UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

REDDIT, INC., Plaintiff, - against - SERPAPI LLC, OXYLABS UAB, AWMPROXY, and PERPLEXITY AI, INC., Defendants.	Case No.: 25-cv-8736 **FIRST AMENDED COMPLAINT** **JURY TRIAL DEMANDED**

Reddit, Inc., by and through its undersigned attorneys, alleges, upon knowledge with respect to its own actions and upon information and belief as to all other matters, as follows:

NATURE OF THE ACTION

1. Reddit, Inc. brings this action to stop the industrial-scale, unlawful circumvention of data protections by a group of bad actors who will stop at nothing to get their hands on valuable copyrighted content on Reddit. Defendants Oxylabs UAB, AWMProxy, and SerpApi LLC–respectively, a Lithuanian data scraper, a former Russian botnet, and a Texas company that publicly advertises its shady circumvention tactics—are data-scraping service providers who specialize in creating and selling tools designed to circumvent digital defenses and scrape others' content. These tools are aimed at bypassing two levels of security: First, evading Reddit's own anti-scraping measures, and second, circumventing Google's controls and scraping Reddit content directly from Google's search engine results. In a very real sense, these Defendants are similar to would-be bank robbers, who, knowing they cannot get into the bank vault, break into the armored truck carrying the cash instead. Meanwhile, Defendant Perplexity AI, Inc., an artificial intelligence ("AI") company that is more akin to a "North Korean hacker," uses at least one of its Co-Defendants' tools in order to get the Reddit data it desperately needs to fuel its

"answer engine." It is apparently willing to do anything to obtain that data—that is, anything *other than* enter into an agreement with Reddit directly, as some of its competitors have done. Reddit brings this action to stop each of these bad actors.

2. Founded over 20 years ago, Reddit is one of the largest repositories of human conversation in existence. Every day, over 100 million unique users engage in discussions across its hundreds of thousands of interest-based communities (or "subreddits"), a continuous stream of real-time and creative copyrighted works. This vast corpus of authentic human discourse is "like manna from heaven" for the growing number of companies vying for dominance in the AI market. These AI companies, worth up to hundreds of billions of dollars, desperately need access to more and more high quality, current data to support their ambitions, and Reddit is a "top-cited source" of data for them.

3. But Reddit has rules. It does not permit unauthorized commercialization of Reddit content absent an express agreement with guardrails in place to ensure that Reddit's and its users' rights are protected. In short, if AI companies want to legally access Reddit data, they need to comply with Reddit's policies. Some of the largest AI companies, like OpenAI and Google, have done just that, entering into agreements permitting them to access Reddit data while ensuring that the Reddit community is protected from abuse. That is not the path Defendants have chosen.

4. The problem for Defendants, however, is that Reddit is protected by technological-control barriers to prevent unscrupulous scrapers from accessing and stealing data directly from its website. It has also issued cease-and-desist letters to companies that have tried

to get around those barriers and even sued an AI company worth $183 billion for misappropriating Reddit data.[1]

5. Recognizing that Reddit denies scrapers like them access to its site, Defendants SerpApi, Oxylabs, and AWMProxy scrape the data from Google's search results instead.[2] They do so by masking their identities, hiding their locations, and disguising their web scrapers as regular people (among other techniques) to circumvent or bypass the security restrictions meant to stop them. For example, during a two-week span in July 2025, Defendants SerpApi, Oxylabs, and AWMProxy circumvented Google's technological control measures and automatedly accessed, without authorization, almost *three billion* search engine results pages ("SERPs") containing Reddit text, URLs, images, and videos.

6. Separately, Reddit caught Perplexity red-handed by using the digital equivalent of marked bills (to use the bank robbery analogy) to track Reddit data and confirm that Perplexity was using Reddit data acquired through the scraping of Google SERPs. Perplexity knows that what it is doing is wrong because Reddit told it so in a cease-and-desist letter. Reddit explicitly told Perplexity that (1) Reddit prohibits commercial use of its content absent agreement, (2) this prohibition applies regardless of the means through which Perplexity obtained the data, and (3) technological control systems were in place within Reddit to prevent Perplexity from taking its data. But rather than respect Reddit's and its users' rights, what Perplexity has done in response is simply come up with increasingly devious schemes to circumvent Reddit's security systems and policies. In fact, Perplexity's citations to Reddit increased *forty-fold* after Reddit told it to

---

[1] On June 4, 2025, Reddit filed a Complaint against Anthropic for breach of contract, unjust enrichment, trespass to chattels, and unfair competition. *See Reddit, Inc. v. Anthropic, PBC*, Case No. CGC-25-625892 (Super. Ct., S.F. Cty. June 4, 2025), Dkt. 1.

[2] Since filing its Complaint in October 2025, Reddit has identified evidence suggesting that one of the Defendants may also be scraping content directly from Reddit's website.

stop. And as an advertised client of SerpApi, there can be little doubt where and how Perplexity is getting its illicit Reddit data.

7. Congress has enacted laws to prevent exactly what Defendants are doing: circumventing or bypassing technological measures that effectively control access to copyrighted works. *See* Digital Millennium Copyright Act, 17 U.S.C. § 1201, et seq. Each of the Defendants in this action is profiting by evading technological control measures to access Reddit data it knows it does not have permission to access or use. Because Reddit has always believed in the open internet, it takes its role as a steward of its users' communities, discussions, and authentic human discourse seriously. Through this action, Reddit seeks to end Defendants' circumvention of security measures protecting Reddit data, blatant misuse of Reddit content, and disrespect for its users' rights, all of which harm Reddit and its hundreds of thousands of authentic human communities.

PARTIES

8. Plaintiff Reddit's mission is to empower communities and provide human perspectives to everyone. Reddit is a corporation organized under the laws of Delaware with its principal place of business at 303 2nd Street, South Tower, 5th Floor, San Francisco, California 94107. Reddit maintains an office at One World Trade Center in New York, New York with over 300 employees.

9. Defendant SerpApi is a Texas limited liability company with its principal place of business at 5540 N. Lamar Blvd., Austin, Texas 78756. SerpApi provides products and services to enable web scraping and allows users to bypass or otherwise circumvent technological control measures that prevent large-scale access and web scraping.

10. Defendant Oxylabs is a Lithuanian company with its principal place of business at 32 Svitrigailos St. 03230, Vilnius, Lithuania. Oxylabs provides products and services to enable web scraping and allows users to bypass or otherwise circumvent technological control measures that prevent large-scale access and web scraping.

11. Defendant AWMProxy is a web domain currently registered in California as of March 30, 2025. AWMProxy was previously operated by a Russian entity in connection with a cybercriminal "botnet" that was shut down in 2021.[3] It has apparently resumed operation, selling access to "proxy services" that are used by entities to conceal their location and hide their identity as industrial-scale scrapers, thereby allowing users to bypass or otherwise circumvent technological control measures designed to prevent unauthorized large scale access and web scraping.

12. Defendant Perplexity calls itself an "artificial intelligence powered answer engine." Perplexity is a Delaware corporation with its principal place of business in San Francisco, California. Perplexity maintains an office at 299 Park Ave., New York, New York 10171 with dozens of employees in its New York office. Defendant Perplexity is a customer of Defendant SerpApi.[4]

JURISDICTION AND VENUE

13. Reddit seeks damages, injunctive, and other equitable relief against each of the Defendants under 17 U.S.C. § 1201, et seq.

---

[3] Krebs on Security, *The Link Between AWM Proxy & the Glupteba Botnet* (June 28, 2022), https://krebsonsecurity.com/2022/06/the-link-between-awm-proxy-the-glupteba-botnet/ (last accessed Feb. 6, 2026).

[4] *See* Ex. A (https://serpapi.com/#features (listing Perplexity as a SerpApi customer) (last accessed October 21, 2025)) and Ex. B (https://serpapi.com/#features (listing Perplexity as a SerpApi customer) (last accessed February 6, 2026)).

14. This Court has jurisdiction over this civil action under 28 U.S.C. §§ 1331, 1338, and 1367.

15. Venue is proper in this judicial district under 28 U.S.C. §§ 1391 and 1400(a).

SerpApi Personal Jurisdiction

16. Defendant SerpApi is subject to the jurisdiction of this Court pursuant to N.Y. C.P.L.R. § 302(a)(1), (2), and (3) and the Due Process Clause of the Constitution of the United States. It has purposefully directed its activities at New York and has purposefully availed itself of the benefits of doing business in New York.

17. SerpApi transacts business in the State and District and supplies goods and services in the State and District. SerpApi prides itself on its Google Search API tool, through which users can access and scrape Google search engine results. SerpApi encourages users to input location-specific information, which enables SerpApi to utilize proxy servers[5] near the location specified to provide improved search results. Its website includes a page with all the "SerpApi supported locations," which contains a "Locations" file.[6] That file includes hundreds of locations in the State of New York, including locations in all five boroughs of New York City, including the boroughs of Queens and Manhattan:

[5] A "proxy server" masks a user's IP address, effectively hiding a user's true location and identity from the website being visited. When a user connects to a website through a proxy server, the website will see the proxy server's IP address instead of the user's true IP address, allowing the user to appear to be browsing from near the proxy server's physical location.

[6] SerpApi, *Supported Locations API*, https://serpapi.com/locations-api (last accessed Feb. 6, 2026).

```
{
  "id": "585069f4ee19ad271e9cb3d3",
  "google_id": 9067608,
  "google_parent_id": 21167,
  "name": "Queens",
  "canonical_name": "Queens,New York,United States",
  "country_code": "US",
  "target_type": "Borough",
  "reach": 3640000,
  "gps": [
    -73.7948516,
    40.7282239
  ]
},
{
  "id": "585069f4ee19ad271e9cb3d4",
  "google_id": 9067609,
  "google_parent_id": 21167,
  "name": "Manhattan",
  "canonical_name": "Manhattan,New York,United States",
  "country_code": "US",
  "target_type": "Borough",
  "reach": 14400000,
  "gps": [
    -73.9712488,
    40.7830603
  ]
},
```


18. In addition, various SerpApi customers are headquartered in or have substantial operations in New York, including Perplexity, KPMG, Morgan Stanley, Shopify, and the United Nations.[7]

19. SerpApi is a "Remote-First Company," permitting employees to work from various locations both within the United States and around the world.[8] One or more SerpApi employees live and work in New York, and conduct activities from New York that give rise to Reddit's claims.[9]

---

[7] SerpApi, *They trust us*, https://serpapi.com/ (last accessed Feb. 6, 2026).

[8] SerpApi, *SerpApi Careers*, https://serpapi.com/careers (last accessed Feb. 6, 2026).

[9] SerpApi's website identifies a senior software developer, who has authored various blog posts on scraping Google data, as being located in New York. SerpApi, *Michael Moura*, https://serpapi.com/blog/author/michael_moura/ (last accessed Feb. 6, 2026); *see also* SerpApi (@serp_api), X (May 3, 2024, at 5:33 P.M.),

20. SerpApi unlawfully accessed and obtained Reddit data on computers located in New York and/or by using proxy servers located in New York, and/or by circumventing technological control measures on servers located in New York.

21. This Court's exercise of personal jurisdiction over SerpApi comports with both the New York long-arm statute and the Due Process Clause of the Fourteenth Amendment. SerpApi maintains various contacts within New York from which Reddit's claims arise, and the maintenance of this suit does not offend traditional notions of fair play and substantial justice.

Oxylabs Personal Jurisdiction

22. Defendant Oxylabs is subject to the jurisdiction of this Court pursuant to N.Y. C.P.L.R. § 302(a)(1), (2), and (3) and the Due Process Clause of the Constitution of the United States. It has purposefully directed its activities at New York and has purposefully availed itself of the benefits of doing business in New York.

23. Oxylabs transacts business in the State and District and supplies goods and services in the State and District. Oxylabs purports to operate "the world's largest ethical proxy network" and offers over 62,000 IP addresses located in New York.[10] Oxylabs also boasts to potential customers that its "[h]igh-quality New York proxies with residential IP addresses will help you scale up your data gathering projects and collect publicly available information efficiently":

---

https://x.com/serp_api/status/1786524554434597360 (last accessed Feb. 6, 2026). This employee can also be seen in this video posted on X (formerly Twitter) touting recent updates to SerpApi's scraping platforms, https://x.com/serp_api/status/1955290218069627300 (last accessed Feb. 6, 2026).

[10] Oxylabs, *Proxy locations in the US by largest IP count*, https://oxylabs.io/location-proxy/usa (last accessed Feb. 6, 2026); *see also* Oxylabs, *New York Proxy*, https://oxylabs.io/location-proxy/usa/new-york (last accessed Feb. 6, 2026).

