Chapter 4

Network: Scraping Sites and Crawling Agents

Michael Veale and Robert Gorwa

Chapter 4 of *Malice, Models and Middlemen: The Full Stack Fight Against AI Misuse*, pages 93–115

Abstract

Modern AI systems, including agentic systems, interact with a wide variety of software tools that allow interfacing with external services and data sources, often via Web scraping or API access. Such tools are also part of misuse pipelines, for example by powering convincing scams or fraud. Powerful intermediaries exist at the network layer between the AI system and these third-party services, most notably among them infrastructural firms such as Cloudflare. In order to respond to the rapid expansion of scraping for AI training and use, these intermediaries have been positioning themselves as new gatekeepers, with bot detection tools such as CAPTCHAs and new cryptographic bot authentication tools. They have been using their market leverage to reshape standards. While these efforts have to date been focused on protecting creative content, news and server load, the same tools may have a powerful role in throttling misuse. At the same time, such infrastructural reshaping comes at a high cost — endangering the openness of the Internet itself.

Preprint, November 2026. Forthcoming from MIT Press, 2027. Page numbers follow the preprint of the full book.

Cite as: Michael Veale and Robert Gorwa, *Malice, Models and Middlemen: The Full Stack Fight Against AI Misuse* (MIT Press forthcoming 2027) ch 4

https://homepages.ucl.ac.uk/~ucqnmve/malicemodelsmiddlemen/chapter/4/

Copyright © 2026 Michael Veale and Robert Gorwa. All rights reserved. This page will be disabled for a period following publication of the physical book.

Reservation. The rightholder expressly reserves all rights to use this work for text and data mining, including under Article 4(3) of Directive (EU) 2019/790 and national implementations. This covers training, fine-tuning or evaluating AI models, and indexing or retrieval for AI systems.

Licensing. No licence is granted for these uses except as stated here.

Retrieval licence. Live retrieval to answer an individual user query is licensed only if the output shows visible credit and a working link to the source, quotes no more than 30 words, and doesn't replace reading the work. The content must not be used for training.

Other rights. Nothing here concedes that any unlicensed use is fair use or fair dealing.

4 NETWORK: SCRAPING SITES AND CRAWLING AGENTS

Modern AI systems, including agentic systems, interact with a wide variety of software tools that allow interfacing with external services and data sources, often via Web scraping or API access. Such tools are also part of misuse pipelines, for example by powering convincing scams or fraud. Powerful intermediaries exist at the network layer between the AI system and these third-party services, most notably among them infrastructural firms such as Cloudflare. In order to respond to the rapid expansion of scraping for AI training and use, these intermediaries have been positioning themselves as new gatekeepers, with bot detection tools such as CAPTCHAs and new cryptographic bot authentication tools. They have been using their market leverage to reshape standards. While these efforts have to date been focused on protecting creative content, news and server load, the same tools may have a powerful role in throttling misuse. At the same time, such infrastructural reshaping comes at a high cost — endangering the openness of the Internet itself.

BOTS BEHAVING BADLY

In June 2024, Perplexity, a provider of AI-augmented search, was queried with a request to summarize a website. To the surprise of the requester, the Web developer, blogger and all-round tinkerer Robb Knight, Perplexity returned a summary of his personal webpage that included recent updates made *after* he had altered the page to both instruct crawlers to go away in general, and to block Perplexity's publicly listed IP addresses in particular. When Knight scanned the logs to see what 'user agent' was listed — the calling card that should be attached to each Web query with information about the browser and operating system of the requester — he found that Perplexity was not properly disclosing its crawlers as it publicly pledged to do. Instead, the firm appeared to be operating scrapers posing as an anonymous user browsing from the Washington, DC suburbs via Chrome on Windows 10.[1]

Meanwhile, developers at the popular online repair community iFixit were surprised to see thousands of queries a minute hitting their websites from many different sources.[2] Labor marketplace Freelancer.com similarly reported receiving 3.5 million visits in just a single afternoon. These scrapers, eventually linked in both cases to Anthropic, became even more

1: Robb Knight, "Perplexity AI Is Lying about Their User Agent" (15 June 2024) https://rknight.me/blog/perplexity-ai-is-lying-about-its-user-agent/ accessed 8 October 2024; Dhruv Mehrotra, "Perplexity Is a Bullshit Machine" (*WIRED*, 19 June 2024) https://www.wired.com/story/perplexity-is-a-bullshit-machine/ accessed 8 October 2024, (noting that Perplexity soon ceased publishing such addresses); Gabriel Corral and others, "Perplexity Is Using Stealth, Undeclared Crawlers to Evade Website No-Crawl Directives" (*Cloudflare Blog*, 4 August 2025) https://blog.cloudflare.com/perplexity-is-using-stealth-undeclared-crawlers-to-evade-website-no-crawl-directives/ accessed 21 August 2026, and replicating this behavior with a more formal research group, see.

2: Jason Koebler, "Anthropic AI Scraper Hits iFixit's Website a Million Times in a Day" (*404 Media*, 24 July 2024) https://www.404media.co/anthropic-ai-scraper-hits-ifixits-website-a-million-times-in-a-day/ accessed 8 October 2024.

aggressive after the websites deployed standard protective mechanisms. The crawlers began to endlessly cycle through unique IP addresses to make their activity almost impossible to reliably stop.[3]

Other community and hobbyist projects around the world were also being impacted. Earlier in 2024, the open-source, community-run software documentation website Read the Docs reported the wasteful and repeated extraction of billions of PDFs — eventually coming to 73 terabytes of material — by an unknown actor, incurring over $5,000 in bandwidth fees.[4] A few weeks later, Read the Docs identified a similar 10 terabyte surge as stemming from Meta-controlled IP addresses.[5] Game designer Edd Coates, the creator of the Game UI Database, a non-commercial labor-of-love video game documentation project, was similarly faced with an onslaught of requests from crawlers operated by OpenAI. These bots were refreshing pages hundreds of times a second, pulling 60 to 70 gigabytes of files in just a few minutes, costing Coates around $850 per day.[6] What exactly was going on?

RETRIEVAL AUGMENTED GENERATION

Readers will undoubtedly know that LLMs can return wildly incorrect answers when queried. Widely publicized examples are manifold, ranging from AI summaries instructing people to eat rocks for their health benefits to LLMs recommending the use of glue to attach cheese to pizza so that it doesn't slide off the crust when you bite into it.[7] Such behavior, popularly called hallucination, is unlikely to simply disappear from models in the future as they improve. A range of theoretical work in computer science compellingly argues that, for many reasons, hallucination is an intrinsic, structural feature of LLMs.[8]

Hallucination is structural in part because training data is necessarily incomplete, noisy and outdated. It contains internal contradictions with no easy method to reconcile them. If you were made to drink the entire Kool-Aid of the modern Internet, you might hallucinate too. Another reason is that language models are trained both to be confidently fluent and, ideally, to deal with uncertainty in a measured and calibrated way. Just as in humans, especially those with a proclivity for bullshit, these two tasks can also find themselves in tension.[9]

3: George Hammond, "AI Start-up Anthropic Accused of 'Egregious' Data Scraping" (*Financial Times*, 26 July 2024) https://www.ft.com/content/07611b74-3d69-4579-9089-f2fc2af61baa accessed 15 November 2024; Anthropic, "Does Anthropic Crawl Data from the Web, and How Can Site Owners Block the Crawler?" (*Anthropic Help Center*, October 2024) https://support.anthropic.com/en/articles/8896518-does-anthropic-crawl-data-from-the-web-and-how-can-site-owners-block-the-crawler accessed 8 October 2024.

4: Eric Holscher, "AI Crawlers Need to Be More Respectful" (*Read the Docs*, 25 July 2024) https://about.readthedocs.com/blog/2024/07/ai-crawlers-abuse/ accessed 8 October 2024.

5: ibid.

6: Darius Rafieyan, "Like Digital Locusts, OpenAI and Anthropic AI Bots Cause Havoc and Raise Costs for Websites" (*Business Insider*, 19 September 2024) https://www.businessinsider.com/openai-anthropic-ai-bots-havoc-raise-cloud-costs-websites-2024-9 accessed 15 November 2024.

7: Megan Cerullo, "Google Admits Its AI Overviews Can Generate "Some Odd, Inaccurate" Results" (*CBS News*, 31 May 2024) https://www.cbsnews.com/news/google-ai-overview/ accessed 19 February 2025.

8: Muhammad Ahmed Mohsin and others, "On the Fundamental Limits of LLMs at Scale" (*arXiv*, 17 November 2025) DOI: 10.48550/arXiv.2511.12869; Sourav Banerjee, Ayushi Agarwal, and Saloni Singla, "LLMs Will Always Hallucinate, and We Need to Live with This" in Kohei Arai (ed), *Intelligent Systems and Applications* (Springer 2025) DOI: 10.1007/978-3-031-99965-9_39.

9: A common refrain in critical scholarship has been to object to the term 'hallucination,' on the grounds that the metaphor anthropomorphizes such systems, and also indirectly presumes consciousness. We have yet to be made aware of a better term —

Beyond hallucination, as AI tools start to be used more frequently for information retrieval in manners that approximate traditional search, they pose a simple accuracy problem: how might models being used today to answer user queries be able to answer questions relating to the news of the day, or to things in the world that came to exist after a model was trained? After all, training generative models from scratch is hugely resource-intensive and expensive. So is re-training them constantly as the world changes just to answer a simple question of recent fact.

The answer to both the accuracy and hallucination problems, for most AI developers, has been to look outside the model rather than just try to pile more information deeper within it.[10] Today, top-of-the-line generative systems are typically situated within extensive software pipelines, where they can interact with personal or proprietary databases, Web-facing search engines or other sources that can be cross-referenced. This practice is commonly called retrieval-augmented generation, often referred to in the plural as 'RAGs.'[11]

Under the RAG method, models are trained and instructed to convert a user's prompt, if the task requires it, into a series of machine-readable commands. These are sent to another tool designed to produce compact data structured to work well with a model's quirks. This data is then — in a process typically hidden from the user — fed into a model's context window, sitting alongside the user's original prompt. The LLM then responds to the user's query based on its trained weights, but the output is somewhat influenced by this additional context. Models can be trained to treat this external content as more authoritative and relevant, although whether that is a good idea will depend on the nature of the content provided and the query.

