AI's Search-Bar Problem

Twenty-five years of search taught us to ask. Agents need to be given the job.



Firecrawl

WHITE PAPER NOVEMBER 2026

Contents

AI's Search-Bar Problem | November 2026 | 16 min read | ~4,000 words

Chapters

No.	Title	Page
00	Radio with pictures	4
01	The search-bar habit	5
02	A job, not a query	7
03	A little knowledge is a dangerous thing	8
04	Give your agent the world	10
05	What leaders should do now	13
06	The next medium	16

Also inside

No.	Title	Page
	Executive summary and key takeaways	3
	End notes	17

Executive summary

Most people still use AI the way they use a search engine. That habit caps what AI can do for them, and it will cap it further as agents take on real work.

In the largest public study of how people use ChatGPT, about half of all messages asked for information or advice. About 40% asked it to produce something.[3] As one team of UX researchers put it, “Information-seeking habits are sticky.”[4]

A search returns an answer. An agent takes a job: it plans, looks things up dozens of times, reads the sources, checks them against each other, and hands back finished work. The work arrives in bursts, many lookups in a few minutes, then quiet.

Handing over the job only works if the agent can reach what the job needs. On 581 real legal research requests, a capable agent working from memory put the official document first less than half the time. An index built from official sources did it about three times in four, and had the official document in its top ten 96.6% of the time.[5] Same request, different reach, different work.

The gap holds beyond law. Across 845 tasks run with the same model and prompts, agents connected to Alexandria, Firecrawl’s library of official data providers, scored 21% higher on answer quality than agents using built-in web tools.[13]

Enterprise research tells the same story from the other side. Nearly eight in ten companies report using generative AI, and just as many report no significant bottom-line impact.[6]

The organizations that get the most from agents over the next two years will do three things. They will hand agents whole jobs instead of questions. They will give those agents the live web plus the official sources behind it. And they will require a receipt for every fact an agent acts on. Chapter 5 sets out five moves leaders can make this quarter.

Key takeaways

01

The search bar is a habit, not a law of nature

About half of what people send to AI chat is still a question.

02

Agents work in jobs, not queries

One finished task can mean dozens of lookups in a few minutes.

03

An agent is only as good as what it can reach

Same model, same prompts, official sources: 21% better answers than built-in web tools.

04

Official sources and receipts make the output usable

You can check it, cite it and act on it.

05

Closing the gap is a management choice

Measure the mix, write jobs, map sources, require receipts, score finished work.

Radio with pictures

When television started, the first things on TV were radio shows. People sat at a desk, smoking, reading the news out loud. NBC's first network evening newscast, the Camel News Caravan, premiered in February 1949 with its anchor reading the headlines beside a lit cigarette, left smoldering in an ashtray because the sponsor asked for it.[1]

Two years later, one of the most respected programs on American radio, Edward R. Murrow's Hear It Now, moved to television and was renamed See It Now.[2] Same team. Same instincts. New box.

We were treating the new medium like the one we were used to. Nobody was wrong to. Radio was what people knew how to make, and what audiences knew how to watch. But it meant we didn't explore the visual medium right away, and it took years to learn what television could be.

The same is true for AI. We're treating it like a search bar.

We type a few words, take the top answer and close the tab. It's the habit twenty-five years of search engines taught us, and it's a fine habit for finding a web page. It's a poor one for working with something that can read, compare, check, write and finish the job.



Stop asking. Start assigning.

This paper is about the gap between those two habits: what it costs, why it exists, and what it takes to close it.

CHAPTER ONE

01

The search-bar habit

Search taught a generation to ask, take the top answer and leave. We carried that habit into AI, and it quietly caps what we get back.

The search box is one of the most successful interfaces ever built. It asks almost nothing of you: a few words and a press of Enter. In return, it hands you a ranked list of places to go. For a quarter of a century, that exchange has shaped how people look for anything, from a lasagna recipe to a court ruling.

We learned to squeeze what we need into keywords, to trust the top result, and to do the rest of the work ourselves. Open the tabs. Read the pages. Compare. Decide. Act. The search engine's job ended at the list. Ours started there.