Large pool of New York proxy IPs

High-quality New York proxies with residential IP addresses will help you scale up your data gathering projects and collect publicly available information efficiently. Oxylabs' pool of Residential Proxies is constantly expanding and helps various businesses access desired geo-restricted websites in the New York region. An average 99.2% success rate will boost your scraping projects without a hitch.

24. Oxylabs also advertises that its "New York proxy network allows you to select proxies from various locations, such as Bronx proxies, Brooklyn proxies, Buffalo proxies, Manhattan proxies, Queens proxies."[11]

25. At least two Oxylabs employees reside in New York.

26. Oxylabs unlawfully accessed and obtained Reddit data on computers located in New York and/or by using proxy servers located in New York, and/or by circumventing technological control measures on servers located in New York.

27. This Court's exercise of personal jurisdiction over Oxylabs comports with both the New York long-arm statute and the Due Process Clause of the Fourteenth Amendment. Oxylabs maintains various contacts within New York from which Reddit's claims arise, and the maintenance of this suit does not offend traditional notions of fair play and substantial justice.

AWMProxy Personal Jurisdiction

28. Defendant AWMProxy is subject to the jurisdiction of this Court pursuant to N.Y. C.P.L.R. § 302(a)(1), (2), and (3). AWMProxy has purposefully directed its activities at New York and has purposefully availed itself of the benefits of doing business in New York.

[11] Oxylabs, *Large pool of New York proxy IPs*, https://oxylabs.io/location-proxy/usa/new-york (last accessed Feb. 6, 2026).

29. AWMProxy transacts business in the State and District and supplies goods and services in the State and District. AWMProxy has previously claimed to have a proxy pool of over 350,000 IP addresses in locations spread throughout the world,[12] including in Brooklyn and New York City.[13] AWMProxy has recently re-established its proxy network, including IP addresses located in New York, to resume its data-scraping operations.

30. AWMProxy unlawfully accessed and obtained Reddit data on computers located in New York and/or by using proxy servers located in New York, and/or by circumventing technological control measures on servers located in New York.

31. This Court's exercise of personal jurisdiction over AWMProxy comports with both the New York long-arm statute and the Due Process Clause of the Fourteenth Amendment. AWMProxy maintains various contacts within New York from which Reddit's claims arise, and the maintenance of this suit does not offend traditional notions of fair play and substantial justice.

Perplexity Personal Jurisdiction

32. Defendant Perplexity is subject to the jurisdiction of this Court pursuant to N.Y. C.P.L.R. § 302(a)(1), (2), (3), and (4) and the Due Process Clause of the Constitution of the United States. It has purposefully directed its activities at New York and has purposefully availed itself of the benefits of doing business in New York.[14]

---

[12] Blackdown, *AWMProxy Review – Affordable Proxies for Micro Tasks*, https://www.blackdown.org/awm-proxy-review/ (last accessed Feb. 6, 2026).

[13] As of 2021, AWMProxy advertised proxies located in Brooklyn and New York City. AWMProxy, *Proxy from United States*, https://web.archive.org/web/20211020154610/https://awmproxy.net/modules/socks/country.php?city=US (last accessed Feb. 6, 2026).

[14] A court within this District recently concluded that personal jurisdiction over Perplexity was proper based on, among other reasons, Perplexity transacting business in New York, noting Perplexity's registration, office, and employees located in New York, its operation of a highly

33. Perplexity is registered as an "active" foreign business corporation in New York and has been registered since April 21, 2023. Perplexity's New York Department of State identification number is 6837808. Perplexity transacts business in the State and District, possesses and maintains an office from which various employees work within the State and District, and supplies goods or services in the State and District, through both the promotion and distribution of its products and the sale of subscription services to customers in this State and District.

34. Perplexity employs numerous individuals in this State and District. These New York-based employees include its Co-Founder and Chief Strategy Officer, business, legal, human resources, product, and engineering staff, as well as staff responsible for content, marketing, and maintaining and improving the customer experience of Perplexity's interactive website and mobile applications, through which it conducts significant business. In addition to Chief Strategy Officer, job titles and descriptions for some of these employees include Commercial Counsel, AI Business Fellow, and Product Marketing Lead.

35. These employees located in this State and District: (a) develop, implement, maintain, and promote the technology and/or business relationships through which Perplexity wrongfully accesses and obtains Reddit's data and then uses that data in the Perplexity platform; (b) support the operation of the web-based and mobile application technology that Perplexity uses to sell and otherwise provide answers and/or information to customers in New York, including by utilizing Reddit's content and services and content authored by Reddit's users; and (c) market Perplexity's products to customers and potential customers, including to business

---

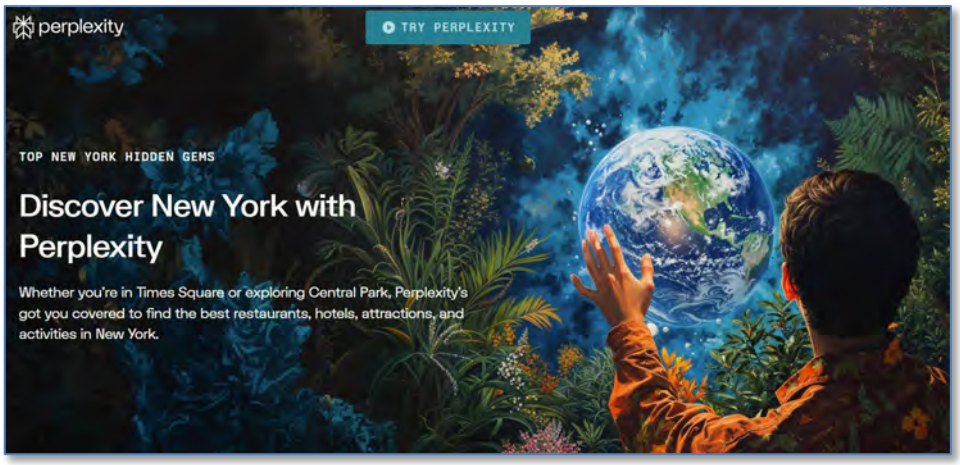
interactive website, and its advertising efforts targeted at New York. *Dow Jones & Co., Inc. v. Perplexity AI, Inc.*, No. 24 Civ. 7984, 2025 WL 2416401 (S.D.N.Y. Aug. 21, 2025).

customers, which results in Perplexity contracting for the sale of goods and services to customers in New York.

36. Perplexity is actively seeking to expand its New York presence by hiring additional employees. Perplexity's career website informs prospective employees that it has an office in New York City. As of October 2025, Perplexity's website listed openings for 27 different positions that were or could be based in its New York City office, including openings for artificial intelligence, machine learning, and various other engineering positions. The responsibilities and duties of these open jobs include (a) developing, implementing, maintaining, and promoting the technology and/or business relationships through which Perplexity wrongfully accesses and obtains Reddit's content and services, and content authored by Reddit's users, and then uses that data in the Perplexity platform; and (b) supporting the operation of the web-based and mobile application technology that Perplexity uses to sell and otherwise provide answers and/or information to customers in New York, including by utilizing Reddit's content and services, and content authored by Reddit's users.

37. Perplexity projects itself directly into New York and aggressively promotes and advertises its products and services in this State and District, including through its dynamic, heavily interactive website (https://www.perplexity.ai and associated subpages) and mobile applications. Perplexity's website targets customers in this State and District with promotional

material tailored to a New York audience, including a webpage inviting visitors to "Discover New York with Perplexity":[15]



38. Perplexity's marketing activities include promoting on its Instagram account using a massive billboard in Times Square from September 2024 which reads "Congratulations Perplexity on 250 million questions answered last month!"[16]

39. Perplexity derives substantial revenue from its website and mobile products through activities and customers in New York in at least three ways: (a) directing targeted advertising to New York users based on New York users' queries; (b) selling monthly recurring Perplexity Pro subscriptions to New York-based customers, and tailoring features to these customers based on their location in New York; and (c) providing its products and services to enterprise customers based in New York.

40. Perplexity's actions to unlawfully access, obtain, and utilize Reddit's data occurred, in part, on computers using proxy servers located in New York, and/or by

---

[15] Perplexity AI, *Discover New York with Perplexity*, https://www.perplexity.ai/encyclopedia/discovernewyork (last accessed Feb. 6, 2026).

[16] Perplexity.ai (@perplexity.ai), Instagram (Sept. 4, 2024) https://www.instagram.com/perplexity.ai/p/C_g2TonSHC5.

circumventing technological control measures and/or accessing data on servers located in New York.

41. This Court's exercise of personal jurisdiction over Perplexity comports with both the New York long-arm statute and the Due Process Clause of the Fourteenth Amendment. Perplexity maintains various contacts within New York from which Reddit's claims arise, and the maintenance of this suit does not offend traditional notions of fair play and substantial justice.

FACTUAL BACKGROUND

I. AI Companies Need Massive Amounts of Data and Target Reddit's Human Content

42. AI companies require staggering amounts of data to operate their products and services, including their large language models ("LLMs"). These companies atomize text and images into individual "tokens" and ingest the tokens into massive AI models defined by billions of parameters.

43. One of the most significant challenges AI companies face is how to procure additional and current data—especially content created by humans—to feed their voracious AI models. Recent trends, including increased competition amongst AI companies both within the United States and across national borders, have only compounded that hunger and driven AI companies to seek more sophisticated, creative, and—for some—underhanded means to obtain data.

44. Reddit is one of the internet's largest bodies of knowledge and information, with billions of posts and comments. As such, Reddit has emerged as an important source of human conversational data. Founded in 2006, Reddit is a community of communities and one of the largest online discussion platforms in the world, with over 100 million active unique users per

day.[17] Reddit comprises nearly two decades of human conversational data organized across interest-based, user-created communities (referred to as "subreddits") spanning subjects on virtually every topic imaginable, from current events to product reviews to sports and entertainment.

45. Reddit's vast corpus of human-generated content is widely seen as invaluable to AI companies. Reddit data is particularly well-suited to training LLMs because "Reddit data and information constantly grows and regenerates as users come and interact with their communities and each other in genuine and authentic ways." Reddit data also provides "real-time access to evolving and dynamic topics."[18] The largest technology companies in the world, as well as a growing list of new AI companies, require tremendous amounts of data to train and operate their AI products, with a high premium for data generated by actual humans. Reddit has become a "top-cited source"[19] of data for these companies. As one AI executive put it, Reddit's body of data is "like manna from heaven."[20]

---

[17] Reddit owns and operates all the "Services" associated with the Reddit community. These include "websites, mobile apps, widgets, APIs, emails, and other online products and services." *See* Reddit, *Reddit User Agreement* (Sept. 24, 2024), https://redditinc.com/policies/user-agreement-september-24-2024#hello-redditors-and-people-of-the-internet-this. Reddit also owns the rights to the compilation of Reddit data, as well as all other elements included in the "Services." *Id.*

[18] Reddit, Inc. Form S-1 Registration Statement (Feb. 22, 2024) at 8, https://www.sec.gov/Archives/edgar/data/1713445/000162828024006294/reddits-1q423.htm.

[19] D. Gallagher, *As AI Changes Internet Search, Reddit Lies in a Sweet Spot*, Wall Street Journal (Aug. 6, 2025), https://www.wsj.com/tech/ai/reddit-rddt-stock-ai-search-2dcc69a4?st=5sRNSY.

[20] S. Needleman, *How Years of Reddit Posts Have Made the Company an AI Darling*, Wall Street Journal (Dec. 10, 2024), https://www.wsj.com/tech/ai/reddit-ai-partnerships-investors-59eef0fe?st=tNB19v.

II. Reddit's Agreements with its Users and its Licensing Partners

46. Reddit's User Agreement defines "Content" as "information, text, links, graphics, photos, videos, audio, streams, software, tools, or other materials ('Content'), including Content created with or submitted to the Services by you or through your Account ('Your Content')."[21] Under the User Agreement, Reddit users retain any ownership rights they have in their content. They also grant Reddit a worldwide, perpetual, irrevocable, non-exclusive, and sublicensable license to use, copy, modify, adapt, prepare derivative works of, distribute, store, perform, and display the content in any media formats and channels now known or later developed anywhere in the world. The license to Reddit includes the right for Reddit to make the content available for syndication, broadcast, distribution, or publication by other companies, organizations, or individuals who partner with Reddit. The license also includes the right to use the content to train AI and machine learning models, as further described in the Reddit Public Content Policy.