Popular chatbot-style cloud AI models like ChatGPT or Claude commonly integrate with Web-facing RAGs to search and browse.[12] And the widespread adoption of this technique is the source of the unabashed Web scraping facing sites like iFixit — a source of extremely valuable and up-to-date information about device DIY repair. A search engine scrapes to produce an index, not in real time for every search.[13] But RAGs use the full content, not just a link and a snippet. In search, human queries are tethered to content requests limited by the number of sites a user clicks on. For RAGs, a human query is tied to a much larger scale of usage that is beyond what a single user would ever normally request within that timeframe, multiplying the network load services have grown used to.[14]

even popularly suggested alternatives, such as 'bullshitting,' channeling philosophers like Harry Frankfurt, seem to suffer from the same problems. 'Hallucination' remains in common use and we use it here for clarity.

10: Mohsin and others (n 8).

11: Patrick Lewis and others, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks" in *Advances in Neural Information Processing Systems* (Curran Associates, Inc 2020) vol 33 DOI: 10.48550/arXiv.2005.11401; Reiichiro Nakano and others, "WebGPT: Browser-assisted Question-Answering with Human Feedback" (*arXiv*, 17 December 2021) https://arxiv.org/abs/2112.09332v3 accessed 8 October 2024.

12: RAGs can also be designed to query the user's own files, whether it be private files stored on cloud services like Google Drive, which integrates with *Gemini*, or the files they have stored on their computer. Advanced users might even have them query the Internet or other third-party services themselves without going through any Internet intermediaries, such as through tools like Langchain. See LangChain, "Tools" (*LangChain Documentation*) https://python.langchain.com/docs/integrations/tools/ accessed 16 November 2024; Harrison Chase, "LangChain" (*GitHub*, October 2022) https://github.com/langchain-ai/langchain accessed 8 October 2024. We focus on networked RAGs here.

13: See generally Joris Van Hoboken, "Search Engine Freedom: On the Implications of the Right to Freedom of Expression for the Legal Governance of Web Search Engines" (PhD thesis, University of Amsterdam 2012).

14: See for example the current (2026) *Deep Research* functionality of chatbots such as Google Gemini and Anthropic's Claude,

AGENTS AND INTERVENTION

While RAGs focus on retrieving information, acting in the world is also a crucial part of modern generative AI. This is not conceptually very different from RAGs, but a focus on what AI systems do in relation to external services raises a new set of issues. Even RAGs for Web searches do send data out rather than just retrieving it — a user's chatbot making a search for biographical information on a certain potential worker still sends the query to a third party, and that query reveals something about the user, what they were interested in and what they were doing. Agentic tools are a step up from that, and are deliberately designed to interact with the broader world.

The daunting term 'agent' is not, at bottom, very complex.[15] The software setup behind these systems is laid out in Figure 4.1. When a user provides input, this input is sent to a generative AI model that engages in a looped process designed to reason about the input, decide on a course of action — which might include triggering an array of different tools — and undertake and evaluate the results of that action. If the reasoning process indicates more action is needed, the loop continues, until the model returns a result that indicates that the loop should stop and a final output should be provided to the user. Software called a 'harness' sits over the model and all of its tools and interactions, coordinating them. Tools are just pieces of software that take certain input and produce certain output, and describe their functioning in short natural language to the model so it can determine what they do, whether they should be used and what results to expect from them. Such tools can, as Figure 4.1 indicates, even involve summoning more agents with similar toolkits to undertake sub-tasks.

This broad toolkit gained significant media attention in early 2026 as OpenClaw, an open-source software pipeline mimicking Anthropic's expensive Claude Code tool, went viral on GitHub.[16] Once set up on a server and given root access to one's device, the software could integrate with many more pieces of software and external services — managing your calendar, handling your email inbox or making online purchases with your cards — acting on prompts a user could message to it while on the go, or on instructions they provided in advance.

The key thing for the purpose of this chapter is the connection between these tools and various external networked services. These connections are the moments where AI systems call for data sources or sites of potential action which are distant from them, typically over the Internet. Such tool use is crucial to agents because without it, they cannot do much autonomously unless the user manually does something with the output they receive.

Software that connects to networked services might take actions that compromise those

which report to the user that hundreds of sites were pulled for a single query. In general, the unusually high server load we describe here appears to be driven by a mix of factors. These include the growing number of scrapers; a decreasing respect for 'politeness' in scraping; a desire for recency (and thus frequent requests) to limit hallucination; the difference between search indexing (one company scrapes, many use the results) and AI tool use (one scrape per customer, on demand); growing numbers of non-human-run systems, often cheaply built and managed by AI; and an increase in services designed to foil antibot systems, such as residential proxy networks. Providers might also query cached datasets of websites, or datasets they have obtained from third parties for this purpose, but this comes with some degree of legal liability on whether they are permitted under intellectual property law to retain and use such content.

15: Agents are actually quite old — what is new is the use of LLMs in the controllers. However, the ease of use of contemporary agents does, in our view, represent a step change for misuse potential.

16: Matthew Gault, "Silicon Valley's Favorite New AI Agent Has Serious Security Flaws" (*404 Media*, 30 January 2026) https://www.404media.co/silicon-valleys-favorite-new-ai-agent-has-serious-security-flaws/ accessed 7 March 2026. The tool was also variously called Moltbot and Clawdbot: it kept being renamed by the creator following copyright pressure from Anthropic.

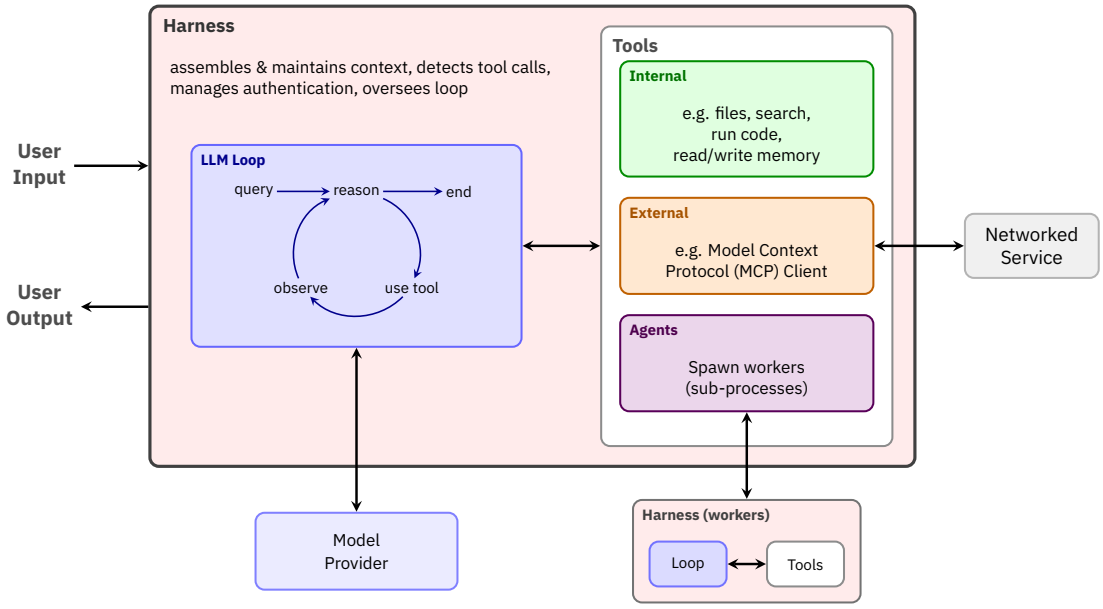

Figure 4.1: The multiple components in an AI 'agent'. Diagram: authors.

services. It might affect the availability of those services — by overloading them, for example — leak the information connected to them or trigger unwanted actions in the real world. These might be purely accidental — OpenClaw users were regularly reported to have lost data or experienced general agentic mishaps from giving their tools unfettered access to their information and their login details.[17] But other actors might make use of these information flows, for example to inject false information or malicious instructions to compromise those systems and those they are attached to.[18] These actors might be engaging in state-sponsored hacking or 'equipment interference' authorized by warrants secretly handed to the AI firms, the tool providers or network actors sitting between them.[19]

The seams between the AI models and the tools they use form another intervention point for varying types of governance. But what kind of misuse might be governed using them?

FROM RAGS TO RICHES

RAGs might conjure images of users searching the Web for product comparisons. Agents might conjure images of software completing a transaction on behalf of a user. But, as the

17: Gault (n 16).

18: Andrea M Matwyshyn, "Exploit Machina" (2026) 59 UC Davis Law Review; Amit Magen Medina and Idan Hen, "Plug, Play, and Prey: The Security Risks of the Model Context Protocol" (*Microsoft Community Hub*, 7 May 2025) https://techcommunity.microsoft.com/blog/microsoftdefendercloudblog/plug-play-and-prey-the-security-risks-of-the-model-context-protocol/4410829 accessed 2 January 2026.

19: See generally Caspar Bowden, *The US Surveillance Programmes and Their Impact on EU Citizens' Fundamental Rights* (European Parliament 2013) (on the compromise of the cloud and the Snowden revelations); Bernard Keenan, *Interception: State Surveillance from Postal Systems to Global Networks* (MIT Press 2025) DOI: 10.7551/mitpress/15654.001.0001; Investigatory Powers Act 2016, part 5.

reader might expect from this tome by now, some uses of these technologies can be significantly darker.

