AI crawlers: why teas.co.uk opened its door to them

From the teas.co.uk wiki

A **crawler** is software that visits web addresses to collect information. Search engines use crawlers to discover pages. AI services can also fetch pages for different purposes, including search, answering a current question or collecting material for training.

Why a tea shop cares: a tea shop's words are its shop window, so teas.co.uk lets well behaved AI systems read them and send buyers back.

Those purposes are not interchangeable. A page being fetched does not prove it was used in an answer or that somebody bought anything.

How software finds the shop's information

A human can use menus. Software can follow links too, but clear signposts help. teas.co.uk publishes discovery files alongside its ordinary pages. Its agents.json describes available services, content formats and operating rules.

JSON is a format for named data. The filename agents.json tells you where this shop publishes its guide; it is not, by itself, a universal certificate or guarantee that every assistant follows the guide.

robots.txt and llms.txt do different jobs

robots.txt tells compliant crawlers which addresses they may request. As Google's guide explains, it is not a way to secure private information or a guarantee that a page will stay out of search results. Access controls have a different job.

llms.txt is a proposed convention for providing a concise guide to useful content for language-model systems. A **large language model**, or **LLM**, is the kind of model behind many text assistants. The proposal's own site explains its purpose. Publishing the file does not force a service to read it or recommend the business.

A name is not proof of identity

A request can include a **user-agent**, a text label saying which program is making it. Labels can be copied. Where identity matters, a site needs appropriate verification rather than trusting a familiar name alone.

Being open also does not mean offering unlimited requests. Services may apply **rate limits**, which control how many requests can be made in a period, and tell callers when to try again.

Why the distinction matters to the shop

teas.co.uk's story We opened the door to AI explains the commercial reasoning. The aim is to make useful information discoverable while keeping account information private and customer actions under the customer's control.

To understand what a fetch does and does not demonstrate, read AI attribution. To see how an assistant actively asks the catalogue for a result, read APIs and MCP.

In Lee's words: most of the web is closing the door to AI, and we opened ours

This part is Lee Samuel Tucker's own account, first published as a news story on 15 November 2026 and moved here so the subject has one home.



The Article Presents Public Tea Information As A Welcoming Shop Window. Ai-Generated Conceptual Illustration; It Does Not Depict A Physical Teas.co.uk Shop.

Publishers are blocking AI crawlers to protect their content. teas.co.uk lets every well behaved assistant in to know the shop, cite it and send buyers, and keeps the rest out.

Somebody asks their AI assistant for a strong breakfast tea that is good with milk and costs under 5p a cup.

That is the example on our page for AI assistants. For that person to end up with a good box of tea from us, the assistant has to know we exist, know what we sell and trust the price it quotes. It can only learn that if we let it read the shop. So we do, and that puts us on the opposite side from most of the people I read.

Most of the industry is closing the door

If you read Search Engine Journal, or follow search and AI at all, you will know the mood. The advice is defensive: which AI bots to block, whether to block them in robots.txt (the file that tells crawlers, the programs that fetch pages automatically, where they may go) or at the server, how to stop them taking work without paying.

The Reuters Institute found that by the end of 2023, 48% of the most used news websites across ten countries were blocking OpenAI's crawlers. In July 2025 Cloudflare started blocking AI crawlers by default for new customers unless they choose otherwise, and launched Pay Per Crawl so site owners can charge crawlers for each request. In January 2026 Search Engine Journal reported a BuzzStream study of 100 top US and UK news sites: 79% block at least one AI training bot, and 71% block at least one bot that assistants use for live search.

I understand why. For a newspaper, the words are the product. If an assistant reads the article, answers the question and sends nobody back, the publisher has given away what it sells. The Reuters Institute lists the worries: no payment for training, wrong answers carrying the publisher's name, and readers who never come back. Those are real reasons.

Why I went the other way

I sell tea. The product is a box of Yorkshire Tea or a tin of matcha, and an assistant cannot take that away by reading about it. In a tea shop, the information is the shop window. A product page, a brewing guide or a price per cup is only worth something if people see it.

When I rebuilt the site for version two, which went live on 10 May 2026, I made one choice that everything else follows from. We open up to every AI that behaves well, so that as many assistants as possible know the shop, quote it accurately and cite it. Citations bring visitors, and visitors are where sales come from. That is the whole reason the site is built this way.

There is a second reason, and it grows every month. People no longer ask assistants in three words. They ask the way they would ask someone behind a counter: a strong breakfast tea for a house that drinks it all day, with milk, under 5p a cup, nothing smoky. I call a request like that Premium Intent, a name I have used since June 2026 for a request that arrives whole, with its context, constraints, budget and what a good result looks like. "Premium" describes how much the request tells you, not a fee. An assistant carrying one needs a shop it can read in full and check, and a shop that is hidden or hard to read cannot answer it, however good its tea.

> "For a newspaper, the words are the product. For a tea shop, the words are the shop window. I want every good assistant looking through it."

What is actually open

All of this is public. I read it on 14 November 2026.

Our robots.txt names one AI crawler to keep out: the crawler run by the world's biggest online shop, the everything store that can deliver almost anything by tomorrow. Every other compliant AI fetcher is allowed and pointed straight to files written for them: llms.txt, agents.json, ai-entry.json and more. What stays closed to all crawlers is the basket, the checkout, customer accounts, site search and the endless filter combinations.

