The Impact of LLM Adoption on Online User Behavior*

Nicolas Padilla[†,1], H. Tai Lam[†,2], Anja Lambrecht[1], and Brett Hollenbeck[2]

[1]London Business School

[2]Anderson School of Management, UCLA

December 2025

Abstract

The adoption of AI tools, and especially Large Language Models (LLMs), has the potential to significantly transform how users engage with information online, potentially serving as substitutes or complements to existing digital resources. We use detailed clickstream data from 2022 and 2023 to examine users' online behavior following the adoption of large language models. We document a significant decrease in online search activity, a typical entry point to content consumption. Online searches drop slowly, suggesting a period during which users learn to use LLMs, but eventually adopters' level of online search in traditional search engines is more than 20% below the pre-adoption period, though there is heterogeneity across types of queries. We then turn to the effect of LLM adoption on website traffic. We document that while frequently visited websites are not affected, smaller websites suffer a significant drop in visits. In line with these results, we then report a significant drop in display ad exposures, especially to consumers with high levels of retail activity, though we do not find a reduction in search ad exposures. Last, we study two distinct categories of websites: education-related websites and user-generated content platforms. We document a significant drop in visits to education-related websites and heterogeneity across user-generated content platforms with a pronounced negative effect on Stack Overflow but no significant effect on Wikipedia, Reddit, and social media. We discuss implications for online content creators, for GenAI firms, and for public policy.

Keywords: *platforms, advertising, AI, large language models*

---

*Nicolas Padilla is Assistant Professor of Marketing at London Business School (`npadilla@london.edu`). H. Tai Lam is Assistant Professor of Marketing at UCLA Anderson (`h.tai.lam@anderson.ucla.edu`). Anja Lambrecht is Professor of Marketing at London Business School (`alambrecht@london.edu`). Brett Hollenbeck is Associate Professor of Marketing at UCLA Anderson (`brett.hollenbeck@anderson.ucla.edu`). We thank Rafael Greminger, Yufeng Huang, Xu Zhang, Ilya Morozov, Andrey Simonov, and participants of the seminars at Kellogg and Chicago Booth for comments. The authors wish to acknowledge the financial support of the Research and Materials Development Fund (RAMD_Lambrecht_A_24/25_8872 and RAMD_Padilla_N_24/25_8876) at London Business School. Sangwook Suh and Yichen Yu provided valuable research assistance. All errors are our own.

†: These authors contributed equally to this work and are ordered reverse alphabetically.

1 Introduction

The rise of AI tools, and especially Large Language Models (LLMs), has the potential to profoundly alter how users engage with online content and, in turn, how such content is created and monetized by platforms and publishers. If users shift queries away from traditional search engines that direct traffic to a wide range of websites toward LLMs that provide synthesized answers and fewer pathways to external content, the underlying revenue model for content creators may be jeopardized. These shifts also raise competitive questions, as LLMs increasingly operate as alternative gateways to information that challenge the central role long held by search engines.

At the same time, LLMs may serve as a complementary tool, whereby the aggregation of information fulfills a distinct function that is complementary to traditional search and online content consumption. For example, LLMs excel at understanding complex queries and prestructuring broad sets of information in an accessible way, but at times struggle at reliably providing specific factual details or credible primary sources. This balance of strengths and weaknesses could result in their adoption, facilitating more efficient initial search, which then results in a greater overall search volume and visits to publisher websites.

Understanding whether LLMs substitute for or complement traditional online search and content consumption is important for evaluating their impact on content providers as well as the competitive structure of the broader web ecosystem.

In this paper, we first investigate how LLM adoption affects online search, a typical entry point for online activity. LLMs' ability to answer questions suggests substitution, but concerns around response quality may constrain this role, and potentially increase overall search volume. Second, to assess whether LLM use substitutes for broader online activity and its impact on web publishers, we evaluate its impact on adopters' online traffic overall, for smaller vs larger websites, and for two distinct content categories that differ in the nature of the information they provide and the roles they play within the online ecosystem: educational websites and user-generated content sites. Third, we study whether LLM adoption affects users' advertising exposure, which may shift web publishers' ability to monetize content and retailers' ability to reach consumers.