North Korean state agents have for many years now sought fraudulent work as IT contractors, pretending to be based domestically and typically taking multiple salaries at once to funnel to the regime.[20] They have been rapid adopters of generative models, using what appears to be a mix of local and cloud services for voice cloning, image generation and creating high-quality job applications.[21] RAGs and similar techniques are being used to allow handlers to operate large numbers of identities and deal with the specifics of many employers — at least for long enough to obtain salaries for the regime.[22]

Relatedly, scam operators, described extensively in Chapter 3, have automated personas to help operate their scams.[23] Convincing romance scams require scammers to generate personas that seem to naturally interact with an array of services: liking Instagram posts, messaging back on WhatsApp, posting to TikTok. Juggling multiple scam targets and personas requires both specific context and the ability to be in more than one place at once. RAGs and AI agents support both ends. When interviewed by researchers, scammers have explained how they deploy databases of chat history to create context-aware LLM responses from everything that is known about the victims.[24] Major platforms have repeatedly reported criminals building, sharing and using RAG tools in order to engage in hacking and social engineering, including to behaviorally profile victims, rank their interests and pull up stolen credentials and files, based on accumulated databases of stolen data.[25] In a 2026 survey of 832 Anthropic accounts banned for cybersecurity breaches, the firm found that 14.9% of these accounts were remotely gathering victim identity information, while 4% were used to create social media accounts for malicious purposes.[26]

Online data scraped and gathered using these tools has heightened value due to the potential for automatic AI-enabled analysis. AI geolocation has been particularly highlighted as an

20: Israel National Cyber Directorate, "Unmasking North Korea's Digital Deception: The Covert Operations of DPRK IT Workers in the Age of AI and Deepfakes" (*INDC*, October 2024) https://www.gov.il/BlobFolder/reports/alert_1813/en/ALERT-CERT-IL-W-1814.pdf accessed 3 January 2026.

21: Microsoft, "Jasper Sleet: North Korean Remote IT Workers' Evolving Tactics to Infiltrate Organizations" (*Microsoft Security Blog*, 30 June 2025) https://www.microsoft.com/en-us/security/blog/2025/06/30/jasper-sleet-north-korean-remote-it-workers-evolving-tactics-to-infiltrate-organizations/ accessed 3 January 2026; Evan Gordenker, "False Face: Unit 42 Demonstrates the Alarming Ease of Synthetic Identity Creation" (*Unit 42*, 21 April 2025) https://unit42.paloaltonetworks.com/north-korean-synthetic-identity-creation/ accessed 3 January 2026, (considering their likely use of local image models for criminal ends).

22: Liam Dermody and others, "How AI Services Power the DPRK's IT Contracting Scams" (*Okta Security*, 24 April 2025) https://sec.okta.com/articles/2025/04/GenAIDPRK accessed 3 January 2026.

23: Simon Moseley, "Automating Deception: AI's Evolving Role in Romance Fraud" (*Centre for Emerging Technology and Security (CeTaS)*, 15 April 2025) https://cetas.turing.ac.uk/publications/automating-deception-ais-evolving-role-romance-fraud accessed 15 April 2025.

24: Gilad Gressel and others, "Love, Lies, and Language Models: Investigating AI's Role in Romance-Baiting Scams" in (USENIX Security '26, 2026) DOI: 10.48550/arXiv.2512.16280.

25: Anthropic, "Threat Intelligence Report: August 2025" (August 2025) https://www-cdn.anthropic.com/b2a76c6f6992465c09a6f2fce282f6c0cea8c200.pdf accessed 7 August 2026, 23–25; Google, "Adversaries Leverage AI for Vulnerability Exploitation, Augmented Operations, and Initial Access" (*Google Cloud Blog*, 11 May 2026) https://cloud.google.com/blog/topics/threat-intelligence/ai-vulnerability-exploitation-initial-access accessed 7 August 2026, (discussing the Graphiti memory system); Microsoft, "AI as Tradecraft: How Threat Actors Operationalize AI" (*Microsoft Security Blog*, 6 March 2026) https://www.microsoft.com/en-us/security/blog/2026/03/06/ai-as-tradecraft-how-threat-actors-operationalize-ai/ accessed 6 August 2026.

26: Kyla Guru, Alex Moix, and Jacob Klein, "Mapping AI-enabled Cyber Threats" (*Anthropic*, 3 June 2026) https://www.anthropic.com/research/attack-navigator accessed 20 August 2026.

emerging tool in scammers' toolkits. It allows posts of even generic-looking holiday photos to be immediately and often fairly accurately placed in a rough location, helping scammers to pretend to be in a trusted or knowledgeable position in relation to a victim, with information that the victim might expect only an organization such as a bank or embassy to have.[27]

BOT OR NOT

Because of the potential misuse of services by bots, or simply the load they place on online services, there are many efforts to try to distinguish between bots and humans on the Internet. To understand these efforts, we have to understand a little more about the different ways there are to interact with services over networks.

PROGRAMMATIC ACCESS

Programmatic access describes any interface aimed at easy automated use. Many services provide an 'application programming interface,' or API for short, for the services they provide.[28] APIs can be created for almost anything you can imagine: they are the way that software talks to other software. For example, while an airline has a website aimed at consumers, it also typically provides APIs allowing travel agents and price comparison websites to rapidly search and book through specialized software they build or procure. It would be an inefficient use of the network to send such software pretty images and videos to entice it, and futile to try to upsell it using psychological dark patterns designed to instill a sense of purchasing urgency. Computers can also process much more simultaneous information than a human, and reply much faster. Thus, APIs are often websites with their visual interface removed, but with more customized options and powerful capabilities.

Many agentic systems programmatically interface with third-party services using tools running the Model Context Protocol (MCP), a widely adopted standard initially proposed by Anthropic. MCP servers inform models in natural language of the various functionalities they offer, accept queries, manage authentication like passwords and keys, and return results tailored for an LLM's context window. Unlike classic APIs, MCP results are optimized for multistage dialogue. For example, if an LLM requests search results, the server is likely to provide hints for the next commands that the model might fruitfully use to find out more based on the features of those specific results.

Programmatic access is far from new, but what is new is the ease with which relative laypersons can create, orchestrate and alter workflows using natural language. LLMs permit more flexible automation, at scale, with little investment, few resources and only limited expertise.

APIs have constraints that often make them less appealing for serious misuse. Because programmatic access typically requires some form of authentication, requests can usually be linked

27: Brooke Seipel, "AI Can Find Your Location 91% of the Time Using Just One Photo" (*McAfee Blog*, 1 July 2026) https://www.mcafee.com/blogs/mcafee-news/ai-travel-photo-location-scams-geolocation-research/ accessed 19 August 2026; Radhika Jhalani and others, "Nowhere to Hide: Privacy Risks and Policy Implications of AI Geolocation" (*Privacy International*, February 2026) https://perma.cc/7HJN-72PN accessed 20 August 2026.

28: Typically using the same technology behind loading any website, an HTTP request, although usually a POST request rather than a GET request.

to an individual account or user. Unauthenticated or open APIs exist, but applications capable of being easily abused do not tend to offer them. Authenticated APIs can typically be monitored using the same or similar tools as those described to manage pooled compute in Chapter 2, as it is not easily possible to stop all types of abuse just by limiting users' permissions to specific actions or access.[29] Detection of malicious activity is therefore important, allowing misused keys to be disabled. Because of this risk of being disabled, and of having activity tracked and monitored, malicious users often want to access services by means other than their locked-down APIs.

BROWSING LIKE A HUMAN

Where APIs are unavailable, insufficient or just too detectable, those who wish to automate tasks online must turn to different methods. They can exploit the fact that Web-enabled services need to be accessible to human users for their own commercial viability. Bots can therefore try to use interfaces designed for humans. This often means they need to appear to be browsing like a human.

This is a less clean and straightforward process than using an API, given that the information is not structured and tagged in a dedicated manner intended for computers to use. Simple attempts to navigate to websites that don't want to be crawled are easily detected. Web browsers are big, slow and clunky. Modern pages involve running a lot of often intensive code. Anyone who has kept too many tabs open in their browser knows that this can make your computer very hot, very fast. Computers can load and try to navigate websites a lot faster if they just look at the underlying HTML rather than trying to run the code and create a fully functioning webpage. Yet operating in this manner makes it easy for servers to see that the visitor is not a 'real person,' and then limit their access. It might also simply not give the bots what they need — the useful content on a site might be designed to only appear once such code is fully executed.

Because of this, bots 'puppet' full-fledged browsers so they can interact with the code after it has been fully rendered. The bot puts on a disguise of sorts, and acts in a way that seems more like a human than a machine.[30] This has become more complicated as time has gone on: if the Web in the 1990s was a series of static noticeboards, the Web in the 2020s is a collection of millions or billions of little software applications. Readers may have visited news websites or blogs which dynamically load the next block of text as they scroll down; computers scraping that content would also need to scroll to permit the download of the entire document. Some websites also try to make themselves deliberately difficult to parse via HTML, for instance in their effort to foil attempts to block ads and trackers.

This whack-a-mole game between Web developers and potential scrapers is not a new development. Businesses have long tried to make automation work in such a puppeteering way

29: This has been a classic challenge of computer security since its earliest days, remains largely unsolved, and necessitates the type of holistic, full-stack approach this book proposes. See James P Anderson, *Computer Security Technology Planning Study* (ESD-TR-73-51, Vol I, Electronic Systems Division, Air Force Systems Command, Hanscom Field October 1972) iv.

30: Software libraries that do this include Selenium, Playwright and Puppeteer. For more on headless browsing and the genealogy of the term 'bot,' see Robert Gorwa and Douglas Guilbeault, "Unpacking the Social Media Bot: A Typology to Guide Research and Policy" (2020) 12(2) Policy & Internet 225 DOI: 10.1002/poi3.184 accessed 11 August 2018.

when code alone doesn't do enough. For instance, enterprise software often tries to lock in users to limit interoperability with competitors, and does not provide users with easy, automatic ways to interact with other systems.[31] The platform business model involves reducing friction between a firm's products and leaving it high in relation to external alternatives.[32] To get around this, corporations seeking to make their processes more efficient have, for decades, reached for approaches known as 'robotic process automation.' Instead of redesigning their clunky, pre-existing processes — for example, the way through which human resources departments in firms need to fill out personnel forms when hiring new staff — organizations might drop a new piece of software into the middle, which is taught to expedite certain processes, including by moving the mouse or typing on the keyboard to interact with software as a human might.[33] This so-called 'swivel chair' automation left the creaking, inefficient legacy systems as they were, but sought to replace the human given the bleak task of pivoting on a stool between two intransigent systems that refused to speak to each other directly.[34] This kind of automation was brittle, and would often break with software updates — such as when buttons changed shape or text boxes were slightly repositioned.