47. The content posted to Reddit includes not only Reddit user content, but also tens of thousands of posts, comments, and other original works authored by Reddit itself. Reddit owns the copyrights in these works, as well as the copyrights in its compilations and arrangements of user-generated content. Both Reddit's and Google's technological protection measures protect access to these copyrighted works and arrangements.[22]

---

[21] *See* Reddit, *Reddit User Agreement*, Preamble & §§ 3, 5 (Sept. 24, 2024), https://redditinc.com/policies/user-agreement-september-24-2024#hello-redditors-and-people-of-the-internet-this.

[22] *See, e.g.*, *Why Is Subreddit? Or, a Brief History of the Subreddit*, r/reddit, Reddit.com (Feb. 15, 2022) (an Admin-authored, 700-word history of the origin of the subreddit), https://www.reddit.com/r/reddit/comments/sta5wl/why_is_subreddit_or_a_brief_history_of_the/. Notably, SerpApi's tool, when used to search "History of the Subreddit," returns a snippet taken verbatim from Reddit's own post: "Back when Reddit launched in 2005, there were no subreddits; the site was just one long list of various links." Likewise, when Perplexity is prompted to provide a brief history of the subreddit, it cites Reddit's post as a source and

48. Although users retain any ownership rights they may have in their own content, Reddit's User Agreement includes certain restrictions. Under the terms of the User Agreement, the user cannot license Reddit user content to anyone else. In addition, users cannot sell, transfer, assign, distribute, host, or otherwise commercially exploit Reddit user content, including their own. Per the Public Content Policy, users can "use Reddit content for non-commercial uses, such as learning and community" but Reddit requires them to "talk to us if you have commercial purposes in mind."[23]

49. The User Agreement defines Reddit's "Services," as "websites, mobile apps, widgets, APIs, emails, and other online products and services." Section 11 of the User Agreement, titled "Intellectual Property," notes that "the Services are owned and operated by Reddit" and that "the visual interfaces, graphics, design, compilation, information, data, computer code, products, services, trademarks, and all other elements of the Services ('Materials') provided by Reddit are protected by intellectual property and other laws. All Materials included in the Services are the property of Reddit or its third-party licensors." The User Agreement limits how users can access the data via Reddit's Services. Users can access and use the Services, but they cannot "[a]ccess, search, or collect data from the Services by any means (automated or otherwise) except as permitted in these Terms or in a separate agreement with Reddit. . . ."

50. The User Agreement grants Reddit many rights as to how to use and protect Reddit user content. For example, the Reddit User Agreement incorporates the Privacy Policy

---

reproduces the same point almost word for word: "When Reddit launched in 2005, there were no subreddits at all, just one shared list of links."

[23] Reddit, *Public Content Policy*, https://support.reddithelp.com/hc/en-us/articles/26410290525844-Public-Content-Policy (last accessed Feb. 6, 2026).

and Public Content Policy by reference and both explicitly authorize Reddit to take measures to protect its users' information and content. The Privacy Policy authorizes Reddit to take "measures to help protect information about you from loss, theft, misuse and unauthorized access, disclosure, alteration, and destruction." This provision authorizes Reddit to require others to protect Reddit user content from misuse or unauthorized access.

51. In addition, the Public Content Policy states that Reddit only enters into licensing arrangements with licensing partners "that allow us to put into place meaningful protections." Those protections include, but are not limited to, the ability to know who is accessing Reddit public content and why; placing contractual restrictions on prohibited use of public content to protect user rights; and ensuring that data licensees honor public content deletions by Redditors and Reddit.[24] Reddit's Public Content Policy makes clear that "Protecting Redditors in our data licensing arrangements is essential."

52. Reddit has entered into partnership arrangements with a variety of companies to license Reddit content. All such agreements include prohibitions on redistribution of the content through selling, syndicating, or transferring to a third party. Partners are obligated to not make the data available to unauthorized third parties. Most of these arrangements forbid AI use cases. Reddit has entered into such partnership agreements with a select list of AI companies that agree to abide by Reddit's policies and restrictions to protect the intellectual property, privacy, and other rights of both Reddit and Reddit users, as set forth in its Public Content Policy.[25] For

---

[24] Reddit, *Public Content Policy*, https://support.reddithelp.com/hc/en-us/articles/26410290525844-Public-Content-Policy (last accessed Feb. 6, 2026).

[25] Reddit's Public Content Policy describes how Reddit protects and licenses public user content. This is to help Reddit's users "understand what happens when [they] create, submit, and make content publicly available on the Reddit platform, and how and why Reddit protects and, where appropriate, licenses public content and related data." Reddit, *Public Content Policy*, https://support.reddithelp.com/hc/en-us/articles/26410290525844-Public-Content-Policy (last

example, on February 22, 2024, Reddit and Google announced a partnership that enabled programmatic access by Google to Reddit content for use in Google's products and services.[26]

III. Reddit and Its Partners Utilize Technological Measures to Control Access and Prevent Scraping

53. Reddit and Google implement technological control measures to effectively control access and prevent scraping[27] of content posted to Reddit.

54. Reddit has and maintains automated anti-scraping systems and teams dedicated to detecting and protecting against unauthorized access to its products and services. Reddit has made significant investments in various anti-scraping tooling and systems that detect and prevent unauthorized access by scrapers and bots, as well as invested in multiple engineering teams and security professionals dedicated to improving such tooling and systems. Where Reddit's tooling and systems detect unauthorized automated access, its systems issue challenges that must be bypassed before providing any content in response to that network traffic. Some of the measures Reddit employs to prevent unauthorized entities from accessing or scraping its data include industry-standard automation-prevention techniques, such as registered user-identification limits, IP-rate limits, captcha bot protection, and anomaly-detection tools. In other words, as Reddit's current Robots Exclusion Protocol file ("robots.txt") says, "Reddit believes in an open internet, but not the misuse of public content." Reddit's robots.txt file provides clear notice to all scrapers that it prohibits all unapproved web crawlers from accessing any part of Reddit's website.

---

accessed Feb. 6, 2026). Reddit published this policy in response to "more and more entities using unauthorized access (for example, by scraping or using data brokers) or misusing authorized access to collect public data in bulk, especially with the rise of use cases like generative AI." *Id.*

[26] *See* Reddit, *Expanding our Partnership with Google* (Feb. 22, 2024), https://redditinc.com/news/reddit-and-google-expand-partnership (last accessed Feb. 6, 2026).

[27] Scraping refers to an automated process for copying content from a website using web crawlers or "bots."

Reddit's User Agreement, which authorized humans agree to when accessing Reddit, also clearly prohibits "scraping the Services without Reddit's prior written consent."[28] Reddit only allows individuals to access its services and content if they agree to Reddit's User Agreement. Reddit prohibits bulk automated access except by separate agreement through defined application programming interfaces ("APIs").

55. Google likewise has anti-scraping systems and teams dedicated to preventing unauthorized access to its products and services. Google operates a search engine which, when a human user submits a query, responds with SERPs that include Reddit data when the search results include Reddit results. Google does not permit, and indeed prohibits, unauthorized automated access to its SERPs. Google's terms of service prohibit "using automated means to access content from any of our services in violation of the machine-readable instructions on our web pages (for example, robots.txt files that disallow crawling, training, or other activities)."

56. Google utilizes a technological access control system called "SearchGuard," which is designed to prevent automated systems from accessing and obtaining wholesale search results and indexed data while allowing individual users—i.e., humans—access to Google's search results, including results that feature Reddit data.[29] Google developed SearchGuard at considerable expense, and it is the product of tens of thousands of person hours and millions of

[28] Reddit, *Reddit User Agreement*, § 7 (Sept. 24, 2024), https://redditinc.com/policies/user-agreement-september-24-2024#hello-redditors-and-people-of-the-internet-this.

[29] Reddit obtained this understanding by subpoenaing Google for information regarding its technological control measures and the identity of any entities circumventing Google's technological control systems to obtain Reddit data. Reddit obtained additional information about Google's technological control measures from the Complaint that Google filed against SerpApi LLC on December 19, 2025. *See Google LLC v. SerpApi, LLC*, No. 25-10826 (N.D. Cal. Dec. 19, 2025), Dkt. 1.

dollars of investment. Its purpose is to prevent unauthorized parties from automatically accessing Google's search results without degrading the Google search experience.

57. SearchGuard prevents unauthorized access to Google's search data by imposing a barrier challenge that cannot be solved in the ordinary course by automated systems unless they take affirmative actions to circumvent the SearchGuard system.

58. Specifically, SearchGuard works by examining each incoming search query to determine whether it is coming from a recognized or unrecognized source. If the source is unrecognized, it then sends a JavaScript "challenge" to validate that the query to Google comes from a real user as opposed to automated software. Google's computer system transmits JavaScript code that calls upon the user's browser to send Google a "solve" for the challenge, that is, to send Google specific information regarding the browser and user generating the request. That process is not confined to automated systems but is presented to all browsers submitting queries, including browsers operated by human users, with access dependent on the browser's ability to provide a valid solution. For human users, the solution is relatively straightforward; their browsers run the JavaScript code and send back the required information seamlessly, without disrupting the user experience. But automated systems that submit automated queries at a massive scale typically cannot access the desired information because they usually cannot solve the SearchGuard challenge.

59. In addition to the above "challenge," Google SearchGuard uses other techniques to control access to Google search results and the content contained within them, including CAPTCHAs.[30] Users seeking access to Google search results, including automated systems and

[30] Like a traditional authentication gate that conditions access on the provision of required information, a CAPTCHA is a challenge-response test designed to distinguish human users from automated programs. The test requires completion of a task, such as recognizing distorted

human users, will encounter one or more aspects of SearchGuard and must overcome any control tests SearchGuard imposes *before* SearchGuard permits access. As a result, SearchGuard effectively controls access to copyrighted works because, in the ordinary course of its operation, SearchGuard requires the application of information to gain access to Google search results and the copyrighted content they contain.

60. Reddit has no agreement or arrangement with any of the Defendants that would permit them to circumvent these technological control measures. Reddit has implemented an API that allows a third party to access Reddit data in bulk so long as the third party agrees to specific restrictions. But access to the Data API requires prior Reddit authorization, user authentication, a unique and descriptive user agent, request rate limits, and limits on data use and retention. Reddit grants Data API users a non-exclusive, non-transferable, non-sublicensable, and revocable license to access and use the Data API in accordance with the terms of use. Unless otherwise allowed, Reddit limits the use of such bulk data through the Data API to "develop, deploy, distribute, and run in" the licensee's app. Reddit also expressly prohibits using content for any other unapproved purpose, such as for training a machine-learning, LLM, or AI model, unless or until the third party has an agreement to do so. Reddit routinely blocks scrapers that attempt to access and scrape data unless they agree to the privacy and data restrictions in Reddit's policies, including their agreement to not use crawled data for AI training. If an entity is interested in using the Data API for a commercial purpose, then it must enter a separate agreement with Reddit. None of the Defendants have entered into any agreement with Reddit with respect to the Data API.

---

characters or matching images, that most humans can perform but automated systems typically cannot, and access is granted only upon successful completion.

61. Google likewise has no agreement or arrangement with any of the Defendants that would permit any of them to circumvent any of these technological control measures. Google provides APIs for trusted partners for accessing search results in bulk but none of the Defendants have entered into any such agreement or arrangement with Google.

62. To the contrary, on December 19, 2025, Google sued Defendant SerpApi alleging circumvention of Google's technological protection measures, including SearchGuard. According to the Complaint, SearchGuard initially blocked SerpApi's access to Google search results and the copyrighted content they contain, but SerpApi developed and deployed means to bypass the technological protection measures. Google alleges that SerpApi has since circumvented SearchGuard on billions of occasions to access Google search results and the copyrighted content they contain, without authorization from the relevant rights holders, including Reddit.

IV. Data Scrapers Unlawfully Access and Obtain Reddit Content to Sell to AI Companies

63. To avoid Reddit's restrictions, Defendants SerpApi, Oxylabs, AWMProxy, and Perplexity have engaged in a scheme to obtain Reddit data by circumventing the above technological measures that control access to Reddit data.[31] Defendant data scrapers SerpApi, Oxylabs, and AWMProxy, in an end-run around Reddit's own technological control measures, are instead circumventing Google's technological control measures to access Reddit content in

[31] Since filing its Complaint in this matter, Reddit identified evidence suggesting that one of the Defendants may also be scraping Reddit data directly from its servers. Specifically, during January 2026, Reddit detected massive increases in the amount of data that certain residential IP addresses were requesting, far in excess of what humans request when accessing Reddit. The responsible entity was somehow able to circumvent Reddit's technological protection measures until Reddit blocked those IP addresses. These tactics are consistent with the Defendants' known behavior—probing Reddit's defenses, exploiting vulnerabilities until Reddit patches them, and then probing again in an endless game of "whack-a-mole." Reddit's investigation is ongoing.