Our analysis draws on a comprehensive panel dataset that tracks detailed desktop browsing behavior over 2022 and 2023. These data allow us to identify users' adoption of LLMs from observed usage patterns and to estimate the impact of adoption on several dimensions of online activity. We implement a staggered difference-in-differences design in which users who adopt later in the sample period serve as the control group for earlier adopters. We further show that traffic to email, retail, and news domains remains unaffected by LLM adoption during our study period; these categories therefore serve as controls for a users' baseline levels of online activity.[1]

Overall, our results suggest that LLM usage substitutes for some – but not all – online activities, and we do not find evidence for any complementary relationship. We document a significant decline in online search queries, which emerges gradually rather than immediately: traditional online search falls by more than 20% from about 20 weeks after adoption, suggesting an initial period during which users learn about the use of LLMs. The decline is pronounced for informational queries, including question-based searches and both short and long queries. In contrast, navigational searches (Blake et al., 2015) using a single branded term do not decline, consistent with the absence of outlinks in LLMs during our study period and the difficulty of using LLMs as navigational tools.

A reduction in search activity may indicate that users turn to LLMs as a substitute for browsing the broader web. We find that LLM adoption does not reduce the total number of URL calls to the most frequently accessed websites, but it does lead to a pronounced decline in traffic to smaller websites. We likewise demonstrate a decline in traffic referred from search to these websites. This pattern underscores a competitive asymmetry: while large, well-established sites remain relatively insulated, smaller sites that are most dependent on search-driven referrals experience, on average, substantial traffic losses. We document that

[1]During the period of our data, LLMs were trained on frozen pre-training data and had no live information retrieval or outlinks, and were therefore unable to provide current news content.

these losses are pronounced in the education category, where LLMs support writing tasks, fact-finding, and skill acquisition (Humlum and Vestergaard, 2025) and affect web publishers monetizing through both subscriptions and B2B sales. By contrast, we find heterogeneous effects for user-generated content platforms, with no effects on Wikipedia and Reddit but drops in traffic to Stack Overflow, suggesting that some types of user-generated information are more easily substituted for than others.

Substitution away from direct website visits has implications for monetization. We find that display ad exposures fall significantly after LLM adoption, especially among users with high levels of retail activity –individuals likely to be particularly valuable to advertisers. In contrast, search ad exposures do not decline significantly, consistent with the stability of navigational searches that often trigger paid search placements.

Our data capture an early stage of LLM adoption when models relied on fixed pre-training data and provided no outlinks. However, we expect our main conclusions to remain relevant given recent developments such as Google's AI Overviews, retrieval-augmented models, and models selectively providing references and outlinks. LLM-generated answers embedded in search results as AI Overviews likely trigger the same substitution mechanism we document and may reinforce traffic losses for smaller sites. Retrieval-augmented LLMs increase the attractiveness of LLMs relative to traditional search, suggesting that substitution may strengthen over time. While some models now offer curated outlinks, these are more limited for users of free versions and typically point to only a small set of sources which may further concentrate traffic rather than restoring broader referral patterns. Taken together, these developments imply that the behavioral and economic forces we observe for 2022–2023 are likely to persist, and potentially intensify, as LLM capabilities expand.

Our findings have implications for online content providers, GenAI firms, and policymakers. For content providers, they highlight the risk that LLMs might draw on web content to produce answers while generating little traffic, thereby weakening the ability to monetize through advertising or subscriptions.[2] As such, online content providers may benefit from considering alternative revenue models, such as charging LLMs for access to content – which is more likely to succeed when offering high-value differentiated content. For GenAI firms, the results underscore the importance of access to high-quality content for training and, increasingly, for retrieval-augmented generation. Yet, online content providers will only be incentivized to continue producing such content if they can monetize it effectively. Our findings also speak to the structure of content creation on the web and, by extension, to regulatory and policy debates. A large share of the digital economy depends on the steady production of high-quality online content, often by smaller or specialized creators. If traffic and monetization decline due to substitution by LLMs, the incentives to produce future content might weaken, with implications for information diversity and the long-run health of the web ecosystem.