When AI chat arrived, it came with a text box that looked a lot like the search box. So we used it like one. And agents, the tools built to do more than answer, have barely reached most people. As WIRED reported this summer, "out in the real world, most people have never touched an AI agent."[7]

What the data says

The best public evidence on how people use AI chat comes from a 2025 study by economists at OpenAI and Harvard. They classified a large sample of consumer ChatGPT messages by what the person was trying to do.[3] About 49% were Asking: looking for information or advice. About 40% were Doing: asking the model to produce something, like a draft, a plan or a piece of code. The remaining 11% were Expressing, people sharing a view or a feeling with no request attached.

Asking isn't wrong. The researchers found people value ChatGPT most as an advisor.[3] But even the doing happens inside the chat: a draft, a plan or a piece of code, handed back for the person to carry forward. One question goes in, one answer comes out, and the person does everything that happens next.

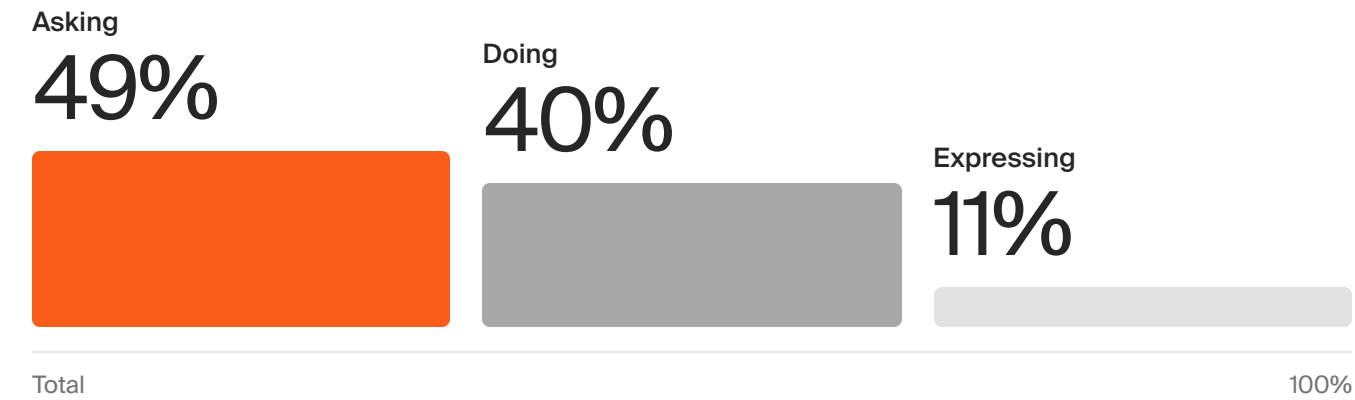
Habits are sticky

Researchers who watch people look for information see the same pull. In a 2025 study, Nielsen Norman Group found that "information-seeking habits are sticky. Once



FIGURE 1

Half of what people send to AI chat is still a question



NBER, 2025 · Share of consumer ChatGPT messages by what the person wanted.

Source: Chatterji et al., "How People Use ChatGPT," NBER Working Paper 34255, September 2025.[3]

someone finds a reliable way to easily find information they need, that method becomes nearly instinctive."[4] That's how new media go. People experiment, then drift back toward what they already know.

What gets left on the table

Look closely at almost any everyday request and the question turns out to be the smallest part of it.

"What's the best price for this TV?" really means: find the best price from a seller I trust, check it's in stock, and tell me whether it's likely to drop before Black Friday.

"Any new data analyst jobs?" really means: find the ones posted this week that match my experience, skip the ones I've already applied to, and draft a note for the best three.

The answer is step one. The value is in steps two through ten, and the search-bar habit leaves those to the person.

CHAPTER TWO

02

A job, not a query

An agent's unit of work is a job. It plans, looks things up, reads, compares, checks, cites and finishes.

EXHIBIT 1

The permit, handed over as a job

"Get my food truck ready to apply for a permit in Austin."

1 Find the city's rules for mobile food vending City code	2 Pull the official application City site	3 Check the fee schedule and required inspections City site
4 Look up the county's health requirements County health	5 Note city and county overlaps and conflicts Cross-check	6 Fill out the application for review, a link per line Finished work

Illustrative · What an agent does with one request, and where each step reads from. Every answer it fills in carries a link to its source.

Take a simple request: get my food truck ready to apply for a permit in Austin.

Typed into a search bar, it returns a list of links, and the person behind the keyboard starts reading.

