[redacted]

IN THE UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE: OPENAI, INC. COPYRIGHT INFRINGEMENT LITIGATION This Document Relates To: All Cases	Case No. 1:25-md-3143-SHS-OTW

DECLARATION OF PADMA PRIYA GAGGARA IN SUPPORT OF MICROSOFT CORPORATION'S MOTIONS FOR SUMMARY JUDGMENT

I, Padma Priya Gaggara, hereby declare as follows:

A. Personal Background

1. I am a Vice President of Microsoft AI, a group within Microsoft Corporation ("Microsoft"), and submit this declaration in support of Microsoft's Motion for Summary Judgment. Unless stated otherwise, all facts stated herein are within my personal knowledge. If called upon, I would be willing to testify as set forth in this declaration.

2. As a Microsoft employee since 2003, I have broad experience ranging from hands-on development of technical components that are integrated in Microsoft's Bing search technology, leadership and management of large engineering teams, and sustained involvement in the development of Microsoft's search and web platforms. Through my work at Microsoft, I have observed the development of Microsoft's Bing search technology and AI innovations, as well as the evolution of industry practices and norms concerning the integration of traditional web search platforms with generative AI technologies.

[redacted]

3. I hold a Bachelor of Technology degree in computer science from Jawaharlal Nehru Technological University (1999) and a Masters degree in computer science from the University of Arizona (2003).

4. I began my career at Microsoft in 2003 as a Software Design Engineer within Windows Live, a discontinued set of web services and predecessor to Microsoft's Bing search engine. I transitioned to the Bing team in 2008, just before the 2009 launch of the Bing search engine, where I served numerous roles leading engineering teams to drive the evolution of Microsoft's Bing search technology and user experience, including developing advanced algorithms to improve the quality of search results, innovative machine-learning powered search features in Bing, and continuous quality initiatives, amongst many others.

5. My responsibilities have progressively evolved to overseeing large-scale engineering teams responsible for critical components of Microsoft's web and search ecosystem, including platforms supporting Bing, Edge, and broader web data infrastructure. Since 2021, I have been responsible for end-to-end management and decision-making concerning Microsoft's web indexing, web crawling, and web-scale infrastructure, which powers the Bing search engine and search-powered products, in accordance with responsible practices in the search engine industry. I am also the individual directly responsible for ensuring that Microsoft's Bing search products and search-powered features honor publisher content controls. I also drive efforts to foster a healthy web ecosystem for the benefit of both web publishers and users.

6. In August 2025 and in January 2026, I sat for depositions in this case as Microsoft's designated corporate witness pursuant to Rule 30(b)(6). My testimony covered numerous topics, including the Bing Index data that Microsoft shared with

[redacted]

OpenAI for experimentation, Microsoft's provision of GPTBot crawling services to OpenAI, and the process of indexing, storing, and/or removing content during the course of operating the Bing Index, among others. I sat for deposition again in July 2026 in my capacity as a non-retained rebuttal expert witness on substantially related topics.

B. Microsoft's Crawling and Indexing Practices, Standards, and Procedures

7. Search engines help users find relevant, comprehensive, and authoritative web content. To do so, search engines rely on a search index. Microsoft's search index, known as the Bing Index, is a structured data store—with a repository of links to webpages, stored webpage content, and invisible metadata associated with a webpage. The Bing search engine and Bing search-powered features rely on the Bing Index to retrieve relevant web content and then provide ranked search results near instantaneously in response to a user's request.

8. Building a search index requires identifying publicly accessible webpages, retrieving information from those webpages, and organizing that information in a way that can be later retrieved in response to a user's search query. Search indexes are built through a well-established pipeline that major search engines have used for decades, including Google (launched in 1998) and Microsoft's Bing search engine (launched in 2009, and predecessor search products operated beginning in 1998).

9. Microsoft conducts its crawling and indexing activities in accordance with best practices for responsible search engine operators. These practices are intended to balance the interests of internet users seeking reliable and useful information with those of web publishers seeking to optimize their content's discoverability through search engines.

[redacted]

a. Microsoft's Bingbot Crawling and Web Indexing Pipeline

10. The first step in building a search index is crawling the web to discover new and updated content. Microsoft's web crawler, known as Bingbot, discovers URLs from every corner of the internet using complex, automated algorithms to determine which webpages to crawl and how often to crawl them.