2000s-era enterprise software didn't usually try to detect and then prevent this kind of puppeteering, but modern online services often do. This leads to complex cat-and-mouse games between headless browsers and the platforms they interact with. Bots that use the mouse too robotically are easily detected, so 'behavioral' scrapers integrate software packages that smooth digital mouse movements into more realistic Bézier curves.[35] A Web agent can also now navigate in a jerky and hesitant way, as though a human were reading the page. A realistic scraper can browse haphazardly, going back, looking around and taking it (somewhat) slowly, rather than immediately trying to navigate to all the links embedded in a page at the same time.[36] Anti-scraping tools try to identify patterns in these systems and block them, but ultimately, the more human-esque an agent's behavior becomes, the higher the chance that this kind of pattern-finding will create false positive matches and inconvenience real human users.

CAPTCHA'D

Firms facing misuse of their systems via APIs and via systems like the Model Context Protocol may try to disable programmatic access, yet blocking agents that mimic legitimate human users is a trickier task. The recent rise of AI agents has therefore rekindled interest in a long-standing problem for Internet policy: the question of how to accurately separate humans from bots so that they can be treated differently on the Web. This has been implemented through

31: Carl Shapiro and Hal R Varian, *Information Rules: A Strategic Guide to the Network Economy* (Harvard Business School Press 1999).

32: Julie E Cohen, *Between Truth and Power: The Legal Constructions of Informational Capitalism* (Oxford University Press 2019) DOI: 10.1093/oso/9780190246693.001.0001.

33: Leslie Willcocks, Mary Lacity, and Andrew Craig, "Robotic Process Automation: Strategic Transformation Lever for Global Business Services?" (2017) 7(1) Journal of Information Technology Teaching Cases 17 DOI: 10.1057/s41266-016-0016-9 .

34: Michael Hammer, "Reengineering Work: Don't Automate, Obliterate." (1990) 68(4) Harvard Business Review 104.

35: Implemented in the Ghost-Cursor package.

36: Implemented in the Pydoll package.

a variety of interlocking technological systems in practice.[37]

The Completely Automated Public Turing Test to Tell Computers and Humans Apart (CAPTCHA) is a cognitive test that readers will surely be familiar with. These are puzzles designed to be human-easy, but bot-hard. With the rise of mass Web scraping by AI companies, people have become better than ever at identifying bicycles, bridges and traffic lights. Even assuming people only see one CAPTCHA every one and a half weeks, approximately five hundred years of human life are spent solving these each day.[38]

CAPTCHAs are no longer especially secure. AI systems can be just as good as, if not better than, humans at recognizing squiggly words, identifying buses or motorcycles in a grainy photo or dragging jigsaw pieces around. The vast majority of modern CAPTCHAs thus don't just check if a user can answer a puzzle, but also track as much as possible about a user's behavior during the process of solving it. For example, what did their mouse movements look like? How fast were the clicks?[39]

The long-running fight between CAPTCHAs and those trying to circumvent them has always come at a cost. CAPTCHAs pose serious accessibility challenges. Even where alternatives to visual puzzles exist to support blind and partially sighted people, some individuals may also have hearing disabilities, or cognitive disabilities that make navigating audio CAPTCHAs difficult. Many others may not be able to solve CAPTCHA puzzles in English, often the only option presented on major Web services.[40] It is therefore unsurprising that some governments advise against the use of puzzle-based CAPTCHAs entirely.[41]

If solving a puzzle offers only a brittle safeguard, could bots be detected by the unique traces that they send out when accessing services? These traces could be passive, sent out automatically as an intrinsic part of browsing, or active, in response to prompts and pokes sent by the services the bot is trying to access. The most obvious candidate for a passive trace is an IP address. All online interactions come with one by design. But IP addresses are not very useful as a detector. They rotate regularly and are shared among many users, as there are not enough of them for each Internet user to have one or more just for themselves.[42]

With every form of technoregulation comes an effort to defeat it. CAPTCHA 'farms' have existed for decades and have generated sophisticated techniques for solving CAPTCHAs on behalf of bots, using automation and paid human clickworkers.[43] Scrapers can hide their

37: Early legal initiatives do exist, such as New York Senate Bill S9934A (New York Stealth Crawler Prohibition Act), which at the time of writing has passed the legislative chamber and is awaiting the Governor's signature (or veto).

38: Thibault Meunier, "Humanity Wastes about 500 Years per Day on CAPTCHAs. It's Time to End This Madness" (*Cloudflare*, 13 May 2021) https://blog.cloudflare.com/introducing-cryptographic-attestation-of-personhood/ accessed 3 March 2026.

39: Andrew Searles and others, "An Empirical Study & Evaluation of Modern CAPTCHAs" in (32nd USENIX Security Symposium (USENIX Security 23), 2023) https://www.usenix.org/conference/usenixsecurity23/presentation/searles accessed 3 March 2026.

40: W3C, "Inaccessibility of CAPTCHA: Alternatives to Visual Turing Tests on the Web" (16 December 2021) https://www.w3.org/TR/turingtest/ accessed 3 March 2026.

41: See e.g. NZ Government, "CAPTCHA — Web Accessibility Guide" (2023) https://govtnz.github.io/web-a11y-guidance/wct/captcha/ accessed 3 March 2026.

42: More specifically, this applies to IPv4, as Internet providers have not moved sufficiently to IPv6, where addresses are bountiful.

43: Hoang Dai Nguyen and others, "C-Frame: Characterizing and Measuring in-the-Wild CAPTCHA Attacks" in *2024 IEEE Symposium on Security and Privacy (SP)* (2024 IEEE Symposium on Security and Privacy (SP), May 2024) DOI: 10.1109/SP54263.2024.00200; Xiwen Teoh and others, "Are CAPTCHAs Still Bot-hard? Generalized Visual CAPTCHA Solving with Agentic Vision Language Model" in *Proceedings of the 34th USENIX Security Symposium (USENIX Security 25)* (2025) https://www.

IP addresses within residential ranges reserved for private individuals, which both disguises them and distributes their traffic, increasing the collateral damage from any blocking attempts (which will then hit regular users). This is often done using a shady business practice known as a residential proxy network. Such networks are made up of inane mobile apps (think notepads, calculators), cheap connected homeware (think smart photo frames) or cheap VPNs. Businesses behind these sell access to individuals' Internet connections, either totally silently or with questionable permissions obtained through terms in the small print. Even where businesses do not do that, such devices and services might be hacked to create involuntary botnets on behalf of a criminal group.[44] In 2024, the FBI shut down a large residential proxy network installed through compromised free VPN apps, which apparently had compromised 19 million IP addresses globally across 200 countries, enabled over $5 billion of Covid-19 insurance fraud and supported various identity theft and child exploitation operations.[45]

More complex antibot systems exist, but often come with significant privacy, tracking and profiling concerns. Google's reCAPTCHA system produces a risk score based on a user's online behavior over time, even taking into account Google credentials if previously logged in. This entails significant tracking and data transfer, which occurs without either the strict necessity or consent required by European law when analyzing users' devices.[46] Several data protection authorities have issued fines against companies using this version of reCAPTCHA following data protection breaches.[47] Perhaps most privacy-invasive of all is the aptly named firm 'Tools for Humanity,' more widely known as World or Worldcoin, which markets an orb-shaped device that exchanges an iris scan for enrollment in an identity infrastructure supposed to verify that the user is a unique human. Tools for Humanity has been subject to widespread regulatory action, accused of disregarding data protection law and required to erase iris data alongside other compliance measures in jurisdictions including Germany, Kenya and South Korea.[48]

Cloudflare's Turnstile system, used 7–8 billion times every day in 2026, provides another option.[49] Turnstile (as of 2026) uses neither puzzles nor longitudinal tracking. It instead relies

usenix.org/conference/usenixsecurity25/presentation/teoh accessed 3 March 2026.

44: Xianghang Mi and others, "Your Phone Is My Proxy: Detecting and Understanding Mobile Proxy Networks" in *Proceedings of the 2021 ISOC Network and Distributed System Security Symposium (NDSS)* (2021) DOI: 10.14722/ndss.2021.24008.

45: USDepartment of Justice, "911 S5 Botnet Dismantled and Its Administrator Arrested in Coordinated International Operation" (29 May 2024) https://www.justice.gov/archives/opa/pr/911-s5-botnet-dismantled-and-its-administrator-arrested-coordinated-international-operation accessed 3 March 2026.

46: Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) [2002] OJ L201/37 (ePrivacy Directive) art 15.

47: Société CITYSCOOT (CNIL, form restr, 16 March 2023, délib n° SAN-2023-003) (France); Société NS CARDS FRANCE (CNIL, form restr, 29 December 2023, délib n° SAN-2023-023) (France); BVwG 13.09.2024, W298 2274626-1/8E (Austria); Apulaistietosuojavaltuutettu 22 May 2023, Dnro 7684/171/22 (Finland).

48: Kris Shrishak, "Privacy Washing through PETs: The Case of Worldcoin" (*Communications of the ACM*, 28 March 2025) https://cacm.acm.org/blogcacm/privacy-washing-through-pets-the-case-of-worldcoin/ accessed 10 August 2026; *Republic v Tools for Humanity Corporation (US) & 9 others; Katiba Institute & 4 others (ex parte applicants)* [2025] KEHC 5629 (KLR); Korean Data Protection Authority (PIPC), " '월드코인' 관계사에 개인정보 보호법 위반으로 총 11 억 400 만 원 과징금 부과 (Total penalty surcharge of KRW 1,104 million imposed on 'Worldcoin' affiliates for violating the Personal Information Protection Act)" (26 September 2024) https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=10608#LINK accessed 10 August 2026.

49: Leo Bacevicius, Ana Foppa, and Marina Elmore, "The Most-Seen UI on the Internet? Redesigning Turnstile and Challenge Pages" (*The Cloudflare Blog*, 27 February 2026) https://blog.cloudflare.com/the-most-seen-ui-on-the-internet-redesigning-

heavily on technical interactions with the user's browser environment in an effort to check whether it is real or simulated. It runs many different pieces of code to see how the browser interprets them and reacts, looking for tell-tale signs that the device may not be physically operated by a human user.[50] This technique is intended to defend against powerful computers that operate hundreds or thousands of virtual browsers. While Turnstile adds friction, it is not very robust against systems that scrape slowly, using real devices. These kinds of systems can be circumvented by open-source tools such as Scrapling, a highly popular and easy-to-run scraping tool heavily used by deployers of agents and LLMs such as OpenClaw, and easily integrated with major commercial AI tools such as ChatGPT or Claude to circumvent their own scraping limitations.[51]

TOKEN GESTURES

The above strategies center on identifying bot-like patterns in order to block them. What if instead of trying to identify the bots, you sought to identify the real people? This is the direction that antibot systems have been heading in recent years. In doing so, they have entangled themselves with other governance intermediaries and platform middlemen.