Google SERPs. Perplexity has similarly used at least one of its Co-Defendants' tools to circumvent or caused others to use those tools to circumvent those same technological control measures on its behalf.

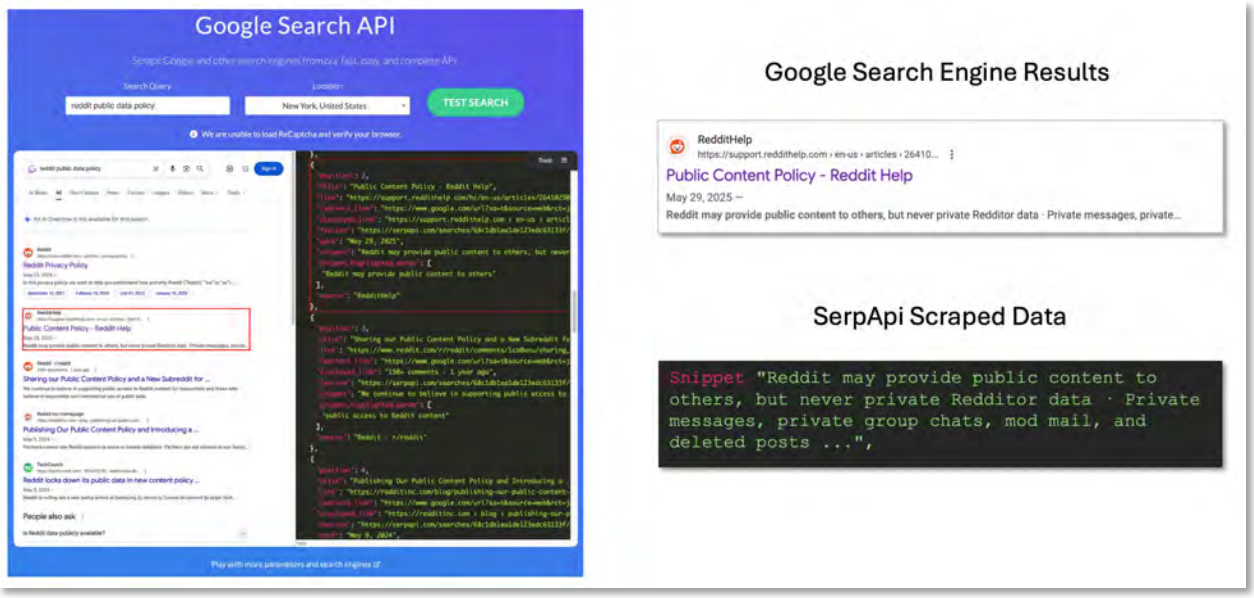
64. As mentioned above, Google controls access to the data that appears in the SERPs, including Reddit data. Google has implemented various measures and security protocols to prohibit access and scraping by unauthorized, automated incursions, including those by SerpApi, Oxylabs, AWMProxy, and Perplexity.[32] The Defendant data scrapers SerpApi, Oxylabs, and AWMProxy, sidestepping Reddit's own technological control measures, have created tools that circumvent Google's security protocols to access, through unauthorized and automated processes, and collect Reddit data from Google SERPs.

65. The Defendant data scrapers allow their customers to establish the parameters governing circumvention activity, including by specifying operational inputs that determine when, how, and at what scale circumvention will occur, and then use the tools themselves. For example, Defendant SerpApi provides customers with a "playground" through which customers can set parameters for circumvention activity, including geographic location, rendering layouts, language settings, advanced filtering, and other operational inputs. Perplexity then used one or more of the Defendant data scrapers' tools to circumvent technological control measures and gain access, through unauthorized and automated processes, to Reddit data.

66. Defendants either use or offer products that make an end-run around Reddit's and Google's restrictions. For example, SerpApi advertises that its product allows users to access and

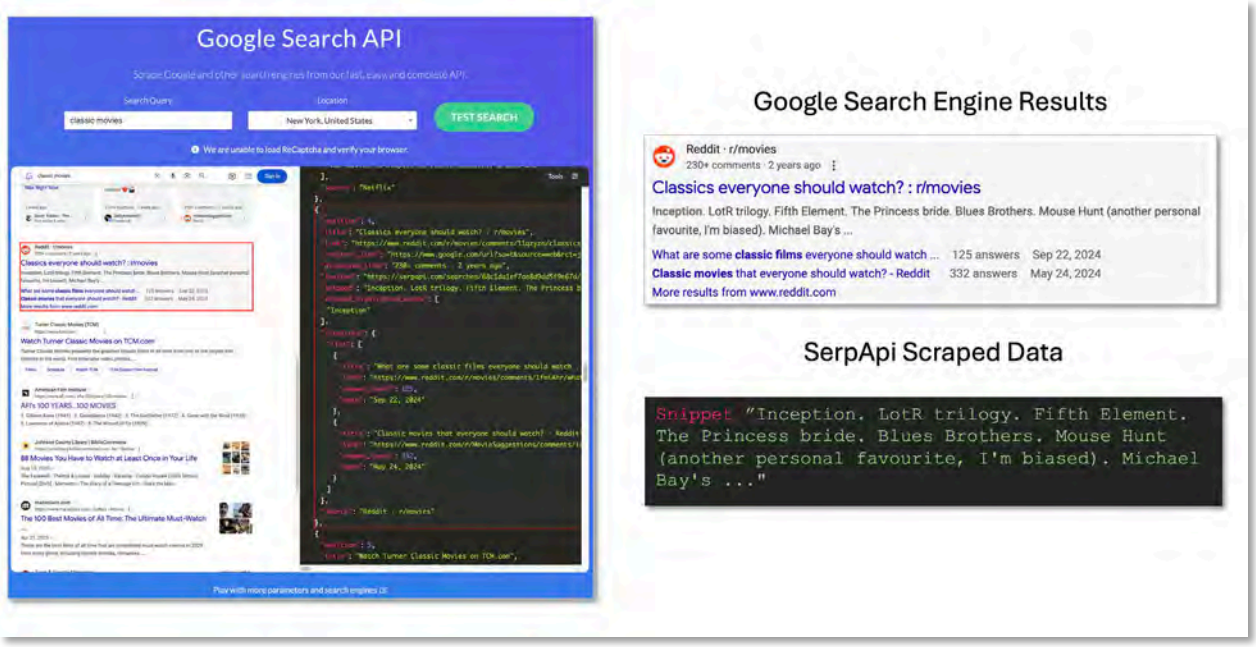
[32] In fact, "Google executives have privately derided SerpApi" and have "tried various techniques to make it harder for the firm to scrape high-quality information through its web crawler[.]" A. Efrati, *et al.*, *OpenAI is Challenging Google—While Using Its Search Data*, The Information (Aug. 21, 2025), https://www.theinformation.com/articles/openai-challenging-google-using-search-data?rc=zsmcbw (last accessed Feb. 6, 2026).

scrape the data that comes back from Google SERPs. Using SerpApi's interface,[33] a search for "Reddit public data policy" generates a list of snippets directly from Reddit, which SerpApi scraped from Google's SERPs:

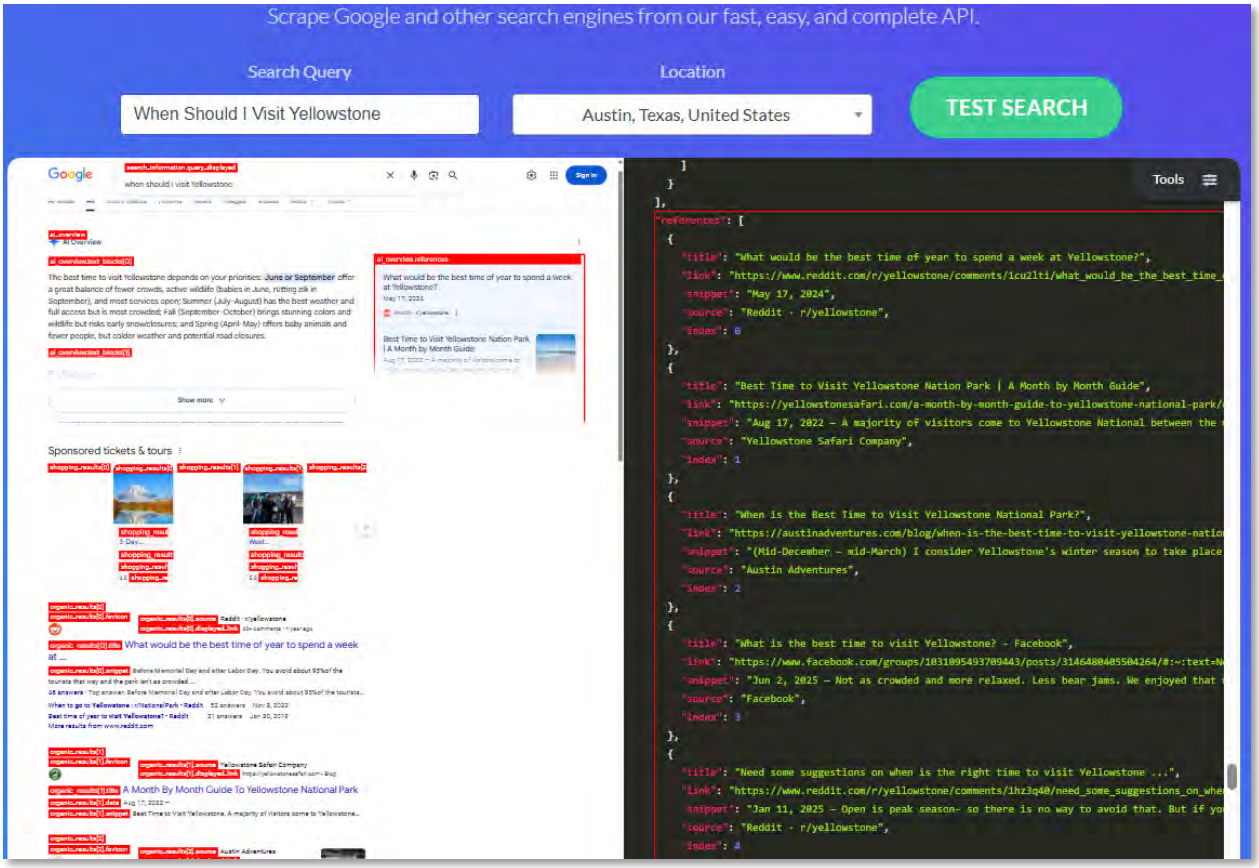


67. SerpApi's tool also accesses and scrapes Reddit data from SERPs even if the query makes no mention of Reddit. For example, a query on SerpApi's interface for "classic movies" also retrieves snippets of Reddit data from the Google SERPs:

[33] SerpApi, *Google Search API*, https://serpapi.com (last accessed Feb. 6, 2026).



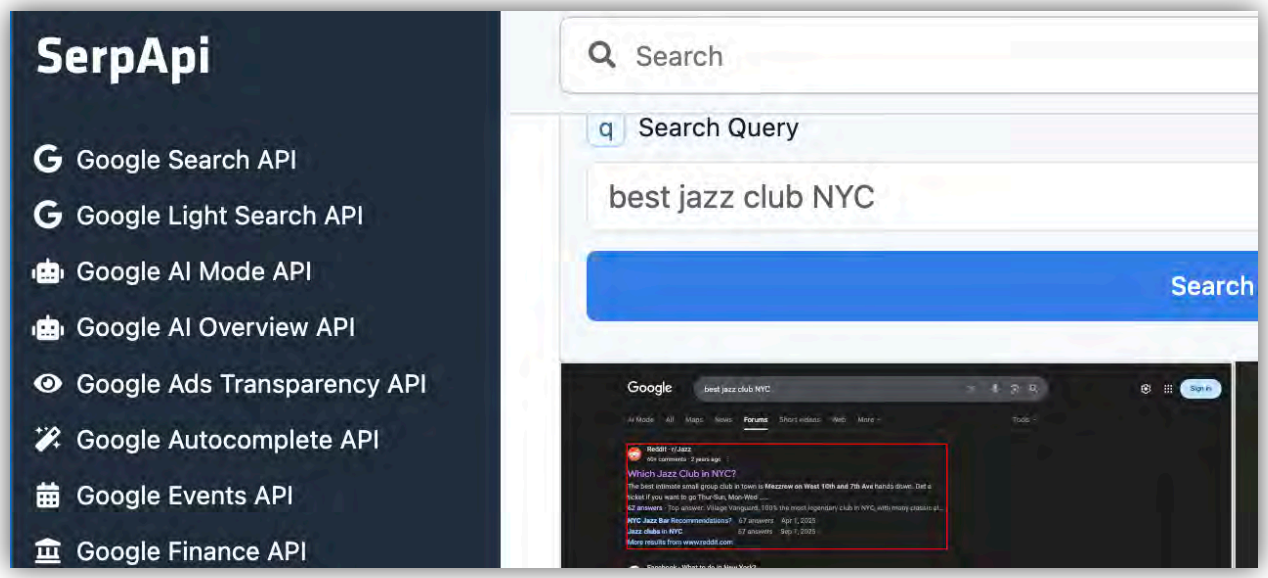
68. SerpApi's tool also extracts Reddit data from the AI overview content at the top of Google SERPs. As one example, entering the search query "When Should I Visit Yellowstone" into SerpApi's tool shows that it scrapes Reddit data from the "AI Overview References," including the post "What would be the best time of year to spend a week at Yellowstone?" as well as references that are not even visible on the page:



69. SerpApi's tool accesses Reddit's arrangements, not merely the raw, user-generated content or Reddit content itself. When Reddit content appears in Google SERPs, the snippets shown often correspond to comments that Reddit has made prominent within a post through its unique arrangement of content. Absent Reddit's arrangements, the same raw comments would be substantially less valuable to a user of the service.