2 Related Literature

Our results contribute to four streams of research. First, our results contribute to the literature on how emerging technologies reshape media and information consumption, either by complementing or by substituting existing tools and behaviors. Digitization has raised questions regarding the extent to which accessibility of content through new technologies, including piracy, substitutes for established services (Oberholzer-Gee and Strumpf, 2007; Rob and Waldfogel, 2007; Telang and Waldfogel, 2018). Recent research suggests that news aggregators can drive traffic to news websites, particularly to outlets with fewer visitors (Calzada and Gil, 2020; Athey et al., 2021) and that snippets of songs on TikTok drive music consumption on Spotify (Bairathi et al., 2024).[3] Together, these results demonstrate that snippets of content – news or music – can plausibly serve as advertising, thereby constituting

[2]Such concerns have led Chegg, a website providing homework answers and inexpensive textbooks to students, to sue Google over the extraction of its content for AI Overviews. See `https://futurism.com/the-byte/google-lawsuit-ai-overview`, accessed April 9, 2025.

[3]Note that results by Cheng et al. (2024) and Winkler et al. (2024) instead suggest a substitutive relationship.

a complementary good (Becker and Murphy, 1993). By contrast, Seamans and Zhu (2014) show that the entry of Craigslist led to a drop of classified ad revenues among local U.S. newspapers, demonstrating how new technologies can threaten existing monetization models. Our paper builds on this research as it demonstrates that LLMs substitute for some online activities more efficiently than for others. Our results also emphasize how this shift has the potential to threaten established monetization through advertising or subscriptions.

Second, our research relates to an emerging literature on the adoption of LLMs and the effect of LLMs on visits to and activity on knowledge-sharing platforms. Fradkin (2025) presents descriptive evidence for the demand for LLMs, suggesting that new models experienced rapid initial adoption that stabilized within weeks. Humlum and Vestergaard (2025) demonstrate that use of ChatGPT is widespread, especially among younger and less-experienced workers. Chatterji et al. (2025) find that the majority of messages on ChatGPT relate to searches for practical guidance or seeking information. Burtch et al. (2024) find that the release of ChatGPT led to a significant decline in website visits to Stack Overflow and question volumes at the site (see del Rio-Chanona et al. (2024) for further evidence) though it had little impact on activity in Reddit communities. Lyu et al. (2025) demonstrate that newly created, popular Wikipedia articles whose content overlapped with ChatGPT saw a greater decline in editing and viewership after the launch of ChatGPT than dissimilar articles. The effect on content creation on such sites is often heterogeneous (Li and Kim, 2024; Shorakaei et al., 2025) though combining the use of LLMs and expert oversight can enhance content quality (Shankar and Sim, 2024). While this literature largely focused on the effect of LLMs on content creation on individual sites, we study how consumers' adoption of LLMs shifts how they interact with online content more broadly. The decrease of traditional online search we document may be one contributing factor behind the significant decline in visits to Stack Overflow that both we and Burtch et al. (2024) find. The drop in display ad exposures we document suggests that such reduced visits can have direct economic consequences for online content providers.

Third, our findings relate to a debate about the impact of AI on creator incentives. Goldberg and Lam (2025) provide causal evidence that generative AI creative goods can substitute for human-generated goods, and crowd out incumbent firms, while intensifying competition to raise variety, improve average quality, and increase overall sales. Yang and Zhang (2024) model how dynamic fair-use provisions, by lowering AI training–data acquisition costs, reshape creators' incentives over time and risk eroding long-run content supply. The fact that we find substitution for smaller but not for large websites suggests a significant amount of heterogeneity in how LLMs will affect incentives for content production in the future and in the competition between content providers and AI platforms.

Fourth, our results relate to the monetization of online content. Sun and Zhu (2013) study the adoption of monetizing blog content through advertising and find that monetization increased content quality and the share of popular content. Lambrecht and Misra (2017) demonstrate the trade-offs online content providers face when monetizing through both advertising and subscriptions. The challenges of monetizing through advertising have lead online content providers to adopt paywalls which can, however, significantly reduce visits (Chiou and Tucker, 2017), especially by heavy users (Pattabhiramaiah et al., 2019), or lead content providers to turn to freemium models where a basic version is offered for free to stimulate the demand for a more advanced paid version (Deng et al., 2023; Lee et al., 2021). For search engines, Larsen and Proserpio (2025) find that using LLMs to interpret consumer search queries can shift CPC in search advertising. Our finding that the adoption of LLMs significantly reduces traffic not only to online content sites that benefit from advertising exposures, but also to those monetizing through subscriptions suggests that multiple different revenue streams may be at risk, potentially disincentivizing the creation of online content more broadly.

