IN THE UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE: OPENAI, INC. COPYRIGHT INFRINGEMENT LITIGATION This Document Relates To: Case No. 1:23-cv-11195 Case No. 1:24-cv-03285 Case No. 1:24-cv-04872	Case No. 1:25-md-3143-SHS-OTW

DECLARATION OF FABRICE CANEL IN SUPPORT OF MICROSOFT CORPORATION'S MOTION FOR SUMMARY JUDGMENT

I, Fabrice Canel, hereby declare as follows:

A. Personal Background

1. I am the former Principal Product Manager of the Bing Web Data platform at 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. I was a Microsoft employee from April 1997 until July 2026 and a Principal Program Manager for Bing. I hold Masters and Ph.D. degrees in Physics and Applied Mathematics from the Université Joseph Fourier.

3. In January 2026, my deposition was taken in this case both as an individual and as Microsoft's designated witness under Rule 30(b)(6). My testimony concerned numerous topics, including Microsoft's policies and practices with respect to web content controls (*e.g.*, Robots Exclusion Protocols and robots meta tag directives), webmasters' use of such controls to communicate Terms of Use, mechanisms for recognizing and

filtering paywalled content in Copilot (formerly known as Bing Chat). My deposition was taken again in this case in July 2026 during the expert discovery period as a non-retained expert witness whose opinions may be presented at trial.

4. From the late 1990s through July 2026, I was responsible for managing teams that discover, select, crawl, process, and index internet content to ensure up-to-date and comprehensive search results from the Bing search engine. The teams located content published on the web; selected which and how much content from each website to attempt to include in the Bing's search index; visited and downloaded the selected content consistent with any permissions provided by web publishers; and processed the downloaded content for potential inclusion in the Bing Index.

5. I was also responsible for developing and managing Microsoft's Webmaster Tools. These comprise a suite of tools that Microsoft provides to web content providers (which I often refer to as "web publishers" or "publishers") and their website administrators (also known as "webmasters") to optimize and monitor how Bing uses their web content. Since February 2026, Webmaster Tools have provided webmasters the ability to optimize and monitor the use of web content in Microsoft's AI knowledge assistant Copilot and similar Microsoft Copilot services.

6. My role on Bing's search frontlines made me an expert on the evolving search industry in general as well as its norms and standards. I was Microsoft's best-known ambassador to the web publisher and search community. I regularly attended and spoke at industry conferences, published many blog posts and articles, and gave frequent interviews about developing and maintaining a healthy and robust web ecosystem that benefits all participants. The topics that I spoke about included, for example, content discovery and indexing, ways for websites to use Microsoft's industry-leading IndexNow

tools to notify Microsoft of website changes, search engine optimization ("SEO"), and, more recently, generative engine optimization ("GEO") as well as evolving views regarding quality user interactions. I maintained constant contact with various web publishers and administrators to educate them about and solicit feedback regarding Microsoft's products and services and to help them optimize their sites for Bing.

B. Search Engine Industry Standards, Practices, and Norms for Publisher Content Controls in Traditional Search

7. Microsoft has operated a search engine since 1998; prior versions were known as MSN Search, Windows Live Search, and Live Search before Microsoft launched the Bing search engine in 2009. Today, Bing is the second largest search engine in the United States and worldwide.

8. Several years after the World Wide Web (the "Web") was introduced in 1991, the first content control standards emerged to let web publishers to communicate to crawlers which parts of their websites not to access. After this foundational development, the search engine industry has continued to provide and expand the controls available to web publishers over how search engines interact with their content. These tools represent well established and broadly adopted foundational standards that all major search engines recognize. Microsoft's Bing search engine has always honored this common set of standardized content controls.

9. These controls developed alongside the search industry. As search engines became an increasingly important way for users to discover online information, web publishers sought additional ways to communicate their preferences on search engines' access to and use of their content. In response, the industry adopted a set of common technical mechanisms that enable web publishers to provide instructions that grant, limit,

or restrict how search engines access and use their web content at various stages of the search pipeline. These mechanisms now form the foundation of the web ecosystem and are a fundamental part of the relationship between web publishers and search engines.

10. Introduced in 1994, the Robots Exclusion Protocol ("REP") is the most widely recognized publisher control. The REP governs automated crawling, *i.e.*, directions to a crawler, such as Microsoft's Bingbot or Google's Googlebot, whether it may retrieve content from that website or specified locations on a website. Web publishers can easily implement the REP by placing a robots.txt file at the root directory of their website, which may identify particular crawlers, with directives as to whether the crawler is allowed or disallowed access to the contents of that website or specified areas of that website. When a compliant search engine crawler, such as Bingbot, discovers a webpage, it first requests the robots.txt file from the website's root directory before accessing any content. The crawler reads these instructions and adjusts its crawling behavior accordingly. If a website's robots.txt file disallows the crawler from accessing that website's content, a compliant crawler will not visit (or "crawl") the webpage. If the robots.txt file allows access, the crawler downloads (or "fetches") the webpage's contents and sends information signals to the search engine's indexer to process, analyzes, and potentially store the content in the search engine's index.