One recurring idea in antibot circles is that devices should present some certificate or credential to prove they have a certain attribute, such as that they are human, or that they are using a specific device. A service could then decide on this basis whether to let them in. Early debates around the potential of this approach in the 1990s and early 2000s revealed enormous privacy issues: services presented with certificates could collaborate to track users over time, and the authorities issuing certificates would amass records of potentially sensitive information about those users.[52] This led to proposals for certificates powered by more advanced cryptography, which would allow users to selectively and reliably prove facts about themselves without revealing more about who they were.[53]

In the network layers of the AI governance stack, the most interest has been devoted to a much-simplified version of this idea, the access *token*. The idea is that individuals could undertake some test to prove an attribute, typically the fact they are a real person. The user would then be issued with a finite supply of tokens, each a little cryptographic reward for being real. Perhaps they logged into their personal banking page, which required a lot of multi-factor checks; perhaps they undertook a series of exhausting and complex CAPTCHA challenges; perhaps they verified their phone number or even talked to someone who confirmed they were indeed a realistic-sounding human; perhaps they computationally demonstrated

turnstile-and-challenge-pages/ accessed 4 March 2026.

50: On how this works generally, see a seminal paper introducing the method by Elie Bursztein and others, "Picasso: Lightweight Device Class Fingerprinting for Web Clients" in *Proceedings of the 6th Workshop on Security and Privacy in Smartphones and Mobile Devices* (SPSM '16, Association for Computing Machinery 24 October 2016) DOI: 10.1145/2994459.2994467. This technology is similar to that used in 'cookie-less' Web tracking (fingerprinting), and therefore users attempting to block tracking often find it much harder to pass tests such as Turnstile and browse the modern Web.

51: Reece Rogers, "OpenClaw Users Are Allegedly Bypassing Anti-Bot Systems" (*Wired*, 25 February 2026) https://www.wired.com/story/openclaw-users-bypass-anti-bot-systems-cloudflare-scrapling/ accessed 3 March 2026.

52: Stefan A Brands, *Rethinking Public Key Infrastructures and Digital Certificates: Building in Privacy* (MIT Press 2000).

53: Brands (n 52); Merel Koning and others, "The ABC of ABC: An Analysis of Attribute-Based Credentials in the Light of Data Protection, Privacy and Identity" in J Balcells (ed), *Internet, Law & Politics: A Decade of Transformations.* (Huygens Editorial 2014) http://hdl.handle.net/2066/129933 accessed 16 December 2022.

they have a specific, untampered-with device, with a particular identifier — or, perhaps, they jumped through a mixture of all of these hoops. Unlike certification schemes, tokens do not need to be connected to the identity of the holder. If these tokens could be linked to a challenge that hit the right balance between laborious and user-friendly, then there should be enough for humans browsing the Web at human speed to spend them as they go, while bots seeking to go beyond these speeds would soon hit a rate limit.

Together with academics, Cloudflare proposed and launched such a system in the late 2010s. Dubbed 'Privacy Pass,' it was initially motivated by a desire to help users of the Tor anonymity network browse without hitting a flood of CAPTCHAs that made the darkweb browser unusable.[54] As the bot issue accelerated in the 2020s, Cloudflare gathered various corporate allies and proposed an Internet standard based around this technology.[55]

A problem remained: what would stop those wishing to scrape maliciously from acquiring and hoarding tokens from a wide variety of sources?[56] Such scrapers might recruit a gig worker to undertake the relevant challenges where it was cheapest, slotting right into the model of CAPTCHA-solving services, which offer just that — thousands of solves for a few dollars.[57] The risk of hoarding tokens could be somewhat reduced if tokens were linked specifically to the service that the user wished to visit or scrape, or could not all be spent at the same place, but schemes that enable this without requiring individuals to constantly solve challenges, massively sacrifice network speed or risk decryption by future quantum computers, have proven complicated to design and implement.[58]

As another way of stopping this trade and hoarding, operating system providers have started to tie Web authentication tokens to an individual's device. Instead of stamping them with an identifier, the tokens would be placed in a cryptographically sealed box on the phone — the secure enclave that we introduced in Chapter 3 — which cannot be opened to transfer them to others.

Apple's Private Access Tokens (PATs) system is the key example of this concept in the wild. Integrated into iOS 16 and macOS 13 (released in 2022), PATs are designed to replace CAPTCHAs, proving that a request to access a website comes from a legitimate device,

54: Alex Davidson and others, "Privacy Pass: Bypassing Internet Challenges Anonymously" (2018) 3 Proceedings on Privacy Enhancing Technologies 164 DOI: 10.1515/popets-2018-0026.

55: Alex Davidson, Jana Iyengar, and Christopher A Wood, "The Privacy Pass Architecture" (*IETF RFC 9576*, June 2024) DOI: 10.17487/RFC9576.

56: ibid s 7.1.

57: Marti Motoyama and others, "Re:CAPTCHAs: Understanding CAPTCHA-Solving Services in an Economic Context" (2010) https://www.usenix.org/event/sec10/tech/full_papers/Motoyama.pdf.

58: Two schemes are relevant here at the time of writing in mid-2026. The first, *Anonymous Rate-Limited Credentials*, is to, once a user completes a challenge (like a CAPTCHA or a demonstration they hold an untampered device), issue a temporary token-generation machine that a user can take around the Web with them. When a user visits a website, they can use the machine to generate up to N tokens (N is a number set by that specific website) and can then generate a fresh N (set by the next website) at the next website, and the next, without ever returning to be challenged. The website cannot tell that distinct tokens up to N came from the same person (so it cannot be used to track them). What it knows instead is that no machine can produce more than N, so that whoever has exhausted their allowance cannot return until they complete the challenge again. The second, *Anonymous Credit Tokens*, can be thought of as (after a CAPTCHA/untampered device challenge) giving the user a 'pre-paid card' with access credits on it (e.g. 100 credits), to which a website can charge a customizable toll (e.g. 20 credits), and then hand back a totally fresh card with 20 fewer credits than before (or, if it dislikes what the user just did, hand back nothing at all!). Neither approach yet has a cryptographic construction that is both efficient and robust to future quantum computers. In the case of the first, such a computer would allow someone to store all of this authentication data and then one day link the token spends together to reconstruct browsing patterns. At the time of writing, both have stalled or been discontinued in the Internet Engineering Task Force (IETF), seemingly for these quantum-related reasons.

and a real person, without revealing that person's identity to the website. They are issued by Apple, and such issuing requires that an individual have a modern, untampered device and an iCloud account that is in 'good standing.'[59] At the time of writing, Private Access Tokens allow users to bypass certain CAPTCHAs — Cloudflare, for instance, has enabled users with PATs to automatically skip through its Turnstile system[60] — but are not required to navigate the Web. Yet as alternative methods of bot identification continue to fail, and scraping and malicious agents become more common, more providers could move in the direction of Apple, and accessing the open Internet could become far more difficult for the billions of people around the world not using the latest, authenticated, untampered-with devices.[61]

Private Access Tokens and similar protocols are intensely controversial as a result. In 2023, Google announced its 'Web Environment Integrity' API. Google proposed that websites be given the tools to ask a user's browser or operating system whether or not the user was operating in a 'trusted' environment — a browser that had not been tampered with, for example in an effort to automate it. The attestation that this was true would come from the operating system, and be entangled with technologies such as Apple's Secure Enclave to try to ensure it could not be forged. Critics quickly noted that effective attestation would become an oligopolistic affair, as only the underlying operating system could really make this kind of claim about non-tampering. The keys to who could, or could not, access the Web would thus be given to Apple, Google and Microsoft. Individuals who did not want to use the Internet through the services of a dominant player would get a vastly inferior online experience.[62] Simultaneously, these kinds of attestations assumed that the operating system was able to detect and restrict such automation, which implied significant control and monitoring by operating systems vis-à-vis the software running on user devices. In effect, Web Environment Integrity brought to the network layer the issues of trusted computing and gatekeeper control of operating systems, already discussed extensively in Chapter 3.

Web Environment Integrity could potentially be used to prevent users from installing ad or tracker blockers, or using alternative browser engines or tools — all threats to the major players — lest they be blocked from the services they were trying to access.[63] The proposal posed conflict of interest questions as well: a key motivator for Google to propose Web Environment Integrity was to limit ad fraud, which costs the company significant amounts of money, and is a byproduct of its surveillance-based personalized advertising system where advertisers pay

59: Good standing is not defined, although given the context, we can assume accounts flagged for spam or fraud would not meet this criterion. See Tommy Pauly, "Replace CAPTCHAs with Private Access Tokens" (*Apple*, 2022) https://developer.apple.com/videos/play/wwdc2022/10077/ accessed 2 March 2026.

60: Reid Tatoris and Maxime Guerreiro, "Private Access Tokens: Eliminating CAPTCHAs on iPhones and Macs with Open Standards" (*The Cloudflare Blog*, 8 June 2022) https://blog.cloudflare.com/eliminating-captchas-on-iphones-and-macs-using-new-standard/ accessed 8 March 2026.

61: Similar efforts have been announced by other firms; see Google's 'Play Integrity' system (using, on Pixel devices, the Titan M chip), and Microsoft's 'Trusted Platform Module.'

62: Eric Rescorla, "Private Access Tokens, Also Not Great" (*Educated Guesswork*, 29 August 2023) https://educatedguesswork.org/posts/private-access-tokens/ accessed 12 August 2026; Dennis Jackson, "PACT: Anonymous Credentials for the Web" (*Mozilla Hacks*, 23 June 2026) https://hacks.mozilla.org/2026/06/pact-anonymous-credentials-for-the-web accessed 12 August 2026.

63: Cory Doctorow and Jacob Hoffman-Andrews, "Your Computer Should Say What You Tell It To Say" (*Electronic Frontier Foundation*, 7 August 2023) https://www.eff.org/deeplinks/2023/08/your-computer-should-say-what-you-tell-it-say-1 accessed 5 March 2026.