3 Empirical Setting and Data

3.1 Empirical Setting

LLMs are a class of deep learning models trained on vast amounts of textual data to perform a wide range of language tasks. Their development progressed rapidly over the past decade, resulting on November 30, 2022, in the release of ChatGPT by Open AI. In the following year, multiple updates were released, including in March 2023, GPT-4. In May 2023, the paid "Browse with Bing" feature was released in Microsoft's search engine.

Until the end of 2023, our observation period, almost all available LLM versions were trained on fixed, pre-training datasets and lacked external retrieval updating or outlinks.[4] These models were employed across a wide range of personal and professional applications. Typical use cases include writing assistance and drafting documents, including digital content such as for blogs, emails or marketing content, as well as translation or support with writing of creative texts. Further, LLMs would support code-generation, code explanation and de-bugging as well as data analysis. A further context where LLMs provide value is education. Use cases include having complex topics explained (e.g., in sciences or programming), creating study aids such as flash cards or quizzes, and summarizing content. LLMs can also help students with outlines, grammar, or translations, and students using LLMs for academic cheating, such as for essay writing and take-home exams, became a significant concern. [5]

[4]As an exception, "Browse with Bing" had live-updating but was only available as a paid version between May and July 2023, and released again in October 2023, and does not appear to account for a great share of usage in our data.

[5]One survey suggests that in the academic year 2022-2023, 30% of US college students used ChatGPT for schoolwork while 46% of this group says they somewhat or very frequently used the tool, see `https://www.intelligent.com/one-third-of-college-students-used-chatgpt-for-schoolwork-during-the-2022-23-academic-year/`, accessed May 12, 2025.

3.2 Data

We use the Comscore Web-Behavior Panel dataset for the years 2022 and 2023. The data set records 1,179,088 users' detailed URL-level data for their desktop browsing. From this dataset, we observe: (1) all URL calls, and the timestamps of those URL calls, made to websites through the desktop of household panelists, (2) all search queries performed by these panelists on search engines,[6] and (3) Comscore's categorization of the accessed websites. We note that we observe not only URLs of websites displayed in the browser URL bar, but we also observe other URLs that the browser calls when displaying a website. This is important as it allows us to observe ad impressions that are shown to panelists when loading a webpage, since these impressions trigger URL calls to known display ad networks. Further, our data document the content of search queries using traditional search engines, such as when a user conducts a Google search, as search queries are recorded as part of the URL when search results are returned. It does not, however, record queries submitted to LLMs. Instead, if a user visits an LLM, we only observe the URL call, for example, to chatgpt.com, but do not observe the content of the submitted queries.

We rely on users who adopted LLMs in early December 2022 through the end of 2023.[7] The inclusion of users who adopted after our observation period allows us to have a control group of not-yet-adopters for those individuals who adopted towards the end of the period. We focus on users' online activities between November 2022 and October 2023, excluding traffic outside that one-year period. Including activity during the four weeks prior to the first cohort's LLM adoption allows us to observe pre-treatment activity.

We construct a balanced weekly panel of adopters in two steps. First, we filter for users who had at least four days with any web traffic in each month from November 2022 to October 2023 (our sample period).[8] This removes users who have effectively left the sample but also

[6]Even though we observe all search instances, we do not observe the exact query used as some private information is masked (e.g., location).

[7]We later in this section explain in detail how we define adoption.

[8]We explore robustness of our results around the sample criteria in the Web Appendix A.

allows for the possibility of holidays or extended periods of leave. This intermediate sample contains 74,940 individuals. Second, we define a user as adopting LLMs if we observe three consecutive weeks of at least one instance of LLM usage each. In our analysis, we restrict attention to users who have adopted LLMs, dropping those who never adopt. We define as the adoption date for each user the first of the three consecutive weeks of at least one instance of LLM usage.[9] This leaves us with a sample of 2041 households that adopted LLMs during the relevant time period, with households adopting LLMs between December 5th, 2022, and December 17th, 2023.

3.3 Descriptive Statistics

We aggregate our data to a weekly level. Table 1 summarizes the number of observations in our sample. We observe each of the 2041 panelists in our main sample for 52 weeks, which gives us a total of 106,132 week-user observations. Users adopt LLMs at different points during our data period. For our main sample, 64,543 observations fall into the pre-adoption period and 41,589 observations fall into the post-adoption period.