70. When the SerpApi tool is used to circumvent Google's SearchGuard, it accesses not merely individual Reddit comments, but also Reddit's selection and arrangement of which comments to highlight. On January 24, 2026, when the SerpApi tool was used to access the Reddit content through Google, it accessed the comment that Reddit made most prominent rather than accessing any arbitrary comment within the post. In doing so, SerpApi accesses not only the content itself, but also Reddit's arrangement of Reddit data. In the below example, the SerpApi

tool accessed the comment that Reddit arranged to display most prominently within the "*Which Jazz Club in NYC?*" post.



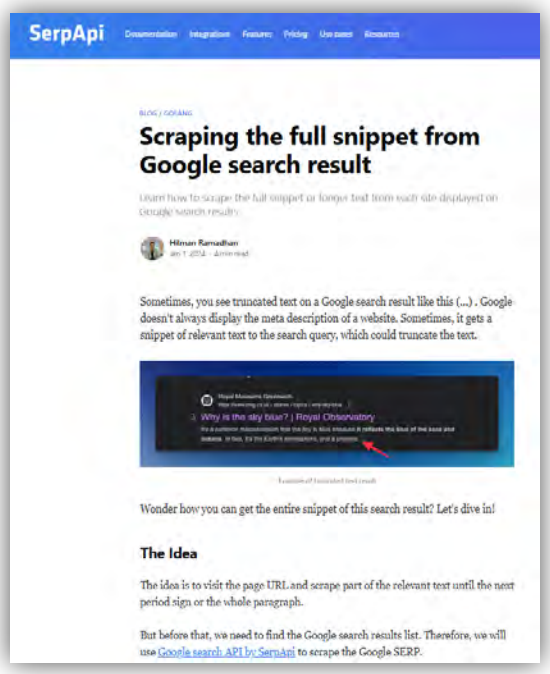
71. In addition, as the below example shows, SerpApi does not access Reddit data like an ordinary, human user conducting a Google search would. Instead, SerpApi customers access this content from Google SERPs in a JSON format. JSON is a structured, machine-readable format that organizes information into labeled data fields, rather than presenting content as a webpage intended for human viewing. Through this format, SerpApi customers access Reddit content as structured data that identifies particular elements of the content.

72. The JSON format accessed by SerpApi customers includes Reddit content arrangements. The JSON content accessed by SerpApi may include as a standalone field the

comment that Reddit's arrangement has decided to display most prominently within a post. In the above example, the "snippet" field includes the comment that Reddit's arrangement made most prominent within the "*Which Jazz Club in NYC?*" post. That is no accident. Reddit's arrangement is valuable for machine training, automated analysis, and/or Perplexity's "answer engine" because it allows systems to ingest not only the raw text of user comments, but also to benefit from Reddit's unique arrangement of content, which indicates which types of responses may be more useful or informative in response to particular questions.

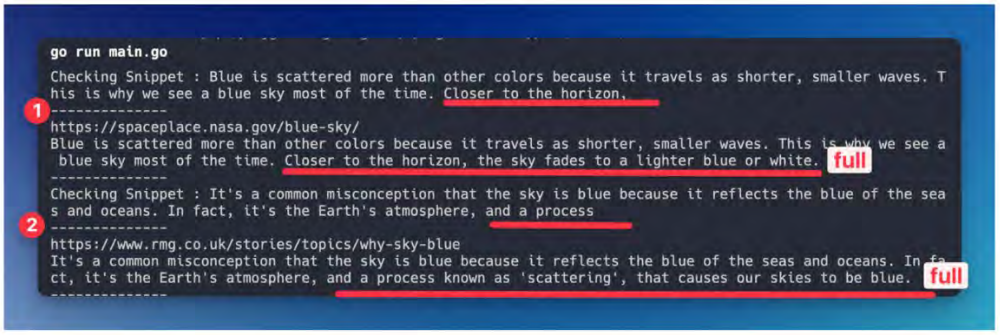
73. SerpApi also explains that the text it retrieves from Google SERPs is not limited to truncated snippets. As SerpApi describes on its website, users may "see truncated text on a Google search result" because "Google doesn't always display the meta description of a website."[34] SerpApi nevertheless advertises on its website how to use its service to collect a full snippet that would include longer portions of text:

[34] SerpApi, *Scraping the full snippet from Google search result*, https://serpapi.com/blog/scraping-the-full-snippet-from-google-search-result/ (last accessed Feb. 6, 2026).



74. For example, SerpApi presents the following illustration of a truncated text result.

75. In that same page, SerpApi explains how its customers can extract additional text. According to SerpApi, users can "scrape part of the relevant text until the next period sign or the whole paragraph." SerpApi then provides the following example demonstrating how its services can be used to obtain "a full snippet from Google."



Result of a full snippet from Google

76. As the above example illustrates, in the instances that Google displays a truncated snippet, SerpApi enables its customers to obtain the full, untruncated text. SerpApi also advertises that its tool is effective for machine learning and AI models because users can use its tool to "scrape text" and "extract the text behind each webpage."[35]

77. Google's terms of service prohibit "using automated means to access content from any of our services in violation of the machine-readable instructions on our web pages (for example, robots.txt files that disallow crawling, training, or other activities)."[36] The scraping examples noted above of Google's SERPs occur through Defendants bypassing these access restrictions.

78. Despite Google's technological control measures, SerpApi, Oxylabs, and AWMProxy all have products or services that circumvent these measures – indeed, both Oxylabs and SerpApi boast about their ability to bypass these protections. Perplexity is also circumventing these measures by using (at least) SerpApi's tool.

---

[35] SerpApi, *Data for Machine Learning and AI Models*, https://serpapi.com/use-cases/machine-learning-and-artificial-intelligence (last accessed Feb. 6, 2026).

[36] Google, *Terms of Service* (May 22, 2024), https://policies.google.com/terms?hl=en-US. One observer has specifically suggested that SerpApi's actions "might run afoul of Google's terms of service." A. Efrati, *et al.*, *supra*.

79. For example, Oxylabs is explicit that its scraping service is meant to circumvent Google's technological measures controlling access to SERPs, featuring a page on its website called "How to Scrape Google Search Results" that specifically notes that its service "bypass[es]" these "restrictions":

> [I]f you need to scrape Google search pages automatically and on a large scale, you may run into some difficulties. Using a reliable free proxy list or any other proxy option can help you bypass restrictions and avoid getting blocked while scraping.[37]

Oxylabs then notes that "Oxylabs' Google Search API is constructed to overcome technical challenges and ensure reliable access to publicly-available web data."[38]

80. Oxylabs' website also includes tutorials from its employees on how to use their services to "bypass" sophisticated anti-bot systems:[39]

---

[37] Ex. C, D. Vėsa, *How to Scrape Google Search Results: Python Tutorial*, Oxylabs (Jan. 28, 2025).

[38] *Id.*

[39] Oxylabs, *Learn From Experts and Master Your Web Scraping Skills*, https://experts.oxylabs.io/ (last accessed Feb. 6, 2026).

81. Similarly, SerpApi describes on its website how to scrape Google search results. SerpApi tells its users that with its services, they "don't need to care" about technological control measures that block scrapers, including "HTTP requests, parsing HTML files to JSON (parser), or captcha, IP address, bots detection, maintaining user-agent, HTML headers, [or] being blocked by Google." [40] In order to bypass these technical measures and "*to avoid being detected and blocked by Google*, SerpApi uses a proxy and the latest technologies to mimic human behavior."[41] SerpApi provides its users with tips to reduce the chance of being blocked while web scraping, such as by sending "fake user-agent string[s]," shifting IP addresses to avoid multiple requests from the same address, and using proxies "to make traffic look like regular

[40] SerpApi, *How to Scrape Google Search Results (SERPs) – 2026 Guide*, https://serpapi.com/blog/how-to-scrape-google-search-results-serps-2023-guide/ (last accessed Feb. 6, 2026).

[41] *Id.* (emphasis added).

user traffic" and thereby "impersonate" user traffic.[42] SerpApi markets its tool as providing a way to scrape Google web searches "at scale" for use in training LLM or other AI models.[43] SerpApi says that it can run more than 100,000 searches per hour.[44] SerpApi has also promoted its "Ludicrous Speed" and "Ludicrous Speed Max" features as effective ways to circumvent Google's technological control measures.[45] Ludicrous Speed Max "use[s] 4x the server resources to automatically create numerous parallel requests compared to our already aggressive Ludicrous Speed."[46] Ludicrous Speed Max operates by "reject[ing]" what SerpApi characterizes as "bad HTMLs, CAPTCHA and error pages, and other abnormalities." In other words, SerpApi uses a server-swarm to hide from, avoid, or simply overwhelm by brute force effective measures Google has put in place to ward off automated access to search engine results. SerpApi's website depicts how its server army circumvents Google technological controls on behalf of its "Client":

---

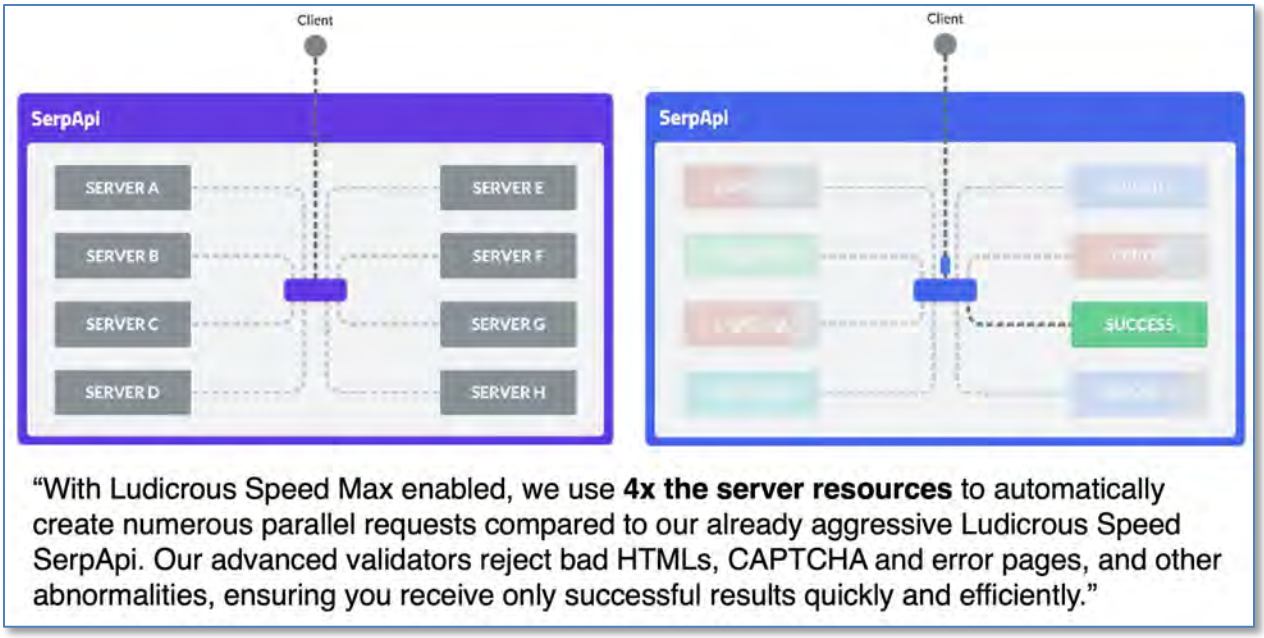
[42] D. Zub, *Reducing the chance of being blocked while web scraping*, SerpApi (Sept. 16, 2021), https://serpapi.com/blog/how-to-reduce-chance-of-being-blocked-while-web/.