11. The Robots Exclusion Protocol ("REP") is a well-established foundational standard governing responsible crawling practices and that all major search engine crawlers follow, including Microsoft's Bingbot. The REP enables websites to communicate crawl preferences to automated crawlers.

12. Web publishers implement the REP by placing a robots.txt file at the root directory of their website with instructions for crawlers indicating whether they may access specified paths. Web publishers may customize such permissions and restrictions for specific crawlers. Bingbot identifies itself in the request's user-agent string so that website owners can identify Bingbot and distinguish it from other crawlers or ordinary users, consistent with best practices for transparently identifying the party responsible for the crawling activity.

13. After discovering a URL, Bingbot reads and follows the crawl directives placed in a website's robots.txt file before accessing the website's content. Where a web publisher uses the "disallow" directive to indicate that Bingbot should not crawl a website or specified portions of that website (known as a "directory"), Bingbot honors that instruction and does not access or download that content.

14. If the robots.txt file allows Bingbot to crawl a webpage, Bingbot downloads (or "fetches") the webpage's contents and sends it to Bing's indexing pipeline, which algorithmically determines whether to add the webpage to the Bing Index. This stage

[redacted]

involves processing and parsing the webpage's contents, metadata, headings, canonical tags, robots meta directives, among other signals.

15. At the indexing stage, search engines abide by the fundamental principle of the web ecosystem that website operators can communicate instructions to search engines about how to handle their content after crawling, and compliant search engines will follow them. Robots meta tags are a decades-old practice that provide a reliable mechanism for web publishers to exercise more granular, page-level control over how search engines should use their content. Meta tags provide instructions to search engines on whether to index a crawled webpage and how the content should be made available to users. Because search engines' processing software cannot ingest freeform language on a website to determine the scope of their crawling permissions, robots meta tags give webmasters a universal and deterministic way to communicate their content control preferences to search engines. In other words, robots meta tags set strictly defined rules that search engines can follow with a high degree of precision, whereas alternative methods of signaling content control preferences cannot guarantee the desired outcome.

16. In addition to following these well-established standard web protocols for search engines, Microsoft offers web publishers ways to apply the standard meta tags, NOARCHIVE and NOCACHE, to exercise more granular control over whether or how they want their content to be used in Microsoft's generative-AI answers grounded in Bing search results, such as Copilot. Specifically, a web publisher can apply the NOARCHIVE meta tag to indicate that their content should not be used for grounding Copilot answers. Alternatively, web publishers can apply the NOCACHE meta tag (either alone or in combination with the NOARCHIVE meta tag) to communicate that their content may be included in Copilot answers but limited to only the URL, snippet, and title. Using these

[redacted]

meta tags to limit or prevent the use of web content for grounding Copilot answers does not impact how that content appears or ranks in traditional Bing search results. In other words, a web publisher can apply the NOARCHIVE and/or NOCACHE meta tags, and their content will continue to appear in Bing search results the same way it always has, while instructing Bing that this content, either in full or in part, should not be used for grounding answers in Microsoft's generative AI experiences, such as Copilot.

17. This complex algorithmic pipeline continuously operates at an enormous scale to account for continually changing web content. Bingbot crawls tens of billions of URLs daily, each of which are processed and selected on a page-by-page basis. [redacted]

18. The Bing Index aims to cover a broad range of online content to fulfill Bing's purpose of helping users locate useful information within the internet's vast expanse.

b. Crawling and Indexing Limitations

19. Bing does not guarantee that every crawled page will be indexed, displayed in search results, or retained indefinitely.

20. First, the Bing Index does not necessarily contain meaningful content for every URL indexed. This may be because the nature of the URL lacks content for indexing (known as a "definitive"). For example, "gmail.com," a known URL that will appear in search results when relevant, does not contain textual content that would otherwise be indexed. In other cases, Bingbot may discover (and therefore "knows") a URL but the webpage's robots.txt files disallow Bingbot from accessing or downloading its contents.

[redacted]

At other times, the contents of a webpage may be removed from the Bing Index if the webpage's robots.txt directives later change to disallow Bingbot's access, even if that webpage's content was crawled and indexed at an earlier point in time. In other instances, the robots.txt directives may allow Bingbot to access the contents of a webpage but apply the NOINDEX meta tag directive, which prevents Bing from indexing those contents (and that webpage will likewise not appear in search results). These website behaviors are frequently observed with "transient" websites that seek to evade discovery by search engines, such as those hosting illicit content, including sources of pirated content.

