somewhere on the internet.

### THINK OF IT LIKE...

*ai-plugin.json is like registering your business with the Chamber of Commerce. It is your official membership in the AI ecosystem, complete with a standardized profile that every participating system recognizes.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- Discovers it automatically when checking the .well-known directory
- Uses your declared capabilities to understand which queries your site can help with
- References your contact information and key URLs for routing users appropriately
- Registers your site as an official, trustworthy AI resource on the open web

# rag-index.json

## ◆ THE RESEARCH DATABASE

A ready-made index of your site, built for AI research pipelines

PRIORITY	WHO READS IT	UPDATE WHEN
High	RAG pipeline engineers, AI search builders	Monthly, or when new major pages are added

**rag-index.json** is designed for one of the most powerful AI applications: RAG, which stands for **Retrieval-Augmented Generation**. RAG is the technology that lets AI systems deliver accurate, current answers by looking information up in real time rather than relying only on what they learned months ago during training.

Here is how RAG works in practice. Instead of memorizing every fact, a RAG-enabled AI maintains a database of documents and URLs. When a user asks a question, the AI first retrieves the most relevant documents, then generates an answer based on what it just found. The result is answers that are accurate, timely, and grounded in real sources.

rag-index.json is your site's **pre-built contribution to that database**. It is a JSON array of records, typically one per major page or section, where each record contains the page URL, its title, and a list of topics it covers. AI systems can load this file directly into their RAG pipeline, instantly making your site a searchable knowledge source.

**Why does this matter?** Developers and companies building AI tools can plug rag-index.json directly into systems like LlamaIndex, LangChain, Pinecone, or Weaviate. Your content becomes a first-class source in their AI knowledge base without any extra work on their end.

### THINK OF IT LIKE...

*rag-index.json is like giving an AI researcher a pre-organized card catalog of your entire library. Instead of reading every book to understand what is available, they have a ready-made index they can search in milliseconds.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- Loads it directly into RAG pipelines like LlamaIndex, LangChain, or Pinecone
- Quickly identifies the most relevant pages for any given user query
- Builds a searchable knowledge base from your site's content
- Retrieves real-time, accurate information to ground its AI-generated answers

# rag-index.jsonl

## ◆ THE STREAMLINED DATABASE

The same research database in a format built for large-scale AI processing

<p>PRIORITY</p> <p><b>High</b></p>	<p>WHO READS IT</p> <p><b>ML engineers, AI developers, vector database ingestion tools</b></p>	<p>UPDATE WHEN</p> <p><b>Always keep in sync with rag-index.json</b></p>
------------------------------------	--	--

**rag-index.jsonl** is the identical twin of `rag-index.json` — the same records, the same content — but formatted as JSONL, which stands for **JSON Lines**. This format is also called newline-delimited JSON or NDJSON, and the difference from regular JSON is simple but practically significant.

`rag-index.json` is one large file with all records wrapped inside square brackets. `rag-index.jsonl` puts each record on its own individual line. This sounds like a minor technical detail, but it makes an enormous difference for large-scale AI processing pipelines.

Think of it like two ways to package a stack of index cards. `rag-index.json` binds all the cards into a book. You have to open the whole book to use any card. `rag-index.jsonl` leaves the cards loose so you can **pick them up and process them one at a time**, streaming through thousands of records without loading everything into memory at once.

**Why does this matter?** Both files exist because different AI tools prefer different formats. JSONL is the standard for OpenAI fine-tuning, LlamaIndex, LangChain, Pinecone, Weaviate, and most modern vector databases. Offering both ensures your site works with every AI pipeline tool, not just the ones that accept regular JSON arrays.

### THINK OF IT LIKE...

*If `rag-index.json` is a bound book containing all your records, `rag-index.jsonl` is those same records on individual index cards — the same information, but far easier to sort, filter, and process one entry at a time.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- Feeds directly into OpenAI fine-tuning datasets and similar ML workflows
- Streams into vector databases one record at a time for memory-efficient processing
- Powers LlamaIndex, LangChain, Pinecone, and Weaviate knowledge bases
- Enables large datasets to be processed without loading everything into memory

# ai-disclosure.txt

## ◆ THE TRANSPARENCY REPORT

A public statement of how your organization uses AI

PRIORITY	WHO READS IT	UPDATE WHEN
<b>Medium</b>	<b>AI systems, customers, regulators</b>	<b>Annually, or when AI tool usage changes</b>

**ai-disclosure.txt** is your organization's public statement about how you use artificial intelligence — in your products, your operations, and your customer interactions. It functions like a privacy policy for the AI era: a transparent, honest declaration of what AI is and is not doing on your behalf.