[43] SerpApi, *Data for Machine Learning and AI Models*, https://serpapi.com/use-cases/machine-learning-and-artificial-intelligence (last accessed Feb. 6, 2026).

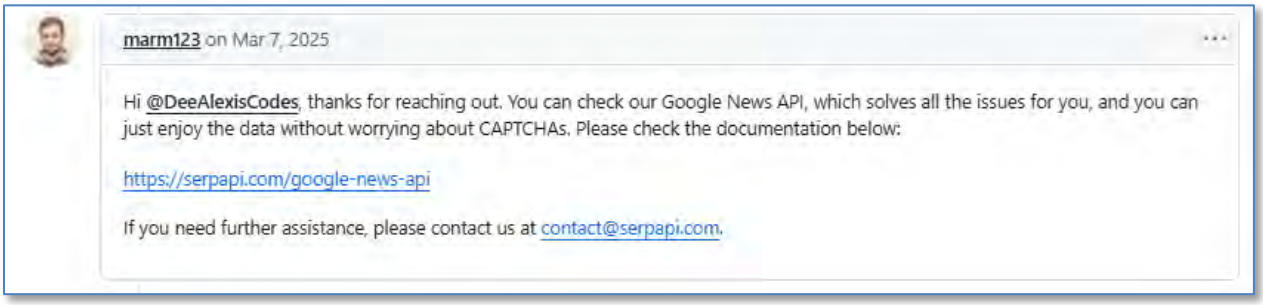
[44] SerpApi, *FAQ*, https://serpapi.com/faq (last accessed Feb. 6, 2026).

[45] R. Schafer, *Google Now Requires JavaScript*, SerpApi (Jan. 17, 2025), https://serpapi.com/blog/google-now-requires-javascript/.

[46] SerpApi, *Ludicrous Speed Max*, https://serpapi.com/ludicrous-speed-max (last accessed Feb. 6, 2026).



82. SerpApi's employees have also assured users that its tools are able to bypass Google technological protections. When one potential user posted a request to SerpApi's public road map that "I need to bypass Google CAPTCHA," a SerpApi employee responded the very same day, reassuring the user that the requested functionality was already present: "You can check our Google News API, which solves all the issues for you, and you can just enjoy the data without worrying about CAPTCHAs":[47]



[47] GitHub, *Search API – I need to bypass Google CAPTCHA*, https://github.com/serpapi/public-roadmap/issues/2466 (last accessed Feb. 6, 2026).

SerpApi's CEO and Founder, Julien Khaleghy, has described SerpApi's circumvention methods as including a tool for "creating fake browsers using a multitude of IP addresses that Google sees as normal users."[48]

83. SerpApi advertises on its website that AI companies can use SerpApi's tools to help them "power chatbots with fresh search results for more reliable answers" and "[f]eed search data into AI pipelines to minimize hallucinations:"[49]

> **Real-time SERP Data**
>
> Google and other search engines' data can ground Large Language Models (LLMs), reduce hallucinations, and power more accurate AI assistants.
>
> - AI Assistants: Power chatbots with fresh search results for more reliable answers.
> - Retrieval-Augmented Generation (RAG): Feed search data into AI pipelines to minimize hallucinations.
> - Knowledge & Research Tools: Provide users with fact-checked sources directly inside AI-powered platforms.
> - Automated Agents: Enable autonomous agents to query the web safely and at scale.

84. In addition, SerpApi offers its users a "Zero Trace Mode" so that SerpApi will not even store search parameters, search files, or search data on its servers. In its own words, "once your search is done, it's as if it were never here."[50]

---

[48] S. Palazzolo, *Meet the Google-Scraping Startup Used by ChatGPT, Cursor and Perplexity*, The Information (Aug. 28, 2025), https://www.theinformation.com/articles/meet-google-scraping-startup-used-chatgpt-cursor-perplexity?rc=zsmcbw.

[49] SerpApi, *AI Search Engine API*, https://serpapi.com/use-cases/ai-search-engine-api (last accessed Feb. 6, 2026).

[50] SerpApi, *Zero Trace Mode*, https://serpapi.com/zero-trace-mode (last accessed Feb. 6, 2026).

85. According to information obtained through a subpoena issued to Google, the three Defendant scraping companies have been circumventing or bypassing Google's technological control measures to access and scrape Reddit data from Google search results at an industrial scale:

Total Google SERPs with Reddit Data Automatedly Accessed Without Authorization in a Two-Week Period		
	July 1-6, 2025	**July 7-13, 2025**
SerpApi	784,006,306	1,057,744,585
Oxylabs	333,022,219	448,235,070
AWMProxy	217,954,929	264,232,011

86. The three companies are only able to obtain Reddit data from Google SERPs because they circumvent the technological control measures Google has in place to prevent such unauthorized automated access and scraping. As discussed previously, Google has technological control measures to validate that a query comes from a real user and not automated software. SerpApi, Oxylabs, and AWMProxy (and their customers and licensees) employ various measures to avoid, bypass, disable, or impair these measures, including by forging credentials or manipulating Google's control measures into incorrectly identifying their scraping services as individual users, not automated scrapers, and/or evading or otherwise circumventing Google's control measures by submitting repeated queries through large volumes of proxy servers.

V. AI Companies Like Perplexity Unlawfully Circumvent Technological Protections to Access and Obtain Reddit Content Despite Promising Not to Do So

87. Perplexity is an AI company that purports to be "the world's first answer engine" by gathering insights from "top tier sources." In reality, Perplexity does nothing groundbreaking. Its answer engine simply uses a different company's LLM to parse through a massive number of Google search results to see if it can answer a user's question based on those results. But

Perplexity can only run its "answer engine" by wrongfully accessing and scraping Google's own search results (and Reddit data) from Google's own search engine.

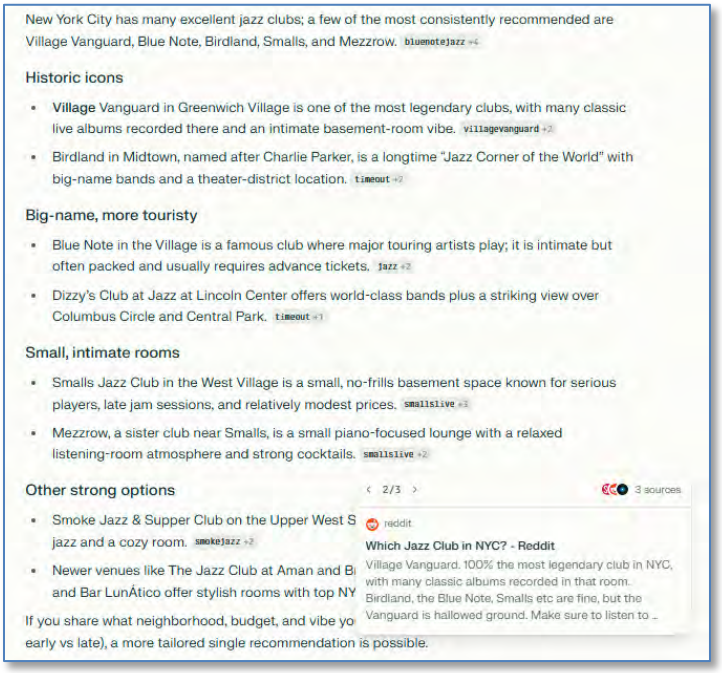
88. Perplexity admits that one of its "top tier sources" is Reddit, noting in an August 2025 blog post that "Reddit has emerged as the most cited domain across AI models globally" based on an analysis of ChatGPT, Google AI Overviews, and Perplexity citations.[51]

89. Perplexity created its "answer engine" by utilizing content from original sources such as Reddit and placing that data into a massive "retrieval-augmented generation" ("RAG") database, which Perplexity then combines with another company's LLM to repackage original source data into "answers" to user queries. In other words, Perplexity's business model is effectively to take Reddit data from Google search results, feed them into a third party's LLM, and call it a new product. While that business model has somehow translated into a $20 billion valuation, it has not resulted in a willingness to pay for what others (including Google) have.

90. As one example, when SerpApi scrapes the Google SERPs for "When Should I Visit Yellowstone" and extracts the Reddit post "What would be the best time of year to spend at Yellowstone," *that same post* then shows up as a source for Perplexity's "answer" to the same question:

[51] Perplexity, *Reddit overtakes Wikipedia as top source for AI model citations* (Aug. 5, 2025), https://www.perplexity.ai/discover/you/reddit-overtakes-wikipedia-as-UUKumG5zRRiU8oZHcu_gDA (last accessed Feb. 6, 2026).

91. As another example, the Reddit post ("*Which Jazz Club in NYC?*") scraped by SerpApi shows up as a source for one of Perplexity's answers for where to see jazz in New York City:



92. Reddit expends substantial monetary and labor resources to design, build, and maintain the systems that structure content on its platform, including dedicated engineering teams and significant technical infrastructure required to continuously update that structure based on new posts, comments, voting activity, and other user inputs. That structured organization of content is a valuable commercial advantage belonging exclusively to Reddit, because it reflects which responses are most relevant, useful, or salient in relation to a given question or topic, information that is not apparent from raw content alone. Accordingly, Reddit's structuring provides Defendants with independent commercial value, apart from any copying or reproduction of content, by materially improving the performance of their products. Defendants accessed Reddit data to obtain the most current versions of Reddit's content structures, thereby capturing the benefits of Reddit's investments without incurring the costs of creating or maintaining comparable systems. Accessing Reddit's structured content at the scale and speed necessary for Defendants' businesses requires evasion of the protection systems safeguarding it, resulting in the misappropriation of Reddit's labor, skill, and expenditures and causing special damages, including the loss of partnerships it would have entered but for these unfair acts.

93. In May 2024, Reddit sent a cease-and-desist letter to Perplexity, demanding that it stop accessing Reddit data. Reddit further demanded that Perplexity cease using Reddit data unless it entered a license agreement expressly authorizing that use. Reddit also noted that Perplexity must stop using Reddit data for commercial purposes regardless of the means by which Perplexity obtained Reddit content. At that time, Perplexity told Reddit that it was not using Reddit content to train any AI models and that it would respect Reddit's robots.txt directives to not scrape Reddit content. Perplexity claims that it "respects robots.txt directives,"

and "only crawls content in compliance with robots.txt."[52] What Perplexity did not tell Reddit was that it was working with SerpApi and others to obtain Reddit data from Google SERPs and that it was using the Defendants' scraping tools to access that data by circumventing the technological protections Google had put in place to protect Reddit's data.

94. After Reddit sent its cease-and-desist letter to Perplexity, the volume of Reddit data cited by Perplexity did not decrease. Just the opposite: Perplexity *increased* the volume of citations to Reddit forty-fold.[53] The increase was so dramatic that an outside observer hypothesized that the increase was due to Perplexity entering a licensing deal with Reddit and thereby obtaining full access to Reddit's data. In truth, there is no license between Perplexity and Reddit but rather a scheme by Perplexity to obtain Reddit's data through the circumvention of the technological measures protecting Reddit data.

95. Perplexity or its agents are working with one or more of SerpApi, Oxylabs, and AWMProxy to circumvent the technological control measures protecting Reddit data to access this content in bulk from Google SERPs.

96. To confirm this hypothesis, Reddit created a "test post" – the equivalent of a digital "marked bill" – that could only be crawled by Google's search engine and was not otherwise accessible anywhere on the internet. The posts contained an uncommon hexadecimal string, was only served to Google, and was withheld from other sources, such as Microsoft's Bing search engine. More specifically, the test post was served with unique robots.txt instructions that prohibited indexing by *everyone* except Google's specific crawler. Within

---

[52] Perplexity Support, *How does Perplexity follow robots.txt*, Perplexity, https://www.perplexity.ai/help-center/en/articles/10354969-how-does-perplexity-follow-robots-txt (last accessed Feb. 6, 2026).

[53] N. Bluman, *Did Perplexity Sign a Deal with Reddit?*, Xfunnel (Apr. 15, 2025), https://www.xfunnel.ai/blog/perplexity-reddit-rumor.

hours, queries to Perplexity's "answer engine" produced the contents of that test post when searches were run in Perplexity for the uncommon hexadecimal string. The only way that Perplexity could have obtained that Reddit content and then incorporated it in its "answer engine" is if it and/or its Co-Defendants on Perplexity's behalf scraped Google SERPs for that Reddit content and Perplexity then quickly incorporated that data into its answer engine.