Typed into a chat window used like a search bar, it returns a tidy paragraph about the kinds of permits food trucks usually need. Probably right. Possibly out of date. Definitely not a permit.

Handed to an agent as a job, it looks different. The agent finds the city's rules for mobile food vending. It

pulls the official application from the city's own site. It checks the fee schedule and the list of required inspections. Then it looks up the county's health requirements, which live somewhere else and follow different rules, and it notes where the two overlap. Finally it fills out the application for the person to review and sign, with every answer linked to the page it came from.

The person gets a filled-out application ready to sign instead of a paragraph to start from. Same request. Very different afternoon.

A job is many lookups

Six steps on paper turns into dozens of lookups in practice. A person with a search bar makes one query and reads. An agent working a job makes all of them in a burst, then goes quiet when it's done.

Tools designed around one human query at a time will keep tripping over this. Limits, pricing and result formats that make sense for a person skimming ten links make much less sense for an agent reading a hundred pages in four minutes.

Where this is heading

Gartner predicts that by 2028 at least 15% of day-to-day work decisions will be made autonomously through agentic AI. It also expects more than 40% of agentic AI projects to be canceled by the end of 2027, over rising costs, unclear business value and weak risk controls.[8] Both can happen at once.

Agents will do more of the work, and many projects will stall because nobody defined the job, gave the agent what the job needed, or checked what came back.

The rest of this paper is about the second and third of those.

CHAPTER THREE

03

A little knowledge is a dangerous thing

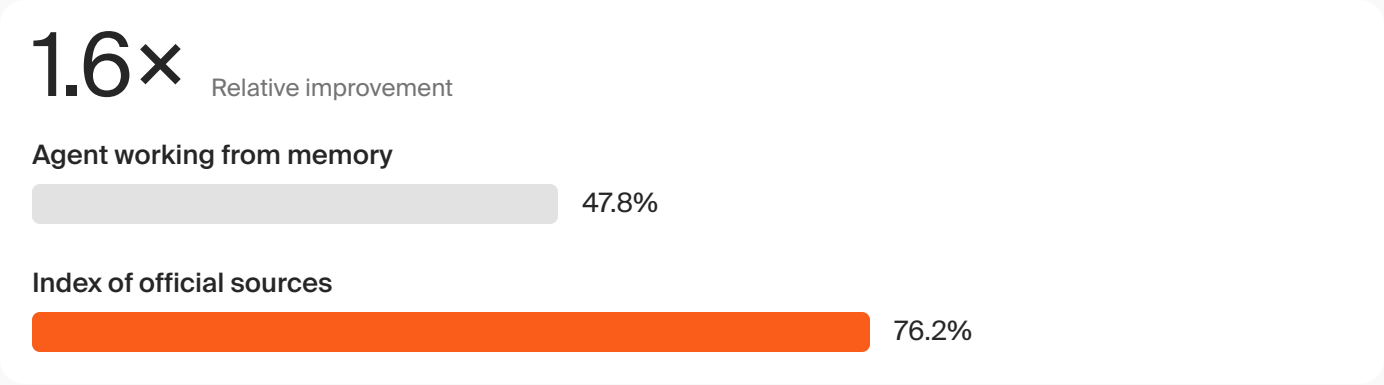
An agent's work is only as good as what it can reach. Give two equally capable agents the same job and different sources, and you get different work.

Here's a thought experiment. Give three agents the same question. They're equally capable. The only difference is what each one is connected to: one has nothing but its own training, one has a general web search, and one has an index of official documents. They'll come back with three different answers. If they're doing a job rather than answering a question, they'll come back with three different pieces of work.

That's the knowledge gap. We tend to talk about AI quality as a property of the model. For agents doing real work, it's just as much a property of what the agent can

FIGURE 2

Reach the official source and the right document comes first 1.6 times as often



Firecrawl, 2026 · Share of 581 legal research requests where the official document came first.

Source: Firecrawl Legal Information Retrieval Benchmark 2026, November 3, 2026. "Index of official sources" is the Firecrawl Government & Legal Index.[5]

read. Change what an agent can reach and you change what it hands back.

The cost of guessing

In law, the price of a wrong source has already been measured. Stanford researchers found that general-purpose chatbots hallucinated on 58% to 82% of legal queries, and that even AI tools built for legal research gave incorrect information on at least one query in six.[9] An answer that sounds right and cites the wrong authority is worse than no answer at all.