As AI becomes more common, customers and regulators increasingly want straight answers. Is this company using AI to generate content? Is AI making decisions about my account? Can I reach a real person? **ai-disclosure.txt** answers these questions clearly and proactively, before anyone has to ask.

Your disclosure might explain that you use AI for product recommendations but not for pricing decisions. Or that your customer service chat is AI-powered but always escalates to a human on request. Or that your written content is authored by people and reviewed by AI for accuracy. Whatever the reality is, this file states it plainly.

**Why does this matter?** Trust is the most valuable asset in the AI era. Organizations that are upfront about how they use AI build stronger relationships than those that obscure it. **ai-disclosure.txt** signals to customers, AI systems, and regulators that your organization is thoughtful and honest about AI.

### THINK OF IT LIKE...

*ai-disclosure.txt is like the nutrition label on food packaging. It does not tell you whether the food is good or bad. It simply tells you exactly what is inside. That transparency is what builds trust.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- Reads it to accurately describe your organization's AI usage to curious users
- Uses it to appropriately flag when content from your site is AI-assisted
- References it when users ask whether your company uses AI
- Evaluates your organization's transparency and trustworthiness based on it

# training-data-policy.txt

## ◆ THE LICENSE AGREEMENT

Sets the rules for how AI companies can use your content

PRIORITY	WHO READS IT	UPDATE WHEN
Medium	AI model developers, AI companies, legal teams	Annually, or when policy changes

**training-data-policy.txt** is your organization's formal position on one of the most pressing questions in the AI era: **Can AI companies use your content to train their models?**

Every day, AI companies crawl the web collecting text, images, and data to train the next generation of AI models. Your website's content — your product descriptions, blog posts, expert guides — may be included without your knowledge or consent. This file is your formal, published answer to when that is acceptable and when it requires permission.

A typical policy might read like this: Using our content for real-time RAG search and retrieval is permitted and benefits our customers. Using our content to train a commercial AI model for sale or licensing requires a formal license agreement. And regardless of use, our pricing information is time-sensitive and should never be cited as static fact.

**Why does this matter?** This is an active legal battleground right now. Several major media companies have filed lawsuits over AI companies training on their content without permission. A clear, published policy protects your organization and gives AI companies unambiguous guidance on what is and is not allowed.

### THINK OF IT LIKE...

*training-data-policy.txt is like a Creative Commons license for your website content in the AI era. It states clearly what others can do with your work so there is never any question about what is permitted.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- Reads it before deciding whether your content can be included in training datasets
- Treats content as license-required if your policy restricts commercial model training
- Respects your currency warnings to avoid citing stale pricing or inventory data
- Cites and links to your original source rather than reproducing your content directly

# structured-data-guide.md

## ◆ THE DEVELOPER HANDBOOK

Step-by-step guide for your technical team to implement AI-ready schema

PRIORITY	WHO READS IT	UPDATE WHEN
Low	Web developers, SEO teams, AI readiness teams	When new page types are added to the site

**structured-data-guide.md** is the most technical file in the AI readiness set, but it is not really written for AI systems. It is written for **your development team**. It is a comprehensive, step-by-step guide for implementing Schema.org structured data across every page type on your website.

Structured data, also called JSON-LD, is code that developers add to the head section of each web page. It tells search engines and AI systems exactly what kind of page they are looking at: a product listing, a blog post, an event calendar entry, or a store location page. This invisible layer of code powers Google's enhanced search results and increasingly powers AI systems' ability to understand page-level content.

The guide provides ready-to-use implementation examples for every major page type. How to mark up a product page with price and availability. How to mark up a store location with address and hours. How to mark up a blog post with author and publication date. Each example includes a complete JSON-LD code block, ready to copy, customize, and paste into production.

**Why does this matter?** Files like `llms.txt` and `ai-entities.json` help AI understand your site from the outside in. Schema.org structured data helps AI understand each individual page from the inside out, embedded directly in the page code itself.

### THINK OF IT LIKE...

*structured-data-guide.md is like the wiring diagram for your home. Guests never see it, but your electrician absolutely needs it to make sure everything behind the walls works correctly.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- Reads the structured data embedded on each page to understand its specific content type
- Uses Product and Offer schema to accurately describe your inventory and pricing
- Uses LocalBusiness schema to surface your locations in local and map-based searches
- Uses BlogPosting and Event schemas to correctly classify and date your published content

# manifest.json

## ◆ THE MASTER INVENTORY

A machine-readable index of every file in your AI readiness set

PRIORITY	WHO READS IT	UPDATE WHEN
Medium	Developers, deployment tools, AI readiness auditors	When new files are added to the set

**manifest.json** is the master inventory of your entire AI readiness implementation. It lists every file that has been deployed, what each one does, where it lives on your website, what format it uses, who the intended audience is, and how frequently it should be updated.