Google per click. In the face of intense backlash, Google dropped the proposal in 2023.[64] Yet with Private Access Tokens, this direction of travel has not gone away. Instead of proving they are human, users are being asked to prove that they have little control over their devices, and therefore have no tools that could reliably be used to change them — whether that be for legitimate tinkering or misuse.

Rather than relying on hardware to attest that someone is real, another approach could be to provide signals based on networks of reputation. Such a proposal has been forwarded by Mozilla, the creator of the Firefox browser, which does not want to be excluded or have its browser design forcibly altered by operating systems linking deeply with specific hardware.[65] There is some promise to this approach, but it is extremely early, and has significant technical challenges to overcome for it to be workable and provide meaningful guarantees against abuse.

THERE'S A GOOD BOT

If tokens, however acquired, become the norm for humans to access the Web, then how do bots access online services? If bots could access such tokens easily and at scale, then the whole notion of tokens as an effective form of network governance would break down.

This might seem like an odd question. Why not block bots entirely? While there are many malicious bots out there, users around the world appear to have a significant appetite to use bots for online activities. There are good reasons not to shut this down. If bots are polite with respect to website preferences and careful with the load they place on services, this could indeed be a totally legitimate development. Scraping on behalf of users is the fastest-growing type of scraping.[66] Bots can work on behalf of users in useful ways, for example, to seek out goods and services, compare prices and negotiate with sellers.[67] Scraping can be a crucial tool for research, and promote the public accountability of increasingly important online platforms.[68] More generally, the current state of the platformized Web involves ordinary people facing increasingly dehumanized and 'enshittified' customer service interactions, as generative AI chatbots try to shift these labor costs onto users.[69] In this context, there is arguably an emerging consumer protection case for individuals to be able to deploy their own counter-automation in response, such as to navigate websites or even place fully automated phone

64: Google, "Android Developers Blog: Increasing Trust for Embedded Media" (2 November 2023) https://android-developers.googleblog.com/2023/11/increasing-trust-for-embedded-media.html accessed 5 March 2026.

65: Jackson (n 62).

66: David Belson, "The 2025 Cloudflare Radar Year in Review- the Rise of AI, Post-Quantum, and Record-Breaking DDoS Attacks" (*The Cloudflare Blog*, 15 December 2025) https://blog.cloudflare.com/radar-2025-year-in-review/ accessed 5 March 2026.

67: Michal S Gal and Niva Elkin-Koren, "Algorithmic Consumers" (2016) 30(2) Harvard Journal of Law & Technology (Harvard JOLT) 309.

68: Christian Sandvig and others, "Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms" in *Data and Discrimination: Converting Critical Concerns into Productive Inquiry: A Preconference at the 64th Annual Meeting of the International Communication Association* (Seattle, 2014); Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC [2022] OJ L265/1 (DSA) art 40(12) (enshrining a form of scraping rights for authorized research projects); Jacquellena Carrero, "Access Granted: A First Amendment Theory of Reform of the CFAA Access Provision" (2020) 120(1) Columbia Law Review 131.

69: Agathe Balayn and Seda Gürses, "Beyond Debiasing: Regulating AI and Its Inequalities" (*European Digital Rights (EDRi)*, 2021) https://perma.cc/4UAV-3UFB accessed 26 September 2021, 93.

calls to return products or negotiate down their bills.[70] As with so many of the tools that we talk about in this book, agentic AI systems (and the crawlers underpinning them) are dual-use technologies.

This situation means that segmenting humans from bots and simply banning the latter is likely to be a non-starter. Instead, even if humans are segmented from bots, for the bots that remain, various networked gatekeepers will want to know: do you, visitor, have good intentions?

RISE OF THE ROBOTS.TXT

Allowing the good bots while keeping out the bad is not a new idea. In the 1990s, website operators were also struggling to deal with a deluge of annoying and disruptive scrapers and crawlers. Just like today, bots swamped servers with rapid-fire requests and created many issues for Web admins, frequently breaking code and interacting with systems that they weren't meant to access. In response, at the end of June 1994, a mailing list of Internet hobbyists and developers — bringing together "the majority of [web] robot authors and other people with an interest in robots" — came up with a simple solution.[71] System administrators could place a text file named *robots.txt* in the main directory of their domain.[72] In that file, one could provide instructions to Web agents, stating what they should not do and which parts of a website should not be accessed and crawled. For example, deploying the following code snippet would state the policy that all bots (the asterisk) are unauthorized to access any part of the website (the slash, meaning everything after the domain name):

```
User-agent: *
Disallow: /
```

Unlike the CAPTCHA, *robots.txt* is not technoregulation in and of itself. A text file is not a magic spell that can unilaterally change the practices of bots that are designed to ignore it. It classically relies on respect — a voluntary norm, a best practice, good 'netiquette.' Until 2022, it was not even an official standard.[73] But it was, in the terms we introduced at the start of the book, a governance *loop*. It is a technical artifact that other stack actors might later *hook* onto to achieve their own, or collective, ends.

Even some of the Internet's heroes do not respect *robots.txt*, and for quite good reasons. The Internet Archive's scraper, which has accumulated copies of hundreds of billions of websites since the late nineties, no longer honors it when building the snapshots behind its *Wayback*

70: See e.g. automated tools designed for citizens to exercise rights or engage in disputes, such as DoNotPay or Kudos AI. See relatedly, on tools to rebalance power within systems optimizing for company aims only, Bogdan Kulynych and others, "POTs: Protective Optimization Technologies" in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (FAT* '20: Conference on Fairness, Accountability, and Transparency, ACM 2020) DOI: 10/ggjf5m.

71: Martijn Koster, "A Standard for Robot Exclusion" (30 June 1994) http://www.robotstxt.org/orig.html accessed 16 November 2024.

72: This means, for example, *example.com/robots.txt*.

73: Martijn Koster and others, *Robots Exclusion Protocol* (Request for Comments, RFC 9309, Internet Engineering Task Force 2022) DOI: 10.17487/RFC9309.

Machine.[74] The UK Conservative Party is in part to blame for this. In 2013, the party both deleted ten years of speeches from its website and added a new, restrictive *robots.txt* file. This new file meant that, under the Internet Archive's policy and configuration at the time, the library did not just refrain from scraping going forward, but scrubbed that site's past archiving history too.[75] Confronted with this, the Archive changed its policy soon after. The public interest may well be served by exemptions to no-crawling norms, especially where powerful actors try to make certain information less visible and accessible for their own purposes.

In light of the server-pounding described earlier in this chapter, websites have generally sought to become more restrictive on permission to access. The number of *robots.txt* objections in Web domains commonly used for AI training rose by more than 500% in 2023–24, and this number has continued to grow.[76] The terms of service of many of these websites also now explicitly do not consent to scraping by AI companies.[77] Reputable sources forbid AI disproportionately more than poor-quality sources.[78] Many general-purpose model and agentic technology providers have been found disregarding *robots.txt* entirely.[79]

How can such objections be enforced? Law is a classic way of making norms binding, and there are various instruments allowing people to automatically object to the collection and use of content and data. European copyright law gives an exception to intellectual property rights for 'text and data mining.' Unless carried out for the purpose of scientific research, rightsholders can prevent this exception from being engaged for their content using a machine-readable signal.[80] This typically relies on private enforcement through the courts, but the AI Act places a complementary regulatory obligation grounded in public law on general-purpose model providers to comply with such signals.[81] A mirror provision, known as the 'right to ob-

74: Mark Graham, "Robots.txt Meant for Search Engines Don't Work Well for Web Archives" (*Internet Archive Blogs*, 17 April 2017) https://blog.archive.org/2017/04/17/robots-txt-meant-for-search-engines-dont-work-well-for-web-archives/ accessed 9 October 2024.

75: Jill Lepore, "Can the Internet Be Archived?" (*The New Yorker*, 19 January 2015) https://www.newyorker.com/magazine/2015/01/26/cobweb accessed 9 October 2024.

76: Shayne Longpre and others, "Consent in Crisis: The Rapid Decline of the AI Data Commons" in *Proceedings of the 38th International Conference on Neural Information Processing Systems* (NeuRIPS '24, 10 December 2024) DOI: 10.48550/arXiv.2407.14933 5; Belson (n 66). The growth is perhaps leveling off, and a small proportion of sites have become selectively more permissive, likely due to licensing arrangements with the big AI firms. The Longpre paper also has an error rate of about 10% in parsing *robots.txt* files. See, making both claims, Enze Liu and others, "Somesite I Used To Crawl: Awareness, Agency and Efficacy in Protecting Content Creators From AI Crawlers" (IMC '25, Association for Computing Machinery 15 October 2025) DOI: 10.1145/3730567.3732913 82–83.

77: Longpre and others (n 76) 6.

78: Nicolas Steinacker-Olsztyn, Devashish Gosain, and Ha Dao, "Is Misinformation More Open? A Study of Robots.Txt Gatekeeping on the Web" (WWW '26, Association for Computing Machinery 12 April 2026) DOI: 10.1145/3774904.3792625.

79: Katie Paul, "Multiple AI Companies Bypassing Web Standard to Scrape Publisher Sites, Licensing Firm Says" (*Reuters*, 21 June 2024) https://www.reuters.com/technology/artificial-intelligence/multiple-ai-companies-bypassing-web-standard-scrape-publisher-sites-licensing-2024-06-21/ accessed 8 October 2024; Jess Weatherbed, "Anthropic's Crawler Is Ignoring Websites' Anti-AI Scraping Policies" (*The Verge*, 25 July 2024) https://www.theverge.com/2024/7/25/24205943 accessed 8 October 2024; Elena Yovcheva, "Why We Blocked Bytespider and ClaudeBot" (*Servebolt*, 27 May 2024) https://servebolt.com/articles/servebolts-decision-to-block-bytespider-and-claudebot/ accessed 8 October 2024. As discussed below, with the private enforcement by Cloudflare, larger model providers appear to respect these preferences more than they used to, to avoid being blocked. See Liu and others (n 76). However, agentic tools are still commonly disregarding *robots.txt*. See Gabriel Lopez-Fonseca and others, "Do Generative AI Assistants Respect Robots.txt? Tracing Web Access Beyond Visible Answers" (*arXiv*, 19 July 2026) DOI: 10.48550/arXiv.2607.14447.