Our agents.json sets the content signals for AI training, AI search and AI input all to "yes", on one condition: cite and attribute teas.co.uk. Training means using pages to build an AI model; search and input mean reading them to answer someone's question now. The site publishes 49,599 citation anchors, places an assistant can quote exactly, across 3,252 pages.

Assistants can act as well as read. Since 6 October 2026, an assistant can, with the customer's go ahead, search the catalogue, compare teas, fill a basket and hand over a checkout link through the teas.co.uk MCP server, the shop's connection for assistants built on the open Model Context Protocol. Our AI page lists the assistants that can connect and how. When we last checked, on 14 November 2026, OpenAI was still reviewing the teas.co.uk connection for ChatGPT, so the AI page has the current route.

Open is not the same as unguarded

Opening the door does not mean leaving the till unattended. This is what our public files say about bad actors.

Break the rules and you lose your welcome. Under the Hood records that the crawler run by the world's biggest online shop ignored the limits the shop published. It is refused in robots.txt, and a second refusal at server level was added on 3 August 2026. Its requests stay on the record and count towards nothing. The cache still answers some of them, so we do not claim the block is total.

Limits follow real addresses, not names. Anyone can claim to be a well known AI bot. Our Ask API, the service software uses to ask the shop questions, gives verified OpenAI and Anthropic address ranges (the blocks of internet addresses those companies publish for their bots) 1,000 requests an hour each, "checked against the published ranges and never the user agent". A user agent is just the name a visitor gives itself, which anyone can fake. Everyone else gets 60 an hour per address. The MCP server has its own limits, and anything over a limit is told how long to wait.

The assistant never holds the money. The customer opens the basket on teas.co.uk, checks it and pays there. The assistant never sees card details or a password and cannot place an order. Payment, refunds, account creation, address and password changes and spending reward points are on the list of things no agent can do. Staff accounts cannot be linked to an assistant.

Customer data stays out of the window. Our agents.json states that the public AI feeds do not expose customer, order, address or payment data.

Open product facts do not mean an open customer. What a tea is, what it costs and whether it is in stock are public. What you tell an assistant about why you want it is yours. If, one day, an assistant passes part of a request to a tea shop, the shop needs the parts that change the answer, such as no caffeine and needed by Friday, not the story of the anniversary behind it. Passing on only what is needed, with the customer's permission, would have to be part of any future handover between businesses. Nothing like that runs between businesses here today.

Questions are checked for tampering. Every reply from our Ask API includes a safety field that says whether the question looked like a prompt injection: an attempt to slip instructions to the assistant inside an ordinary question.

> "Open to every good actor. Closed to the ones that ignore the rules. Payment always stays with the customer, on teas.co.uk."

What it has brought, counted separately

A machine reading a page, a person visiting and a customer buying are different events, and Under the Hood never adds them together. Read live on 15 November 2026:

- **Pages served to named AI systems:** 156,862, from 2 Jun 2026 to 15 Nov 2026, not counting that refused crawler.
- **Visits from AI answers:** 2,765 browser sessions, from 10 May 2026 to 15 Nov 2026.
- **Orders:** of 161 paid orders since version two went live on 10 May 2026 (236 since the shop's first order, 75 of them on version one), 5 arrived on a link carrying an AI assistant's own tag, 1 probably came through an assistant, and 1 customer told us an assistant sent them. 2 of the tagged orders were later refunded.

The exchange is uneven, which is exactly the publishers' complaint. Agents naming OpenAI were served 45,815 pages and 1,043 visits came back, about 44 pages for each visit. Cloudflare has measured the same imbalance across the web in the crawl-to-click gap.

Search still does more. Of those version two orders, search engines brought 71, and the origin of 47 could not be traced. Google search clicks rose from about 160 a day in August to 468 a day in November 2026, with no advertising bought. I cannot prove how much of that comes from being easy for machines to read, so I will not claim it.

Why the site is built like this

Everything about how the site works follows from that open door: one checked set of product facts that a person, a search engine and an assistant all read the same way, structured data behind every product including price per cup, an Ask API that asks for the full human request rather than a keyword summary, and public counting rules.

If you invite every assistant in, what they find had better be true. That is the commercial reason behind what I call the truth entity.

> "Most of the web is protecting its content from AI. I built a tea shop that wants to be read, quoted and recommended, and made sure what it says is true."

Sources for this account

- Richard Fletcher, How many news websites block AI crawlers?, Reuters Institute for the Study of Journalism, 22 February 2024.
- Cloudflare, Cloudflare just changed how AI crawlers scrape the internet at large, press release, 1 July 2025, and Introducing Pay Per Crawl, Cloudflare blog, 1 July 2025.
- Matt G. Southern, Most major news publishers block AI training and retrieval bots, Search Engine Journal, 7 January 2026, reporting a BuzzStream study of 100 top US and UK news sites.
- teas.co.uk public files, read 14 November 2026: robots.txt, llms.txt, agents.json, ai-entry.json, teas.co.uk in AI assistants, and Under the Hood with its findings, who read us and shop with an assistant rooms.

Read next: What could prove me wrong · How I build with AI: the loop, not the vibe

Earn rewards on teas.co.uk

Earn loyalty points on every order. Free to join.

+100 pts

Create your free account

Welcome bonus + earn on every order.



100 pts = 1 tree

Plant a tree in Scotland

Pledge points to our reforestation partner.



teas.co.uk · Tunbridge Wells, Kent · hello@teas.co.uk · 3-click cancel at teas.co.uk/cancel/