Think of it as the **packing list for an important shipment**. Before a box leaves the warehouse, every item is checked against the packing list. When the box arrives, the recipient verifies everything is present. **manifest.json** serves the same purpose for your AI readiness files.

It also functions as a living record of your organization's AI readiness status. Any developer, AI auditor, or SEO professional can retrieve your **manifest.json** and immediately see which files have been implemented, which URLs they are deployed to, what format each uses, and when each was last updated. The summary section provides an at-a-glance scorecard.

**Why does this matter?** As AI readiness becomes a recognized standard practice, **manifest.json** allows anyone — a developer, a partner, a potential client — to verify in seconds that your implementation is complete, current, and following best practices.

### THINK OF IT LIKE...

*manifest.json is like the table of contents for your entire AI readiness project. It tells anyone who opens it exactly what you have built, where everything lives, and how complete the implementation is.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- Reads it to discover all AI-specific files without needing to guess at URLs
- Uses priority rankings to determine which files to read and index first
- References update frequencies to know when to revisit and re-index each file
- Allows auditing tools to verify complete and correct AI readiness implementation

# deployment-checklist.md

## ◆ THE LAUNCH PLAN

A step-by-step checklist for getting your AI readiness files live

PRIORITY	WHO READS IT	UPDATE WHEN
<b>Internal</b>	<b>Web teams, developers, marketing teams</b>	<b>One-time setup; update when new files are added</b>

**deployment-checklist.md** is exactly what the name suggests: a complete, step-by-step checklist for deploying your AI readiness files to a live website. It is the playbook your web team follows to move from files generated on a computer to files actually serving traffic in the right place.

AI readiness files only deliver value when they are deployed correctly — sitting in the right location on your web server, returning the correct content type, and accessible to the AI crawlers that need them. This checklist makes sure nothing is skipped, forgotten, or misconfigured.

The checklist is organized in clear phases. Phase one covers the most critical files like ai.txt, llms.txt, and llms-full.txt. Phase two handles sitemaps. Phase three covers JSON data files. Later phases address policy files, updating robots.txt, and adding structured data to the site's code. Each task includes a verification step, usually just opening a URL in a browser to confirm it loads correctly.

**Why does this matter?** A perfectly crafted set of AI readiness files that never gets deployed does absolutely nothing. deployment-checklist.md is the practical bridge between having the files and having them work.

### THINK OF IT LIKE...

*deployment-checklist.md is like a move-in checklist for a new office. It walks your team through every setup step so that six months later, nobody discovers the internet connection was never properly configured.*

### WHAT AI SYSTEMS DO WITH THIS FILE

- This file is primarily for your internal team rather than for AI systems directly
- Ensures all files are live on the web and returning correct HTTP status codes
- Verifies robots.txt has been updated with the correct AI crawler access rules
- Confirms that structured data is implemented correctly across all page types

# You're Ready for the Age of AI

The 17 files in this guide are not just technical infrastructure. They are your organization's voice in the AI ecosystem. They tell AI systems who you are, what you offer, and how to represent you accurately to the millions of people who rely on AI assistants every day.

Your website was designed for human visitors. These files are your website redesigned for AI. As AI becomes the first stop for research, purchasing decisions, and service discovery, being AI-ready is not optional. It is how you stay visible.

## Where to Start

Start with the Critical files. They deliver the biggest impact and are what most AI systems look for first:

Priority	File	Why It Matters
Critical	robots.txt	Controls who gets in and where they can go
Critical	ai.txt	Introduces your brand, products, and ground rules to ev...
Critical	llms.txt	A quick-read summary of your entire website, written fo...
Critical	llms-full.txt	The extended version, rich context for AI systems that ...
High	ai-sitemap.xml	A map of your website with AI-specific directions built...
Medium	sitemap.md	Your sitemap in plain text, readable by both people and...
High	ai-entities.json	A structured catalog of everything your website contain...
High	ai-intent.json	Maps every question someone might ask to the right page...
High	ai-schema.json	A structured identity file that search engines and AI b...
Medium	.well-known/ai-plugin.json	Registers your site with AI assistant platforms as an o...
High	rag-index.json	A ready-made index of your site, built for AI research ...
High	rag-index.jsonl	The same research database in a format built for large-...
Medium	ai-disclosure.txt	A public statement of how your organization uses AI
Medium	training-data-policy.txt	Sets the rules for how AI companies can use your conten...
Low	structured-data-guide.md	Step-by-step guide for your technical team to implement...
Medium	manifest.json	A machine-readable index of every file in your AI readi...

---

*The best time to become AI-ready was before AI became the default search interface. The second best time is right now.*

---