Measuring it where there's a right answer

Law is a good place to measure this, because legal questions usually have a right document. A statute says what it says. A city code has an official home. Either the agent finds it, or it doesn't.

In November 2026 we published the Legal Information Retrieval Benchmark 2026: 581 legal research requests, each rewritten from a real search people ran against Firecrawl's Government & Legal Index, and each matched in advance to the one official document that answers it.[5] No model grades the results. Scoring is an exact match on the official page.

Working from memory, a capable agent put the official document first less than half the time. An index built from official sources did it about three times in four, 1.6 times as often, and had the official page somewhere in its top ten results on 96.6% of requests. It answered in under half a second at the median.[5]

Why it matters more when agents act

A wrong source costs more in a job than in a chat. When a person reads an answer, the person is the last check. When an agent acts on it, by filling in a form, sending a summary to a client or updating a record, the wrong page travels further before anyone notices.

A stale fee schedule in a chat reply is a typo. In a permit application, it's a rejected filing and a lost week.

The hard-to-find layer

Much of what matters most sits a little out of reach. Registries live behind search forms. Filings live in databases. Research papers live behind indexes. Rules live in city code portals designed for people with patience and a free afternoon. General web search rewards what ranks, and official sources are often the pages that don't.

That's the part of the internet agents need most for serious work, and the part they're least likely to reach by default.

CHAPTER FOUR

04

Give your agent the world

Hand an agent the live web plus the official sources behind it, with a link back to every one, and its work changes.

If the knowledge gap is the problem, the fix sounds simple: give agents more to read. It's a little more specific than that. Agents doing real work need three things from the knowledge they use.

Reach

The live web is the starting point: today's prices, this morning's job postings, the injury report that dropped an hour ago. Behind it sits the material serious work depends on. That means statutes, regulations and city codes. Company filings and fundamentals. Life science papers. Public registries and directories. Open-source code and its documentation. An agent that can reach all of it, directly, doesn't have to guess which of those it's missing.

Receipts

Every fact an agent acts on should come with a receipt: a link to the source, a citation, the jurisdiction it applies to where that matters, and structured fields the agent can work with directly instead of re-reading a page. Receipts turn an answer you'd double-check into work you can review in seconds.

Built for how agents work

The search industry is already moving this way. As one investor analysis put it last year, "Web search for the past 30+ years was built for humans. Now it's being rearchitected for agents."[10] Reach and receipts only help if the agent can use them at agent speed. That means answers in milliseconds rather than seconds, limits that expect bursts, and clean formats an agent

EXHIBIT 2

Anatomy of a receipt

One fact the agent will act on

Source link	The official page it came from
Citation	The code section, filing or paper
Jurisdiction	Where the rule applies
Fields	Structured values the agent can use directly

Sources a job draws on

Law
Statutes, regulations, city codes, court opinions

Finance
Company filings and fundamentals

Research
Life science papers

Records
Public registries and directories

Code
Open-source code and documentation

Web
Everything else on the live web

What travels with every fact an agent acts on, and the kinds of sources a job draws on.

can read without wading through page furniture. When knowledge is slow or awkward to reach, agents do what people do. They fall back to whatever's easiest, and the job quietly gets worse.

What it looks like at scale

At Firecrawl, we bring these sources together in Alexandria, a library of official data providers and our own indexes that agents reach alongside the live web. Two examples show the shape of it. Firecrawl's Government & Legal Index covers more than 30 million data points of US law, from statutes to city codes and court opinions, and agents can search it for free.[11] Our life science research collection gives agents more than 41 million life science papers, from drug discovery and clinical trials to gene studies, free for agents in popular AI assistants.[12] Those are two collections among many. The principle is the same for filings, registries and code: reach the source, keep the receipt.

Reach shows up in the work. Across 845 tasks run with the same model and prompts, agents using Firecrawl's library of official data providers scored 21% higher on answer quality than agents using built-in web tools.[13]

For humans and their agents

None of this should require people to learn a new tool. Most people will never configure a data source, and they shouldn't have to. Their agent picks it up, the way a phone picks up a new network. The person asks for the job; the agent knows where to look.