Throughout, we measure user activity by aggregating the number of URL calls made by the browser to different websites, rather than by attempting to infer time spent on each site. The latter will be affected by users' idiosyncratic online behavior, including switching between tabs or interspersing browsing with other activities, and requires ad hoc data-processing choices, such as selecting thresholds for inter-event gaps.[10] In contrast, URL-call counts provide a cleaner and less ambiguous proxy for activity.

In Table 1, we display summary statistics for the sub-samples we use in our analyses. We restrict these sub-samples to panelists who have at least 10 URL calls in the entire observation window on the relevant main dependent variable. We find that 1886 panelists had at least

[9]We explore robustness of our results to different numbers of consecutive weeks of LLM usage when defining adoption in Web Appendix A. As expected, the more consecutive weeks we require to define adoption, the smaller the sample, leading to wider confidence intervals. Additionally, as we show in Table B.1 in the Web Appendix, overall activity increases when we require more weeks of consecutive LLM usage for adoption, relative to the full sample.

[10]For an example of these, see Appendix A.7 in Greminger et al. (2023).

10 URL calls to education-related websites during our sample period.[11] For our analysis of Wikipedia, Stack Overflow, and Reddit we rely, respectively on the sample of 1634, 287, and 1488 panelists that visited each domain at least 10 times during our observation period.

As expected, the number of observations for each sub-sample fluctuates with the number of panelists.

Variable	Main	Education	Wikipedia	Stack Overflow	Reddit
Unique panelists	2,041	1,886	1,634	287	1,488
Weeks in data	52	52	52	52	52
Total observations	106,132	98,072	84,968	14,924	77,376
Pre-adoption	64,543	59,706	51,120	8,246	46,245
Post-adoption	41,589	38,366	33,848	6,678	31,131

Table 1: Size of traffic, search, and ad impressions sample

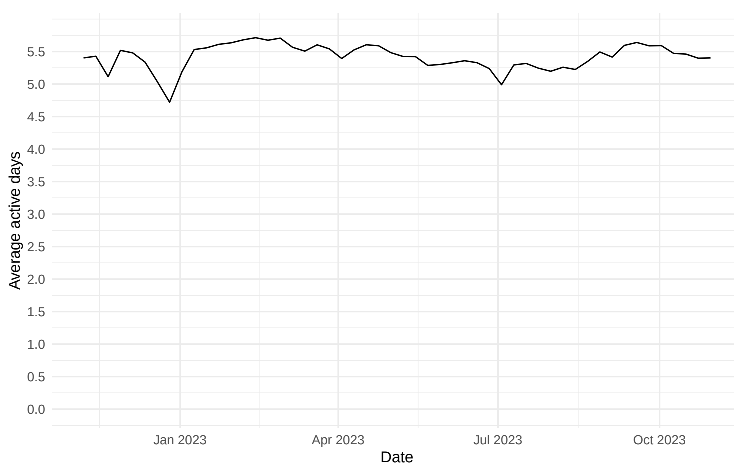

Figure 1: Average number of active days per week.

Figure 1 captures the average number of active days per week across all 2,041 users. It demonstrates that this number lies fairly consistently between 5 and 5.7, with a more pronounced drop during the winter holiday period and another drop in early July, likely a result of the 4th of July holiday. This pattern suggests that users in our data regularly use the device that contributes data to our panel.

[11]We filter for education-related websites as follows: We rely on Comscore's categorization of a website as being education-related and then manually check the top 107 domains, ranked by their share in URL calls to ensure these are indeed education-related. These domains each have at least 0.1% of URL calls in the education category and in total account for 85.5% of URL calls in the category. We remove six domains that are not education-related

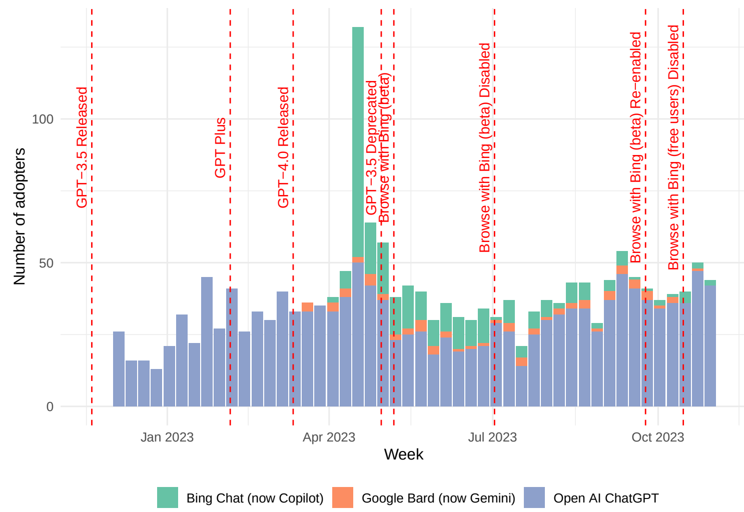

Figure 2: Adoption per week (first week of three consecutive weeks of usage).