97. Public reports suggest Perplexity will often resort to this sort of "stealth behavior" to gain access to content it covets. Perplexity has been known to modify its user agent, change operational IP addresses, and ignore or seek to evade directives against crawling of other websites. As just one example, Cloudflare recently noted in an update on its blog that "Perplexity is using stealth, undeclared crawlers to evade website no-crawl directives."[54] Cloudflare's August 2025 update goes on:

> We see continued evidence that Perplexity is repeatedly modifying their user agent and changing their source ASNs to hide their crawling activity, as well as ignoring—or sometimes failing to even fetch—robots.txt files.[55]

98. These practices led Cloudflare's CEO to say that Perplexity's practices better resemble "North Korean hackers" than the practices of a "reputable AI company."[56]

99. Perplexity has continued with its improper and unlawful attempts to access and obtain Reddit data by bypassing or otherwise circumventing the technical measures in place at Google to prevent access to, and widespread scraping of, Reddit data from SERPs, including by

---

[54] G. Corral, *et al.*, *Perplexity is using stealth, undeclared crawlers to evade website no-crawl directives*, Cloudflare (Aug. 4, 2025), https://blog.cloudflare.com/perplexity-is-using-stealth-undeclared-crawlers-to-evade-website-no-crawl-directives/.

[55] *Id.*

[56] M. Kan, *Cloudflare: Perplexity AI Acts Like North Korean Hackers, Ignoring Scraping Blocks*, PC Magazine (Aug. 4, 2025), https://www.pcmag.com/news/cloudflare-perplexity-ai-acts-like-north-korean-hackers-ignores-scraping.

entering into contracts or other agreements with the other Defendants to use their tools to access and scrape Reddit data from SERPs.

100. Perplexity is doing more than simply receiving data from SerpApi. Rather, Perplexity is conducting the actual queries itself. As SerpApi explains in its API documentation, "Our Google Search API allows ***you*** to scrape results from the Google search page."[57] Perplexity circumvents Google's security protocols and gains access, through unauthorized and automated processes, to Reddit data by using (at least) SerpApi's tools or directing others to use those tools.

VI. Defendants' Actions Have Harmed Reddit

101. Defendants' activities are unauthorized and unlicensed by Reddit and are not governed by the guardrails that Reddit insists on to protect Redditors and their communities.

102. Defendants' conduct harms Reddit in various ways. Reddit depends on the contributions of its many communities of Redditors, who care about how their content will be treated when it is posted to Reddit. When they post their content to Reddit, Redditors place great trust in Reddit; they expect Reddit to respect their wishes and take its role as stewards of the content seriously. By bypassing all lawful means for obtaining this content, Defendants also bypass all the strict restrictions placed on Reddit's licensing partners that protect Redditors and their anonymity and privacy.

103. For example, Reddit's licensing agreements with third parties preclude those parties from using Reddit data to target Reddit users through advertisements, profiling, facial recognition, or another improper purpose. As articulated in its Public Content Policy, Reddit

[57] SerpApi, *Google Search Engine Results API*, https://serpapi.com/search-api (emphasis added) (last accessed Feb. 6, 2026).

believes in an open internet, but it "do[es] not believe that third parties have a right to misuse public content just because it's public."

104. Reddit also requires its licensing partners to respect Reddit users' deletions. Users can delete their own content using a self-service tool or they can request that Reddit delete the content. These deleted posts will no longer display publicly on Reddit. Likewise, Reddit requires its licensed business partners to stop using deleted posts and content and provides dedicated tools to facilitate this. One commentator described this policy as a "big win for your privacy."[58] But these privacy and content protections are stripped away when third parties circumvent protection measures to access Reddit data without the user's or Reddit's permission. Defendants' conduct directly and negatively impacts Reddit's reputation, its users' trust, and its ability to operate its business as a community of communities.

105. The privacy protections in Reddit's Public Content Policy are material to Reddit's business model. Users expect Reddit to comply with their deletion requests and do not want to see content they have deleted showing up in the outputs of Reddit's licensed business partners. This concern is not speculative: in a single month, millions of posts are deleted from Reddit by the users who posted them. And each deletion is in turn transmitted to Reddit's licensed business partners for removal. Content accessed by Defendants (and subsequently incorporated into Perplexity's answers), however, is not.

106. Reddit's ability to offer these protections is particularly important given the often sensitive, and personal nature of the content posted on the platform. As Reddit has long

[58] A. Beaty, *Why Reddit's new content policy is a big win for your privacy*, ZDNET (May 9, 2024), https://www.zdnet.com/article/why-reddits-new-content-policy-is-big-win-for-user-privacy/.

recognized, pseudonymity is a core principle to Reddit,[59] and it fundamentally shapes how users engage with the site. By allowing participation without attachment to a real-world identity, Reddit enables candid participation across the platform, including in discussions involving mental health, fertility, addiction, trauma, personal relationships, and other sensitive subjects, as well as in everyday conversations that users would be less willing to engage in under their real names. This pseudonymous participation underlies how users create, share, and manage content on Reddit and is central to the trust users place in the platform.

107. Users commonly rely on deletion for a variety of reasons, including to protect their privacy and their real-world identities. As one Redditor put it, "Sometimes I accidentally overshare something too identifying and only realize it later."[60] As another explained, "Whenever I tell a personal story I am a little worried that someone will recognize me from it. So then when I write some other stuff I sometimes think, hmmm do I want my friends or family to know I wrote that?"[61] More broadly, users routinely delete posts they no longer wish to have publicly visible, whether because circumstances have changed, a discussion has run its course, or the content no longer reflects what they intended to share. Reddit has built its platform, policies, and license agreements around these user expectations, and it takes active steps to reflect user deletions across Reddit and its licensed content through its partnerships.

---

[59] Reddit, *Testing Verified Profiles on Reddit* (Dec. 10, 2025), https://redditinc.com/news/testing-verified-profiles-on-reddit (last accessed Feb. 6, 2026).

[60] *Why do you delete your Reddit posts or comments?*, Reddit, r/AskReddit (Oct. 24, 2023), https://www.reddit.com/r/AskReddit/comments/1o6mqeb/why_do_you_delete_your_reddit_posts_or_comments/.

[61] *If Reddit is anonymous, why do so many people delete their questions/answers?*, Reddit, r/randomquestions (Oct. 15, 2025), https://www.reddit.com/r/randomquestions/comments/1o7nuwd/if_reddit_is_anonymous_why_do_so_many_people/.

108. Defendants' unauthorized access to Reddit content causes concrete injury to Reddit by breaking the system of protections Reddit has in place to uphold its promises to users. Reddit has designed and operates technical systems, policies, enforcement processes, and licensing controls that govern how content is accessed, including content users have requested be deleted. When Defendants access Reddit content outside those systems, Reddit loses the ability to give effect to subsequent user-deletion requests in accordance with Reddit's policies. As a result, Reddit's policy commitments cannot be fully carried out, the value of the resources Reddit has devoted to carrying out its Public Content Policy is diminished, and Reddit's ability to manage its platform in the manner it promises users is impaired.

109. Similarly, Reddit's business and reputation are damaged by Defendants' misappropriation of Reddit data and circumvention of technological control measures because Defendants deprive Reddit of the ability to control (a) what entities have access to its data, (b) how that data is used, and (c) whether Reddit data is handled in a manner consistent with its Privacy Policy, Public Content Policy, and User Agreement. Reddit has entered into licensing agreements with several AI companies for access to the same data that Defendants improperly access and scrape. Permitting Defendants' conduct to continue will discourage those entities from abiding by their existing agreements, including the guardrails in those agreements that protect Redditors, and encourage them and others to instead use similar techniques as Defendants to gain access to the content from Reddit. Defendants' conduct also diminishes the value associated with Reddit's licenses with its licensing partners. Indeed, Google's Complaint against SerpApi confirms that SerpApi's unauthorized access to Reddit's content "threatens to disrupt Google's relationship with the rights holders who look to Google to prevent the misappropriation of the content Google displays." Reddit has lost licensing fees or other

commercial opportunities it would obtain from these and new arrangements, and its reputation and business relationship with existing licensees is harmed by third parties gaining access to the data outside of Reddit's established licensing program. In short, Defendants' actions undermine Reddit's licensing business.

110. As a result of Defendants' actions, Reddit has also been harmed because Reddit data is being accessed without Reddit's permission. By circumventing Google's technological control measures, the Defendants' access has allowed it to unlawfully reproduce and distribute Reddit data without Reddit's authorization or approval.

111. Reddit has also been harmed because its own structured data has been accessed by Defendants without authorization, including the structured data scraped by Defendants at an industrial scale. Reddit has always authorized its partners, including Google, to protect this content from misuse and prevent unauthorized access to that data.

112. Reddit is further harmed by Defendants' conduct because it is forced to invest significant resources into hardware, software, and personnel to improve its technical security systems, surveillance, and anti-scraping efforts to try and prevent the unauthorized accessing and use of Reddit data. It also forces Reddit to require its partners to invest in similar ever-increasingly expensive efforts to protect against Defendants' malfeasance.

113. These and various other harms to Reddit will continue unabated so long as Defendants' schemes for unlawfully accessing and scraping Reddit data persist.

COUNT I – DIGITAL MILLENNIUM COPYRIGHT ACT
Circumvention of Technological Control Measures (17 U.S.C. § 1201(a)(1)(A))

[All Defendants]

114. Reddit repeats and realleges the allegations in the previous paragraphs as if fully set forth herein.

115. Reddit and Google have implemented technological measures that effectively control access to Reddit content. Both companies use advanced technological techniques, as described above, to control unauthorized, automated access to their server systems. These measures, in the ordinary course of their operation, limit the freedom and ability of users to access Reddit content, including by prohibiting automated entities from accessing SERPs and scraping search engine results that include Reddit content.

116. Defendants' actions violate 17 U.S.C. § 1201(a)(1)(A), under which no person shall circumvent a technological measure that effectively controls access to a copyrighted work. Defendants have circumvented these measures in one or more ways, including:

a. Avoiding or bypassing Reddit's measures entirely in order to obtain Reddit's content and services, and the content authored by its users, that appear in Google search results; and

b. Avoiding, removing, deactivating, impairing, and/or bypassing SearchGuard and Google's other technological control measures by using devices, systems, processes, and/or protocols, including large-volume proxy networks, to improperly gain access to Google search results.

117. Defendants have also substantially aided, assisted, or contributed to circumvention of these measures, with knowledge that circumvention would occur, by entering into contracts or agreements with one another aimed at circumventing technological control measures in order to access and scrape Google search results featuring Reddit's content and services, or content authored by Reddit users; by supplying products or services that enable that circumvention; and/or by paying for or otherwise financing the use of such products or services.

118. Pursuant to 17 U.S.C. § 1203(a), Reddit brings this action because it has been injured by Defendants' violations of 17 U.S.C. § 1201(a)(1)(A). Reddit has been injured in one or more ways, including that Defendants have unlawfully circumvented technological measures which effectively control access to Reddit's content and services, and the content authored by its users, and in so doing have accessed and acquired Reddit data outside the legitimate licensing channels Reddit has established containing appropriate contractual guardrails to protect Reddit's users and its users' rights.

119. As a result, Reddit has suffered damages, including, but not limited to, lost profits and business opportunities, reputational harm, and loss of user trust.

120. Reddit is entitled to recover from Defendants the actual damages it has suffered because of their statutory violations, as well as any additional, non-duplicative profits they have earned from them. In the alternative, Reddit may elect to recover statutory damages of no less than $200 and as much as $2,500 for each of Defendants' many statutory violations.

COUNT II – DIGITAL MILLENNIUM COPYRIGHT ACT

Trafficking of Technology, Product, Service, or Device for Use in Circumventing Technological Measure Controlling Access (17 U.S.C. § 1201(a)(2))

[SerpApi and Oxylabs]

121. Reddit repeats and realleges the allegations in the previous paragraphs as if fully set forth herein.

122. Defendants SerpApi's and Oxylabs' actions violate 17 U.S.C. § 1201(a)(2), under which:

> No person shall manufacture, import, offer to the public, provide, or otherwise traffic in any technology, product, service, device, component, or part thereof, that:

> (A) is primarily designed or produced for the purpose of circumventing a technological measure that effectively controls access to a work protected under this title;
>
> (B) has only limited commercially significant purpose or use other than to circumvent a technological measure that effectively controls access to a work protected under this title; or
>
> (C) is marketed by that person or another acting in concert with that person with that person's knowledge for use in circumventing a technological measure that effectively controls access to a work protected under this title.