That's the way Firecrawl is built to be used. More than 1.5 million users have signed up, and more than 150,000 companies build with Firecrawl, including teams at 67% of the Fortune 100.[14]

What changes

When an agent can reach the world's knowledge with a receipt on every fact, three things change for the people relying on it. They can **check** the work in seconds by following the links. They can **cite** it in front of a client, a regulator or a boss. And they can **act** on it, because they know where each fact came from and which rules it falls under.

That's the difference between an answer you'd want to double-check and work you'd put your name on.

CASE STUDY



Retell

Phone agents that answer from today's docs

Retell runs AI voice agents that handle business phone calls on their own: a borrower checking an application, a lead on an outbound sales call, a customer asking how a feature works.[15]

15M

pages a month

PROBLEM	On a live call, the agent has seconds to answer, and the caller hears whatever its knowledge base says. If a pricing page or policy changed last week and the agent missed it, the answer is wrong.
WHAT THE AGENT DOES	Each customer's knowledge base is just a list of URLs. Firecrawl reads the docs, help center, API reference and PDFs behind them and turns them into content the agent can use. Keeping it current means rerunning the job.
RESULT	About 500,000 pages a day, with no ingestion pipeline for Retell to build or staff. New customers are set up from a list of links, not an engineering project.

> **"Firecrawl made our knowledge base creation workflow easy, handling different kinds of websites reliably."**
>
> **Zexia Zhang**, Co-founder & CTO, Retell



TRY IT IN FIVE MINUTES

Hand your agent one job today

1 Connect

In Claude, ChatGPT or Cursor, add the Firecrawl connector and sign in. It's free to start, with no keys to copy. From a terminal:

```
$ npm install -g firecrawl-cli
$ firecrawl login
```

2 Hand it the job

```
Use Firecrawl to get my food truck ready to apply for a
permit in Austin. Use official city and county sources.
Fill out the application for me to review, with a link to
the source for every answer.
```

3 Check the receipts

Open three links at random. Each one should land on the official page the answer came from. Then swap in a job from your own week.

CHAPTER FIVE

05

What leaders should do now

The gap between asking and assigning won't close with the next model release. It's a choice about how your team works, and you can make it this quarter.

Television didn't find its feet because the cameras got better. It found them when producers stopped filming radio scripts and started asking what pictures could do.

Enterprise research suggests AI is at the same stage. McKinsey found that nearly eight in ten companies report using generative AI, yet just as many report no significant bottom-line impact, and that about 90% of function-specific use cases remain stuck in pilot mode. Its diagnosis is blunt: "Today, AI is bolted on."[6]

The same shift television made is available to any team using AI today, and it starts with how work is designed, not with the next model. Here are five moves.

1 Measure the mix

Sample a week of your team's AI use and sort each request into two piles: asking (a question that ends in an answer) and assigning (a job that ends in finished work). If more than half your use is asking, you have room to grow.

→ **Try** Pull 50 requests from one team. Count. Share the split at your next staff meeting.

2 Write jobs, not prompts

Pick the five tasks your team repeats most and write each as a job: the goal, the sources, the checks and what "done" looks like. Store them as setup guides or skills agents can reuse, so nobody retypes the method.

→ **Try** Turn one weekly report into a written job this week, and let an agent run it next week.

3 Map what your agents can reach

List the sources each job depends on: live web pages, laws and codes, filings, research, registries, code. Each one your agents can't reach directly is a gap where they read a different internet from the one the job needs.

→ **Try** For your highest-stakes job, name the three sources a careful expert would check first. Can your agent reach all three? If not, connect a source library such as Alexandria and ask again.

4 Require receipts

Make it a rule: anything an agent acts on links back to where it came from. No link, no action. Receipts make review fast and audits possible, let trust build on evidence instead of hope, and make work easier to hand off.

→ **Try** Add “cite a source for every fact” to every written job from Move 2.

5 Score finished work

Stop counting prompts per seat or answers per day. Those reward the search-bar habit. Track time to done and rework. If time to done falls and rework stays flat, the agent is working. If rework climbs, go back to Move 3.

→ **Try** Pick one job and record time to done for four weeks.

A first 30 days

WEEK 1

Measure

Sample and sort a week of requests. Share the asking-to-assigning split.

WEEK 2

Write

Turn three repeated tasks into written jobs with sources, checks and a definition of done.

WEEK 3

Connect

Connect your agents to the sources missing from your map. Add the receipts rule to every job.