11. Robots Meta Tags are another longstanding category of web content controls. Meta tags were first introduced in 1996 to address the need for more granular, page-level control over how search engines index and/or display content. Web publishers can place a meta tag (or combination of meta tags) within a webpage's HTTP response header. After the crawler fetches a webpage, the search engine's indexer processes that content, which involves extracting and parsing metadata, headings, canonical tags, robots

meta directives, amongst other signals. At this stage, the indexer reads these meta tag instructions to determine whether it may store the webpage in the search engine index and surface it in search results. Meta tags may contain a range of directives:

a. The NOINDEX meta tag was one of the first available meta tags. It instructs search engines not to store a page in the search index, and the page should not appear in search results.

b. The NOARCHIVE meta tag was later introduced in the early 2000s. It instructs search engines not to display or otherwise make available to users a cached (or archived) copy of a webpage. The related NOCACHE meta tag also instructs search engines not to make cached/archived copies of webpages available to users.

c. The NOARCHIVE and NOCACHE meta tags have evolved alongside web infrastructure. Historically, when webpages loaded slowly or unreliably, many search engines maintained cached copies of webpages and offered users access to those cached versions alongside a link to the live page. As loading speed and reliability improved, search engines largely stopped providing links to cached-pages in search results. NOARCHIVE and NOCACHE meta tags have since evolved into a way for web publishers to indicate that their content is subscription-based or paywalled. When applying these meta tags, search engines may still index and provide links to the content in search results without compromising the webpage's ranking.

12. Web publishers have widely adopted these content controls to communicate their terms of use to search engine crawlers. Because crawlers can only read the files, meta

directives, and properties of a given webpage, rather than plain language, these longstanding content controls provide an easily implemented, universal way for web publishers to reliably communicate instructions to search engines. Rather than forcing an all-or-nothing choice, these controls provide publishers flexible options to control how their content is made available to search engine users.

13. In my experience, the search engine industry and web publishers have historically favored extending existing content-control standards whenever possible, rather than requiring web publishers to learn and deploy a wholly new control system for each newly introduced search feature.

C. Search Engine Optimization and Monitoring Tools

14. As detailed above, I was primarily responsible for developing Microsoft's Webmaster Tools, a suite of tools for web publishers and webmasters (*i.e.*, website administrators and content managers) can use to exercise additional controls over their web content, monitor data insights, and optimize their website's search performance in Bing. Any webmaster can create a Webmaster Tools account free of charge. Google similarly provides search engine optimization and monitoring tools through its Google Search Console.

15. Through my work assisting with discovery in this case, including collecting search performance data from News Plaintiffs' Bing Webmaster Tools accounts, I understand that each News Plaintiff holds a Bing Webmaster Tools account.

16. Bing Webmaster Tools provides webmasters with additional ways to exercise content controls. For instance:

a. If a webmaster needs to quickly hide a webpage from Bing or Copilot, the webmaster can temporarily block the display of URLs in Bing search results.

This gives webmasters time to make any necessary changes to the webpage's metadata or remove the page altogether.

b. Webmasters can also use the Crawl Control feature to control the speed at which Bingbot makes requests for webpages or set crawling limits. This informs the "politeness standards" that Bingbot follows. "Politeness standards" are responsible crawling practices that are used to ensure automated crawling activities do not disrupt, overload, or degrade websites' performance. Webmasters can also use the Crawl Control tool to increase crawling frequency if they want to ensure that Bing reflects the most up-to-date content.

17. Microsoft's Webmaster Tools also provide data points, SEO tools, and targeted insights that webmasters can leverage to monitor, manage, and optimize their website's performance. For instance, webmasters can track current search performance and historical web traffic trends, monitor crawling activity, analyze keywords and corresponding search volumes, and scan their websites or sitemaps to audit technical issues on demand.

18. Microsoft's Webmaster Tools features are also designed to optimize webmasters' website performance. Each webmaster's account includes SEO reports with page-level recommendations for improving search rankings as well as guidelines for best practices to improve performance.

19. The IndexNow protocol, which I helped create in October 2021, enables webmasters to inform participating search engines of webpages with new or updated content (including meta tag updates or content removals) and to optimize crawling activities. Using IndexNow, web publishers can proactively trigger discovery of content

changes, which is especially important to news publishers who historically rely on freshness to maximize their reach. With the advent of generative AI answers grounded in web results, this is a known best practice for ensuring that content controls are accurately recognized and reflected in Bing search and search-powered features.

D. Typical Practices for Communicating Evolving Content Controls

20. Search engine operators routinely keep webmasters apprised of best practices for managing content and implementing content control mechanisms.

21. Several avenues are available for doing so (many of which I have used in my own experience):

a. I have regularly published blog posts to notify webmasters about any changes or updates to the available standard content controls.

b. I have also written regular blog posts to keep webmasters apprised of SEO best practices for enhancing the discoverability of their content in search.

c. Microsoft may also notify webmasters of updated features and tools through the Notifications Center in Webmaster Tools.

d. Extensive technical documentation is also readily accessible, such as in the Webmaster Tools "Help & How-To" pages, for webmasters to reference on demand.

e. I also regularly attended and delivered keynotes at web publisher conferences, along with other Microsoft executives, including at Pubcon, Search Marketing Expo, brightonSEO, Tech SEO Connect, Botify Connect, and Search Marketing Summit, amongst numerous other webinars, interviews, and industry events.