21. In addition, technical limitations prevent Bingbot from accessing certain types of content. For example, Bingbot's configurations do not allow it to retrieve content that is only available through torrent protocols. Microsoft designed Bingbot to avoid the inclusion of such content in the Bing Index because torrent protocols and peer-to-peer networks are known means of transmitting illicit content, including pirated content.

22. Microsoft takes measures to avoid including unauthorized copyrighted content in the Bing Index and continuously works to remove such content once Microsoft learns of it. To the extent that copyrighted content is included in the Bing Index without the copyright owner's permission, Microsoft follows strict content removal protocols communicated through Digital Millennium Copyright Act ("DMCA") takedown requests. This allows publishers to inform search engines of any content published on a third party's website without the copyright owner's authorization. Governments around the world may also submit requests to remove access to content made available through Microsoft's products or services. When Microsoft receives a valid request, it promptly removes the content from the Bing Index altogether, rendering the content unavailable in Microsoft's internal systems and not retrievable by users through web search results.

[redacted]

Microsoft also strives to be transparent with users and publishers about content removal and publishes regular transparency reports that disclose the takedown requests received and approved in a given timeframe.

23. These limitations reflect Microsoft's design choices to support the web ecosystem, respect publisher preferences, and provide users with safe and relevant search results, consistent with widely recognized best practices in web search.

c. Dynamic Updating of the Bing Index

24. The Bing Index is not static. It is dynamically updated on an ongoing basis to reflect changes on the web and to maintain the quality, freshness, and compliance of Bing's search results.

25. Websites frequently publish new content, modify existing content, remove content, or may at any time alter their permissions. To keep search results accurate and useful, Bingbot regularly revisits webpages that it has previously indexed to check whether the pages have been updated or removed. Dynamic updating helps ensure that newly available content is discoverable in search, that the Bing Index and Bing search results reflect changes that web publishers make to content or webpage permissions, and that content that is no longer available is removed from the index.

26. Crawling frequency determines how frequently Bingbot revisits and refreshes webpages. By default, Bingbot prioritizes crawling known webpages and those sending signals indicating that they may have been updated. Crawling frequency may also vary depending on the nature of a webpage. For instance, the pipeline is engineered to frequently refresh news content (meaning webpages that reflect current events as indicated by signals such as URL age, historical change patterns, site structure, etc.) and sites with a relatively larger internet presence. At the same time, news articles that were

published many years ago and no longer receive continuous traffic are likely to phase out of the frequent recrawling cycle, causing significant gaps in the time that Bingbot recrawls the content. Because Bing search and Bing search-powered features (such as web grounding in Copilot) reflect downstream uses of content in the Bing Index, it is important that webmasters provide signals to Bing if they update the content or meta tags for a webpage with low activity.

27. To optimize content freshness and ensure that Bing accurately accounts for a webpage's most up-to-date content control preferences, Microsoft encourages webmasters to use the IndexNow protocol to manually notify Bing when they create, update, or delete web content. Webmasters can also promote discoverability of their webpages using XML sitemaps that signal freshness or leverage crawlable internal and external linking to signal authority.

28. Bingbot may also adjust its crawl rate in line with "politeness limits" to ensure that its crawling activities do not exceed the website's available resources. These limits may be algorithmically determined based on, *e.g.*, the size of the website or error codes indicating that the website may be overburdened by more frequent crawling. Microsoft gives webmasters tools to control and customize Bingbot's crawl frequency. If a webmaster seeks to limit crawling activity to conserve or balance resources, the webmaster can use the Crawl Control feature available in Microsoft's Bing Webmaster Tools to preset the frequency and/or timing of crawling activities. Alternatively, webmasters can use this tool to increase Bingbot's crawl frequency if they want their content refreshed faster than the automated crawl rate.

29. In my experience, the continuous updating of a search index is a critical aspect of maintaining a useful, accurate, and responsible search engine. These

[redacted]

algorithmic signals reflect a longstanding practice of reciprocal communication between webmasters and search engines, which forms the foundation of a balanced web ecosystem. Collectively, these practices help ensure that the Bing Index reflects the current web, which benefits both users seeking information and publishers seeking to control the availability of their web content.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on September 3, 2026.

Signed by:
Padma Priya Gaggara
A316448A03BB4D5...

PADMA PRIYA GAGGARA