WEEK 4

Score

Compare time to done and rework against week one. Keep what worked; write the next three jobs.

CLOSING

06

The next medium

It took television a few years to stop being radio with pictures. Once it did, nobody wanted to go back to a man reading the news beside an ashtray.

AI is at the same point. The tools can already do far more than answer questions. What's holding most teams back is habit, and the reach of the agents they've deployed. Both can change quickly.

The web was built for people with a search bar. What comes next is being built for agents: the live web and the hard-to-find official sources behind it, brought together and organized for readers that check everything and cite as they go. Call it knowledge infrastructure: the layer that makes agents smarter. At Firecrawl, we're building ours, and we call it Alexandria.

Whatever you call it, the test is the same. Pick one task you'd normally search for. Hand it to an agent that can reach the world's knowledge. Check the receipts.

Give your agent the world

Start with one job this week. Add the Firecrawl connector in Claude, ChatGPT or Cursor, or run `npm install -g firecrawl-cli`, and hand your agent the permit job from Chapter 4. Free to try, no credit card needed.

Connect your agent at alexandria.firecrawl.dev →

End notes

1 "Swayze, John Cameron," Encyclopedia.com. encyclopedia.com/humanities/encyclopedias-almanacs-transcripts-and-maps/swayze-john-cameron. See also "Camel News Caravan," Wikipedia, for the February 16, 1949 premiere. en.wikipedia.org/wiki/Camel_News_Caravan

2 "Edward R. Murrow," Wikipedia: Hear It Now moved to television on November 18, 1951 as See It Now. en.wikipedia.org/wiki/Edward_R._Murrow; HistoryLink.org, "Murrow, Edward R. (1908-1965)." historylink.org/File/23016

3 Chatterji, A. et al., "How People Use ChatGPT," NBER Working Paper 34255, September 2025. nber.org/system/files/working_papers/w34255/w34255.pdf. Summary: OpenAI, "How people are using ChatGPT," September 15, 2025. openai.com/index/how-people-are-using-chatgpt

4 Moran, K., Rosala, M. and Brown, J., "How AI Is Changing Search Behaviors," Nielsen Norman Group, August 15, 2025. nngroup.com/articles/ai-changing-search-behaviors

5 Firecrawl, Legal Information Retrieval Benchmark 2026, November 3, 2026. firecrawl.dev/benchmarks/legal-retrieval. Code and data: github.com/firecrawl/benchmark-legal-retrieval

6 McKinsey & Company (QuantumBlack), "Seizing the agentic AI advantage," June 13, 2025. mckinsey.com/capabilities/quantumblack/our-insights/seizing-the-agentic-ai-advantage

7 Zeff, M., "Why Normal People Aren't Using AI Agents," WIRED, August 6, 2026. wired.com/story/why-normal-people-arent-using-ai-agents

8 Gartner, "Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027," press release, June 25, 2025. gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027

9 Stanford HAI, "AI on Trial: Legal Models Hallucinate in 1 out of 6 (or More) Benchmarking Queries," 2024. hai.stanford.edu/news/ai-trial-legal-models-hallucinate-1-out-6-or-more-benchmarking-queries

10 Cui, J., Li, J., Wang, S. and Zhang, S., "Search Wars: Episode 2," Andreessen Horowitz, November 13, 2025. a16z.com/search-wars-episode-2

11 Firecrawl, Government & Legal Index. alexandria.firecrawl.dev/firecrawl-gov-index

12 Firecrawl, life sciences research collection. alexandria.firecrawl.dev/categories/life-sciences

13 Firecrawl, “Introducing Alexandria and our $75M Series B,” September 22, 2026. firecrawl.dev/blog/introducing-alexandria-series-b

14 Firecrawl company figures, November 2026.

15 Firecrawl, “How Retell Keeps AI Phone Agents Accurate Across 15 Million Web Pages a Month,” November 9, 2026. firecrawl.dev/blog/retell-firecrawl-ai-phone-agents



Firecrawl

Firecrawl is knowledge infrastructure to make your agents smarter. Whatever you're building, one install gives your agent the live web plus the official sources behind it, so it can do the work, not just answer, and every fact it acts on traces back to where it came from.

More than 1.5 million users have signed up, and more than 150,000 companies build with Firecrawl, including teams at 67% of the Fortune 100.

Learn more at alexandria.firecrawl.dev