E. Continued Evolution of Publisher Content Controls Following the Introduction of Generative AI Grounded in Search

22. I was part of the team supporting Microsoft's introduction of Bing Chat and its integration with Bing, first available in preview in February 2023 and generally available in December 2023. Bing Chat was the first major consumer generative AI chatbot with RAG capabilities. Bing Chat combined the capabilities of OpenAI's LLMs with Bing search retrieval, giving it the capability to provide prose answers grounded in web search results.

23. Following the launch of Bing Chat, I regularly communicated with web publishers about Bing Chat and its use of web content to provide helpful answers with clear links to the sources of information and to seek web publishers' feedback. Just weeks after launch, for example, I delivered a keynote presentation at Pubcon Austin 2023, a longstanding and widely publicized SEO conference. There, and in numerous public speaking engagements thereafter, I explained Bing Chat and its continued reliance on standard SEO best practices, including use of tools such as IndexNow. These opportunities to engage with the SEO community at scale facilitated direct lines of communication between Microsoft and web publishers to engage in ongoing discussion.

24. The web publishers' feedback was extremely important to Microsoft, which wants to be a trusted and collaborative partner both in the broader web ecosystem and with the web publishers with which Microsoft has held longstanding relationships. I understand that many web publishers were extremely satisfied with how Bing Chat surfaced their content to users through linked results. Some web publishers providing paywalled or subscription-based content wanted options to exercise choices about the use of their content for grounding Bing Chat answers without changing how their content surfaced in traditional Bing search.

25. To ensure that Microsoft continued to foster a healthy web ecosystem that supported web publishers, I led an effort to develop practical and effective mechanism to provide additional options for web publishers to communicate more specific preferences about the use of their content to ground generative AI products, such as Bing Chat.

26. Microsoft innovated to provide new ways for web publishers to apply the longstanding NOARCHIVE and NOCACHE meta tags to prevent or limit the use of their content for Microsoft's AI generative AI experiences. This industry-leading practice gave web publishers more granular control over the use of their content in AI products without impacting how their content appeared in traditional search results.

27. On September 22, 2023, I published a blog post announcing how webmasters could leverage standard content controls to communicate their preferences about the use of their content in Microsoft's generative AI products:

- Content tagged NOARCHIVE would not be used for grounding content in Microsoft's generative AI products, including Bing Chat and later Copilot, and likewise would not provide a link to that content in responses answers provided to users.
- Content tagged NOCACHE may be used for grounding Microsoft's generative AI products, but limited to only the URL, snippet, and title.
- Content tagged both NOARCHIVE and NOCACHE would be treated in accordance with the above NOCACHE standard.
- Content without either NOARCHIVE or NOCACHE could be used for grounding Microsoft's generative AI answers and benefit from increased ranking opportunities.

28. Importantly, this protocol enabled webmasters to exercise choices about the use of their content in Bing Chat without affecting its discoverability in *traditional* Bing search results. In other words, a webmaster who uses the NOARCHIVE or NOCACHE

meta tags to prohibit or limit the use of their content in Microsoft's generative AI products will not also remove or otherwise limit that same content from Bing search results.

29. This approach reflected practical considerations. Web publishers were already well versed in meta tag controls. The standard meta tags that Microsoft selected matched web publishers' expectations about downstream uses of web content and reflected how web publishers used those tags in practice. For example, the vast majority of web publishers already used NOARCHIVE and/or NOCACHE to indicate paywalled or subscription-based web content, consistent with prior guidance. Because most of the webmasters that wanted to limit or prevent the use of their content for grounding offered paywalled[1] or subscription-based content, their content control preferences were automatically implemented without the need for them to take additional steps. Additionally, search engines already maintained infrastructure to read, interpret, and enforce meta tag directives. Building on existing controls promoted consistency, predictability, and ease of implementation for both web publishers and search providers.

30. Microsoft's September 2023 blog post was widely publicized in the SEO community. In addition to the blog post, Microsoft communicated the extended application of these meta tags to AI through existing channels, including web publisher conferences, individual meetings and direct correspondence with web publishers, mass notifications automatically pushed out to Webmaster Tools account holders, and additional publications providing guidance on best practices (which have persisted through my departure).

[1] At Microsoft, "paywall" is used as an overinclusive term for any content tagged NOARCHIVE/NOCACHE, consistent with longstanding industry practice, to communicate a restriction on access or use. In practice, the term "paywall" can describe many websites that do not restrict access to users who do not have a subscription.

31. Microsoft continuously took steps to ensure that Bing Chat recognized these content controls as intended. Microsoft consistently monitored and promptly addressed any system-level errors that inadvertently resulted in so-called "paywalled" content returned in the search results provided for grounding.

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

Executed on September 3, 2026.



FABRICE CANEL