123. As detailed above, SerpApi provides products and services to facilitate web scraping, including products and services purposefully designed to circumvent effective technological measures that would otherwise prevent access for large scale automated web scraping. SerpApi specifically designs and markets its products and services, including Ludicrous Max and Ludicrous Speed Max, for use in circumventing effective technological measures including Google's SearchGuard, which controls access to search engine results featuring Reddit's content and services, and the content authored by Reddit users. SerpApi's and Oxylabs' offerings have a limited commercially significant purpose apart from circumventing technological control measures, such as Google's SearchGuard. SerpApi and Oxylabs both market their products to others knowing that their products will be used to circumvent technological control measures, such as Google's SearchGuard.

124. Defendant Perplexity has utilized one or more of SerpApi's services or products in order to circumvent Reddit's and/or Google's technological control measures.

125. As detailed above, Oxylabs offers a proxy service that provides web-scraping solutions through a network of proxies that allow for human-like scraping without IP blocking, thereby bypassing or circumventing technical restrictions that would otherwise prevent access for web scraping. Oxylabs specifically designs and markets its products and services for use in

circumventing technological measures including Google's SearchGuard, which controls access to search engine results featuring Reddit's content and services, and the content authored by Reddit users.

126. Pursuant to 17 U.S.C. § 1203(a), Reddit brings this action as someone who has been injured by Defendants' violations of 17 U.S.C. § 1201(a)(2). Reddit has been injured in one or more ways, including that Defendants have unlawfully circumvented technological measures which effectively control access to Reddit's content and services, and the content authored by its users, and in so doing have accessed and acquired Reddit data outside the legitimate licensing channels Reddit has established containing appropriate contractual guardrails to protect Reddit's and its users' rights.

127. As a result, Reddit has suffered damages, including, but not limited to, lost profits and business opportunities, reputational harm, and loss of user trust.

128. Reddit is entitled to recover from Defendants the actual damages it has suffered because of their statutory violations, as well as any additional, non-duplicative profits they have earned from them. In the alternative, Reddit may elect to recover statutory damages of no less than $200 and as much as $2,500 for each of Defendants' many statutory violations.

COUNT III – DIGITAL MILLENNIUM COPYRIGHT ACT

Trafficking of Technology, Product, Service, or Device for Use in Circumventing Technological Measure Protecting Right of Copyright Owner (17 U.S.C. § 1201(b))

[SerpApi and Oxylabs]

129. Reddit repeats and realleges the allegations in the previous paragraphs as if fully set forth herein.

130. Defendants SerpApi's and Oxylabs' actions violate 17 U.S.C. § 1201(b), under which:

> No person shall manufacture, import, offer to the public, provide, or otherwise traffic in any technology, product, service, device, component, or part thereof, that:
>
> (A) is primarily designed or produced for the purpose of circumventing protection afforded by a technological measure that effectively protects a right of a copyright owner under this title in a work or a portion thereof;
>
> (B) has only limited commercially significant purpose or use other than to circumvent protection afforded by a technological measure that effectively protects a right of a copyright owner under this title in a work or a portion thereof; or
>
> (C) is marketed by that person or another acting in concert with that person with that person's knowledge for use in circumventing protection afforded by a technological measure that effectively protects a right of a copyright owner under this title in a work or a portion thereof.

131. As detailed above, SerpApi provides products and services to facilitate web scraping, including products and services designed to circumvent technological measures that would otherwise prevent, restrict, or limit the exercise of a right of a copyright owner. SerpApi specifically designs and markets its products and services, including Ludicrous Speed and Ludicrous Speed Max, for use in circumventing technological measures including Google's SearchGuard, which otherwise prevents the unauthorized and large-scale automated copying and reproduction of search engine results featuring Reddit's content and services and the content authored by Reddit users. SerpApi's and Oxylabs' offerings have a limited commercially significant purpose apart from circumventing technological control measures, such as Google's SearchGuard. SerpApi and Oxylabs both market their products to others knowing that their products will be used to circumvent technological control measures, such as Google's SearchGuard.

132. Defendant Perplexity has utilized one or more of SerpApi's services or products in order to circumvent Google's technological control measures to copy Reddit data.

133. As detailed above, Oxylabs provides products and services to facilitate web scraping through a network of proxies that allow for human-like scraping without IP blocking, thereby bypassing or circumventing technical restrictions that would otherwise prevent, restrict, or limit the exercise of a right of a copyright owner. Oxylabs specifically designs and markets its products and services for use in circumventing technological measures including Google's SearchGuard, which otherwise prevents the unauthorized and large-scale automated copying and reproduction of search engine results featuring Reddit's content and services and the content authored by Reddit users.

134. Pursuant to 17 U.S.C. § 1203(a), Reddit brings this action as someone who has been injured by Defendants' violations of 17 U.S.C. § 1201(b). Reddit has been injured in one or more ways, including that Defendants have unlawfully circumvented technological measures which effectively protect rights in Reddit's content and services, and the content authored by its users and in so doing have accessed and acquired Reddit data outside the legitimate licensing channels Reddit has established containing appropriate contractual guardrails to protect Reddit's and its users' rights.

135. As a result, Reddit has suffered damages, including, but not limited to, lost profits and business opportunities, reputational harm, and loss of user trust.

136. Reddit is entitled to recover from Defendants the actual damages it has suffered because of their statutory violations, as well as any additional, non-duplicative profits they have earned from them. In the alternative, Reddit may elect to recover statutory damages of no less than $200 and as much as $2,500 for each of Defendants' many statutory violations.

COUNT IV – UNFAIR COMPETITION

[All Defendants]

137. Reddit repeats and realleges the allegations in the previous paragraphs as if fully set forth herein.

138. Defendants, through their actions as outlined above, have misappropriated Reddit's labor, skill, expenditures, and/or goodwill, and have displayed bad faith in doing so.

139. By circumventing technological control measures described above in order to access and scrape Reddit data on a large-scale, unauthorized, and automated basis, Defendants have brazenly disregarded technological control measures in which Reddit and Google have invested considerable resources, and in doing so misappropriated real-time Reddit content and services and the timely content authored by Reddit users for Defendants' own commercial gain.

140. SerpApi and Oxylabs have also marketed, sold, and provided products and services for the purpose of circumventing the technological control measures described above in order to access and scrape timely Reddit data on a large-scale, unauthorized, and automated basis, for their own commercial gain.

141. Defendants have taken the actions outlined above in order to secure for themselves an undue competitive advantage, or to damage Reddit's competitive position, including by (a) accessing and scraping timely Reddit data on a large-scale, unauthorized, and automated basis, thereby avoiding the need for any licensing agreement or other commercial arrangement to which Reddit would otherwise be entitled in the marketplace, and reducing the value and desirability of such agreement or arrangements; (b) removing the need for users to access Reddit content directly through Reddit, and thus damaging the commercial utility of Reddit's website, mobile application, and other services; and (c) jeopardizing users' ability to

protect privacy interests in and access to their content, and thus threatening Reddit's ability to attract and retain users.

142. Defendants' conduct violates federal law, as detailed above, and is independently unfair conduct that also violates the state common law of New York.

143. As detailed above, Defendants have acted in bad faith by knowingly circumventing the technological control measures at issue and lauding their ability to do so, and/or by continuing to circumvent the technological control measures after having been blocked from access or otherwise requested to cease their activities.

144. As a result, Reddit has been injured and suffered damages, including, but not limited to, lost profits, business opportunities, reputational harm, goodwill, commercial advantage, and loss of user trust.

COUNT V – UNJUST ENRICHMENT

[All Defendants]

145. Reddit repeats and realleges the allegations in the previous paragraphs as if fully set forth herein.

146. Defendants have been unjustly enriched at Reddit's expense, and equity and good conscience militate against Defendants retaining the benefits they have obtained.

147. Reddit has expended substantial resources to implement technological control measures, policies, and commercial arrangements aimed at preventing access to and scraping of Reddit data on a large-scale, unauthorized, and automated basis, including real-time Reddit content and services and the timely content authored by Reddit users.

148. Through circumventing technological control measures described above, Defendants have gained access to and scraped Reddit data on a large-scale, unauthorized, and automated basis, including misappropriation of real-time Reddit content and services and the

timely content authored by Reddit users, from which Defendants have been unjustly enriched at Reddit's expense.

149. By marketing, selling, and providing products and services for the purpose of circumventing the technological control measures described above in order to access and scrape Reddit data on a large-scale, unauthorized, and automated basis, SerpApi, Oxylabs, and AWMProxy have been unjustly enriched at Reddit's expense. Defendant Perplexity has likewise been unjustly enriched by accessing and obtaining Reddit data through its use of SerpApi and/or the other Co-Defendants' tools.

150. As a result, Reddit has been injured and suffered damages, including, but not limited to, lost profits, business opportunities, reputational harm, goodwill, commercial advantage, and loss of user trust.

151. Defendants' conduct violates federal law, as detailed above, and is independently unjust conduct that also violates the state common law of New York.

152. Equity and good conscience demand that the unjust benefits that Defendants have received should be repaid to Reddit.

COUNT VI – CIVIL CONSPIRACY

[SerpApi and Perplexity AI]

153. Reddit repeats and realleges the allegations in the previous paragraphs as if fully set forth herein.

154. SerpApi and Perplexity have entered into one or more contracts or business agreements for the purpose of circumventing the technological control measures described above in order to gain access to Reddit data on a large-scale, unauthorized, and automated basis, including Reddit content and services and the content authored by Reddit users.

155. Both SerpApi and Perplexity have committed one or more overt acts in furtherance of their agreement, including selling or providing services or products designed to conduct the circumvention described above, paying for and using such services or products, and conducting the circumvention described above. Specifically, Perplexity has contracted with SerpApi for the purpose of using its tool to acquire Reddit data in bulk by circumventing Google's technological control measures, and which Perplexity then uses in its commercial offerings.

156. Each of SerpApi and Perplexity intended by their actions to participate in a plan or scheme designed to circumvent the technological control measures described above in order to gain access to Reddit data on a large-scale, unauthorized, and automated basis, including Reddit content and services and the content authored by Reddit users.

157. Furthermore, Perplexity knows that Reddit objects to Perplexity's unauthorized use of Reddit data. Reddit told Perplexity in 2024 that Perplexity must cease and desist its use of Reddit data in Perplexity's commercial offerings, but Perplexity ignored Reddit's demand, instead obtaining Reddit data through its scheme with SerpApi.

158. SerpApi's and Perplexity's actions led to the commission of one or more torts or other unlawful actions, including those outlined above.

159. As a result, Reddit has been injured and suffered damages, including, but not limited to, lost profits, business opportunities, reputational harm, goodwill, commercial advantage, and loss of user trust.

PRAYER FOR RELIEF

WHEREFORE, Reddit demands a trial by jury and a judgment against Defendants as follows:

a. Injunctive relief enjoining Defendants from:

i. accessing or using Reddit's website, servers, systems, and any data contained therein for the purpose of unlawful data scraping;

ii. accessing or using Google's website, servers, systems, and any data contained therein for the purpose of unlawful scraping of Reddit data;

iii. developing or distributing any technology or product that is used for the unauthorized circumvention of technological control measures and scraping of Reddit data;

iv. using any Reddit data previously obtained through the circumvention of technological control measures in place to prevent the unauthorized and automated scraping of Reddit data; and

v. selling or offering for sale any Reddit data previously obtained through the circumvention of technological control measures in place to prevent the unauthorized and automated scraping of Reddit data.

b. Awarding Reddit damages for all injuries suffered as a result of Defendants' unlawful acts, including:

i. actual, statutory, and/or other compensatory damages as determined by a jury;

ii. disgorgement of Defendants' ill-gotten gains from unauthorized commercial exploitation of Reddit's data; and

iii. pre-judgment and post-judgment interest, in an amount to be determined at trial.

c. Costs of this action, including attorneys' fees; and

d. Such other legal and equitable relief as this Court deems just and proper.

Dated: February 6, 2026
New York, New York

Respectfully submitted,

/s/ William A. Maher
William A. Maher
Justin T. Zimnoch
WOLLMUTH MAHER & DEUTSCH LLP
500 Fifth Avenue
New York, NY 10110
(212) 382-3300
wmaher@wmd-law.com
jzimnoch@wmd-law.com

Reid M. Bolton (*Pro Hac Vice*)
Matthew R. Ford (*Pro Hac Vice*)
BARTLIT BECK LLP
54 W. Hubbard Street, Suite 300
Chicago, IL 60654
(312) 494-4400
reid.bolton@bartlitbeck.com
matthew.ford@bartlitbeck.com

Giovanni Sanchez (*pro hac vice*)
BARTLIT BECK LLP
1801 Wewatta St., Suite 1200
Denver, CO 80202
(303) 592-3100
giovanni.sanchez@bartlitbeck.com

Attorneys for Plaintiff Reddit, Inc.