80: Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC [2019] OJ L62/92, art 4(3). Similar legal constructions are proposed in Hong Kong and Brazil at the time of writing.

81: Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules

ject,' exists in data protection and privacy laws in the EU, UK, India and U.S. states, although the focus here has been on users sending machine-readable signals from their Web browsers to limit the use of tracking data rather than lodging objections alongside personal data they publish online, and not all of these laws apply as broadly as the European provisions do.[82]

However, checking compliance against these legal standards is laborious and costly. Data protection and privacy regulators have tried to enforce Web scraping restrictions with limited success. Clearview AI, a firm notorious for collating enormous volumes of biometric facial templates to sell to law enforcement agencies for the identification of people, has broadly escaped accountability as laws either fail to apply extraterritorially or, after enormous expense, regulators fail to collect the large fines they have levied.[83] The EU AI Act envisages centralized regulators controlling the actions of a few key firms, especially when it comes to general-purpose model developers. These are now directly regulated by the European Commission and indirectly through various voluntary clubs known as codes of practice.[84] Obligations to respect scraping preferences fall on those placing models on the market or making them available, not the long tail of model users.[85] They are likely to bite particularly softly on the users of malicious models pursuing criminal ends, who have little to fear from market regulation, given they do not seek to participate in legitimate markets. And this is all assuming such cases even smoothly get to adjudication and enforcement. If an earlier battle around the voluntary 'Do Not Track' signals is anything to go by, companies will spend an enormous amount of time debating exact definitions and highlighting vagueness to attempt to make any regime of automated, enforced objection seem unworkable before it begins.[86]

Avoiding regulators in Internet law typically means looking for robust technoregulation — self-enforcing solutions baked into code. However, assuming websites are to allow some bots

on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L (AIA) art 53(1)(c); see generally Michael Veale and João Pedro Quintais, "The Obligations of Providers of General-Purpose AI Models" in Gianclaudio Malgieri and others (eds), *EU AI Act: Thematic Commentary* (Hart 2026) DOI: 10.5040/9781509988570.ch-022.

82: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1 (GDPR) art 21, which assumes, as regulators broadly do, that scraping can be based on the 'legitimate interest' legal basis, see European Data Protection Board, "Opinion 28/2024 on Certain Data Protection Aspects Related to the Processing of Personal Data in the Context of AI Models" (*EDPB*, 17 December 2024) https://www.edpb.europa.eu/system/files/2024-12/edpb_opinion_202428_ai-models_en.pdf accessed 18 December 2024; see generally on wider provisions in the U.S. and beyond, including the current situation of 'Global Privacy Control' or GPC Sebastian Zimmeck and others, "Can the GPC Standard Eliminate Consent Banners in the EU?" (2026) 61 Computer Law & Security Review 106332 DOI: 10.1016/j.clsr.2026.106332; note also that unlike a copyright reservation of rights, an objection under European data protection law does not guarantee that the processing stops, but instead forms evidence during a balancing test.

83: See generally Lydia Emmanouilidou, "Clearview: Πώς η σκιώδης αμερικανική εταιρεία συνεχίζει να σκανάρει τα πρόσωπά μας" (*Solomon*, 16 April 2025) https://wearesolomon.com/mag/format-el/erevnes/clearview-pws-i-skiwdis-amerikaniki-etaireia-sinexizei-na-skanarei-ta-proswpa-mas/ accessed 17 August 2026. This has led NGOs to now abandon a strategy of regulatory action against the firm in favor of attempting to find its directors criminally liable, see e.g. noyb, "Criminal Complaint against Facial Recognition Company Clearview AI" (*noybeu*, 28 October 2025) https://noyb.eu/en/criminal-complaint-against-facial-recognition-company-clearview-ai accessed 17 August 2026.

84: Veale and Quintais (n 81).

85: AIA, art 53. A prohibition on using AI systems to scrape the Web to create a dataset of biometric data (AIA, art 5(1)(e)) — directly influenced by the outrage about the activities of Clearview AI — does apply to model users, but is narrow and specific in nature.

86: Irene Kamara and Eleni Kosta, "Do Not Track Initiatives: Regaining the Lost User Control" (2016) 6(4) International Data Privacy Law 276 DOI: 10/gdxwds.

rather than try to block every non-human visitor, it is unlikely that such sites will have enough of a signal themselves to work out which bots are trustworthy and which are not just from appearances. That would be about as reliable as inferring the safety of a driver from their car's color or make. Quickly distinguishing between bots and humans is much easier than distinguishing between good bots and bad bots.

What might work, however, is an entity that could monitor and connect bots' behavior over many different sites. Websites could make strong efforts to block *all* bots by default, segregating human and non-human traffic as best they could, and then, for non-human traffic, only let in the bots that are vouched for, somehow, as 'good,' collating feedback on their behavior for others to rely upon. Any entity organizing and implementing such vouching represents a potential bottleneck to disrupt misuse, and may be able to apply rules and obligations beyond just *robots.txt*, such as detecting automated scams, fraud and other types of illegal abuse.

These entities exist — they are the same entities that produce CAPTCHAs and similar technologies. Cloudflare is the most notable of the bunch.

BOT VERIFICATION

Cloudflare's power emerged in part because bots have always presented risks to websites. Today's robot wars are just the latest chapter in this long-running saga.

A popular Web service cannot survive online without serious protection. This fact has shaped the infrastructure behind the modern Internet. In its early days, individuals could already use computers to send such a large volume of messages so cheaply that it both caused cascades of unwanted 'spam' and threatened the stranglehold on mass communications held by government and the mass media.[87] This dynamic also led to the emergence of various forms of electronic civil disobedience, often termed 'hacktivism.' Hacktivism popularized the use of the generative aspects of computing — its low cost, speed, flexibility and power — for the purpose of protest.[88] From the 1990s until the early 2010s, there were 'netstrikes' on French government websites in retaliation for nuclear experiments in Polynesia, attacks by the group *Anonymous* against the Church of Scientology and *LulzSec*'s attacks against a range of corporations including Sony Pictures and News Corp.[89] This kind of activity was not just limited to 'hackers' and those with technical know-how. Easily downloadable software such as the Low Orbit Ion Cannon was developed to allow lay users to either target sites of their choice or sign their computer up to be part of a voluntary botnet which would take aim at Web properties later selected by hacktivist groups.[90]

These types of distributed denial of service (DDoS) attacks, where servers are overwhelmed by coordinated external traffic, have been mitigated by firms mostly through adding extremely complex network intermediaries between them and their users. The main companies in this area, Cloudflare, Akamai and Fastly, typically started off selling technologies designed to make

87: See generally Finn Brunton, *Spam: A Shadow History of the Internet* (MIT Press 2015).

88: Stefania Milan, "WikiLeaks, Anonymous, and the Exercise of Individuality: Protesting in the Cloud" in Benedetta Brevini, Arne Hintz, and Patrick McCurdy (eds), *Beyond WikiLeaks: Implications for the Future of Communications, Journalism and Society* (Palgrave Macmillan UK 2013) DOI: 10.1057/9781137275745_12 552.

89: ibid.

90: Gabriella Coleman, *Hacker, Hoaxer, Whistleblower, Spy: The Many Faces of Anonymous* (Verso 2014).

the Internet faster, marketing themselves as content delivery networks (CDNs) that would cache pages geographically near to users. Before long, they found that this line of business could also be used to defend servers from malicious traffic. By using algorithmic classifiers and IP reputation databases that try to sift out legitimate from potentially illegitimate traffic, these firms have positioned themselves in a place where they can inject friction (such as CAPTCHAs), impose rules or enforce blocklists, all before the user is even allowed to approach the website's front door.

In 2024, as multiple Web properties began getting bombarded by out-of-control AI scrapers, Cloudflare promised customers "AIndependence" from bots that behave "dishonestly."[91] This has come to be defined as failing to meet the firm's 'verified bots policy,' which includes prohibitions on certain agent activities — including ticket scalping, denial-of-service attacks, various forms of hacking and scraping competitors' prices — and enforces adherence to *robots.txt* signals.[92]

The core requirement of Cloudflare's policy is that bots should identify themselves as such. This is done either by the bot originating from specific IP addresses pre-registered with Cloudflare, or by the bot flashing a cryptographic credential when visiting a site.[93] The authentication does not just signal to Cloudflare that the visitor is a bot; it signals *which* bot, which means that the firm — which guards over 20% of Web domains[94] — can choose whether to open or close its virtual doors.[95]

Under what conditions these virtual doors open and close is not yet settled. Broadly, sites protected by Cloudflare can open the doors to specific categories of registered bots while shutting others out. A bot obtains and maintains its registration status in its specific category by acting honestly and non-abusively, which includes following *robots.txt* files, including parts of the file that state for which purposes scrapers are allowed to use content from varying parts of the page.[96]

Robots.txt as a standard does not allow websites to restrict parts of the website to only certain purposes of scraping. It allows sites to permit certain identities of scraper in relation to certain parts of the site. Scrapers are often used for multiple purposes. Googlebot is one example: its operator has used this bundling to force websites to stay accessible to its AI RAG systems, by stating that any opt-out also means an opt-out of Google search indexing. As a way of dealing with this, Cloudflare has used its network power to redefine *robots.txt* to allow sites to specify which purposes scraping is or is not allowed for, such as Web search, AI training or different types of AI use such as paraphrasing or summarization.[97] Their market share means that they

91: Alex Bocharov and others, "Declare Your AIndependence: Block AI Bots, Scrapers and Crawlers with a Single Click" (*The Cloudflare Blog*, 3 July 2024) https://blog.cloudflare.com/declaring-your-aindependence-block-ai-bots-scrapers-and-crawlers-with-a-single-click accessed 8 October 2024.

92: Cloudflare, "Verified Bots Policy" (*Cloudflare Docs*, 1 July 2026) https://developers.cloudflare.com/bots/concepts/bot/verified-bots/policy/ accessed 1 July 2026.

93: Using the *Web Bot Auth* protocol, originally authored by Cloudflare and submitted to the Internet Engineering Task Force (IETF) for consideration, now at the time of writing an official working group. Thibault Meunier and Sandor Major, "HTTP Message Signatures for Automated Traffic (Internet-Draft)" (*Internet Engineering Task Force*, 18 August 2026) https://datatracker.ietf.org/doc/draft-meunier-webbotauth-httpsig-protocol accessed 25 August 2026.

94: Jin-Hee Lee and Bryan Becker, "Your Site, Your Rules: New AI Traffic Options for All Customers" (*Cloudflare Blog*, 1 July 2026) https://blog.cloudflare.com/content-independence-day-ai-options/ accessed 6 August 2026.

95: Meunier and Major (n 93).

96: Cloudflare, "Verified Bots Policy" (n 92).

97: They first used this power to add a 'crawl-delay' variable to prescribe a minimum time a bot must wait before querying a

do not need a standards body to bless these changes — Cloudflare simply threatens to block bots that do not comply with any sites that make use of the new features it designs.[98] The firm has also been lobbying competition authorities to require 'crawler separation,' where each scraping identity ties to a specific purpose, which would make blocking unwanted purposes much easier for Cloudflare.[99]

Cloudflare is positioning itself as more than just protective software you pay for to keep bots at bay. It has been explicit about wanting to use this power to act as a gatekeeper that could charge AI companies for access to certain websites, splitting the spoils with its customers.[100] The commercial ends of Cloudflare are beyond the scope of this book. What matters, however, is that Cloudflare is aligning technoregulation to focus on the *purpose* of bots, and to create an infrastructure that gives it granular detection and blocking capability.

Such purpose-based bot blocking is exactly the capability which may be called on to detect and prevent abuse arguably much more serious than downloading a few news articles or DIY repair instruction guides.

NETWORKED GATEKEEPERS

There are few intermediaries with enough network visibility and influence to be policemen overseeing scrapers and agents trawling the Web. Cloudflare and its competition are not just well-positioned to potentially do this work, but they are also incentivized to do so, understanding purpose-oriented bot gatekeeping as a major future revenue stream.

This again is not an entirely new role for the firm. Indeed, Cloudflare's infrastructural shielding had already become so important for Internet governance that it created an entirely new form of cyber-activism in the mid-2010s. Coordinated advocacy campaigns could now involve efforts to take certain websites de facto 'offline,' not by actually DDoSing or hacking them, but rather by lobbying services like Cloudflare to withdraw their protective shielding from certain sites — leaving those sites alone in a dark forest of potential attackers.[101] In 2017, Cloudflare cut off its services to the neo-Nazi website The Daily Stormer; in 2019, it removed its shielding from 8chan, an image board popular among white supremacist shooters as a place to post their manifestos; in 2022, it made a similar decision relating to the online sextortion

site again. See Reid Tatoris and Pawel Klimek, "Easily Manage AI Crawlers with Our New Bot Categories" (*The Cloudflare Blog*, 29 September 2023) https://blog.cloudflare.com/ai-bots accessed 9 October 2024; see also the standard in Koster and others (n 73). They then added *content-signal*, which transforms *robots.txt* from a request of who can and cannot scrape to a statement of which scraping purposes are permitted or not. See Cloudflare, "robots.txt Setting" (*Cloudflare Docs*, 3 August 2026) https://developers.cloudflare.com/bots/additional-configurations/managed-robots-txt/ accessed 7 August 2026.

98: On this type of power, see Cohen (n 32) chapter 7.

99: See e.g. Cloudflare, "Cloudflare Response to the CMA Consultation on the Publisher Conduct Requirement Relating to Google's General Search Services" (25 February 2026) https://assets.publishing.service.gov.uk/media/69b970d9f6acd3f43e96129b/Cloudflare.pdf accessed 18 August 2026.

100: Katie Drummond, "Matthew Prince Wants AI Companies to Pay for Their Sins" [2025] Wired https://www.wired.com/story/big-interview-podcast-matthew-prince-cloudflare/ accessed 6 March 2026; Rohin Lohe, Justin Ridgely, and Will Papper, "Announcing the Monetization Gateway: Charge for Any Resource behind Cloudflare via X402" (*Cloudflare Blog*, 1 July 2026) https://blog.cloudflare.com/monetization-gateway/ accessed 7 August 2026.

101: Katherine Alejandra Cross, "In the War Between Harassment and Censorship, No One Wins" (*Wired*, 22 September 2023) https://www.wired.com/story/kiwifarms-tech-ethics-networked-harassment/ accessed 2 August 2024; Seán Looney, "Content Moderation through Removal of Service: Content Delivery Networks and Extremist Websites" (2023) 15(4) Policy & Internet DOI: 10.1002/poi3.370.

and harassment website Kiwi Farms.[102] This kind of 'infrastructural deplatformization' was a blunt form of online content governance, but it was nonetheless effective in disrupting access to certain Web services when few other leverage points existed.[103]

As these firms take strategic roles in governing scraping, the focus may shift from whether the customers they are protecting are legitimate to how they might govern the behavior of users.[104]

An open question remains regarding bots that work on behalf of users — users with good or bad intentions — but scrape via a third party.[105] Is it the authentication of the user, or the bot infrastructure provider, that matters? How is responsibility distributed?

Where bots are run on behalf of users, there are tricky governance challenges. Most obvious are the privacy questions: how can network intermediaries monitor bots that act on behalf of users without also surveilling things about the users themselves in the process? There are also power questions: who benefits from controlling which bots can and cannot access what, and the more fundamental question of who gets a bot in the first place? Finally, there will be new twists on classic content moderation questions: what kind of activity is permissible? Should *robots.txt* or similar settings apply to individuals' bots too? How can policymakers best protect consumers and citizens in this complex, networked and very difficult-to-reason-about world?

Regardless of how things pan out in terms of market structure, Cloudflare is emerging as a key point of visibility. It has large amounts of data on bots, their patterns and origins. Such visibility might help track down the users behind RAGs and agents, or at least allow other actors in the stack to get closer to them. It is very likely that the need for this visibility will only get greater going forward.

⋆

The network, and its interaction with AI tools and plugins like RAG systems, is an under-explored, often missing site of potential governance in AI policy conversations. Nevertheless, there are numerous interesting developments that will be worth monitoring closely in the coming years. Perhaps more than any other layer in our full-stack schema, the network layer illustrates how important an understanding of classic Internet governance and policy questions will be to holistically and responsibly regulate AI misuse.

Not all types of misuse are easily governed at this layer. It is not currently self-evident what roles RAGs and agents will play in some types of misuse going forward, such as the generation of child sexual abuse imagery — although we do not doubt that abusers will have the ingenuity to think of some in the future.

102: It was during this period that Cloudflare removed from its website one of its seven 'warrant canaries': that it has 'never terminated a customer or taken down content due to political pressure.'

103: Suzanne van Geuns and Corinne Cath-Speth, "How Hate Speech Reveals the Invisible Politics of Internet Infrastructure" (*Brookings*, 20 August 2020) https://perma.cc/DEX7-5MTA accessed 10 October 2021; José Van Dijck, Tim de Winkel, and Mirko Tobias Schäfer, "Deplatformization and the Governance of the Platform Ecosystem" (2023) 25(12) New Media & Society 3438 DOI: 10.1177/14614448211045662; Christoph Busch, "Regulating the Expanding Content Moderation Universe: A European Perspective on Infrastructure Moderation" (2022) 27 UCLA Journal of Law and Technology 32.

104: They already face legal pressure in related areas; Cloudflare, as we were finalizing this manuscript, lost an appeal in Italy in relation to creating a blocking regime for alleged piracy websites, which themselves are not Cloudflare customers. See T.A.R. Lazio, Roma, sez. IV, 17 luglio 2026, n. 13201 (ric. nn. 6084/2025 e 10700/2025, riuniti), Cloudflare Inc. c. Autorità per le garanzie nelle comunicazioni.

105: See this playing out in the courts in *Amazon.com Servs., LLC v. Perplexity AI, Inc.*, _ F.4th _ (9th Cir. 2026).

The network layer is one that those looking to disrupt generative AI abuse should be aware of, but interventions there should be pursued carefully. Blocking is a blunt instrument. In this chapter, we have not extensively discussed classic blocking methods — Web blocking, DNS blocking or similar — mostly as they are not very AI-specific and are relatively familiar to law enforcement, telcos and courts. Intermediaries here have some experience as governors, but not much. Cloudflare has been vocal about its unwillingness to use its content moderation powers as openly as, say, a social media company does. This is justified. Network-level blocking is often crude, and typically means taking down an entire service, the good and bad parts together, leading to significant free expression concerns as a result.

But governing bots is not straightforward. Infrastructures in the network layers are already trying to separate humans from bots (formerly via CAPTCHAs, now increasingly through access tokens) and to separate good bots from their bad counterparts. Much of this governance is designed to stop abuse that stems from bots' ability to operate at volumes beyond human scales. It is unclear whether it is well-suited for the type of abuse we focus on in this work, where the quality of the activity rather than its quantity is what makes it so undesirable and urgent to tackle.

If this human-bot separation were to be widely implemented, it could make governance more precise for network gatekeepers. Protocols proposed by Cloudflare, such as the ability to flash a credential to show you are a good or trusted bot, are designed to target very specific actors.[106] Such authentication can hook into governance loops from other parts of the stack, such as information about untrustworthy actors, servers, systems or behavior transmitted via logging systems. If other stack actors can find evidence that certain users or their agents have been behaving badly, then they could share information with groups of CDNs like Cloudflare so that they are blocked on much of the Web. The visibility that network intermediaries can have into certain kinds of crawling behavior may well allow them to pass information to areas of the stack where further services can be disrupted: for instance, to communications intermediaries (Chapter 6) where accounts with the same IP addresses might be posting.

The casualty of these developments, however, might be the open Internet. The picture painted above is, at times, quite bleak. To react to the flood of AI bots, humans wanting to use popular services may be restricted to only accessing them on certain, untampered devices, potentially with some form of proof of identity or humanity. To make use of AI themselves to browse and collate information, they may need to ask permission from a handful of technology giants to help them. There is a serious risk that these developments will consolidate the monopolistic power of certain gatekeepers on the network layer and beyond. Regulators and legislators should balance protecting the victims of AI-enabled misuse with attempting to protect a somewhat flat, open Internet.

106: Meunier and Major (n 93).