3.3.1 LLM Adoption and Usage

We then turn to users' adoption of LLMs. Figure 2 demonstrates that users adopted LLMs throughout the entire observation period, though we find variation in the adoption of LLMs over time.[12] Our data record, on average, 35.1 calls to LLMs per user and week, post-adoption, with a standard deviation of 139.1. In Figure 3, we display the average number of URL calls to LLMs across adoption periods. As one would expect, by construction, we see zero activity the week before adoption (otherwise, adoption would occur a week earlier) and a spike of LLM usage during the period we rely on as defining an adoption event, followed by a drop and a later slight increase. We attribute the large standard errors after around 35 weeks post-adoption to the drop in the number of weekly post-adoption observations. In additional analysis, we find generally similar patterns in LLM usage across cohorts (see Web Appendix B.2).

We display the number of unit-week observations for every number of weeks since adoption across calendar weeks in Figure 4 and across adoption cohorts in Figure 5. These plots

[12]The spike in adoption in the week of April 17, 2023, corresponds to users adopting Microsoft's Bing Chat (www.bing.com/chat, now called Copilot). At that time, Microsoft removed the waiting list to access the GPT-4 powered chatbox for free (Warren, 2023), whose underlying model was not available to free users through ChatGPT.

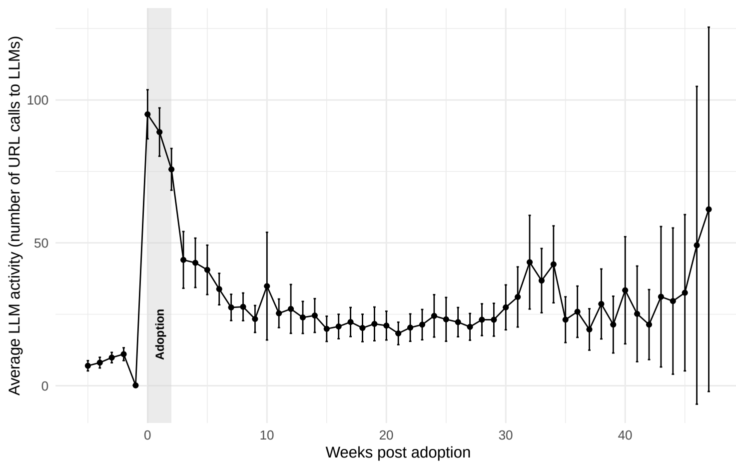

Figure 3: Average number of URL calls to LLMs per post-adoption week. Bars indicate standard errors.

illustrate the amount of data available to estimate average treatment effects on the treated at different points relative to adoption (for details on identification, see Section 4). For most post-adoption weeks in the range of approximately -35 to +35 weeks relative to the week of adoption, we observe a large number of adoption cohorts with a significant number of observations, meaning that estimates at these points average over many units that adopted at different times. This pooling across cohorts contributes to more stable and precise estimates. However, precision diminishes for estimates at longer horizons, particularly beyond +35 weeks since adoption. This is because such post-adoption periods necessarily fall toward the end of the observation window (e.g., late 2023), where fewer units have accumulated sufficient post-adoption time. As Figure 4 shows, the number of observations at these long horizons is substantially lower (visible as fewer darkly colored blocks above the horizontal line at +35), and simultaneously, there are fewer not-yet-treated units contributing to the control (i.e., fewer observations below the horizontal line at 0). This sparsity helps explain the wider confidence intervals of estimates pertaining to 35 or more weeks after adoption across most of our results. Similarly, Figure 5 demonstrates that few panelists adopted in the initial weeks when LLMs became available, further contributing to the low numbers of adopters for whom we observe 35 or more weeks post-adoption.

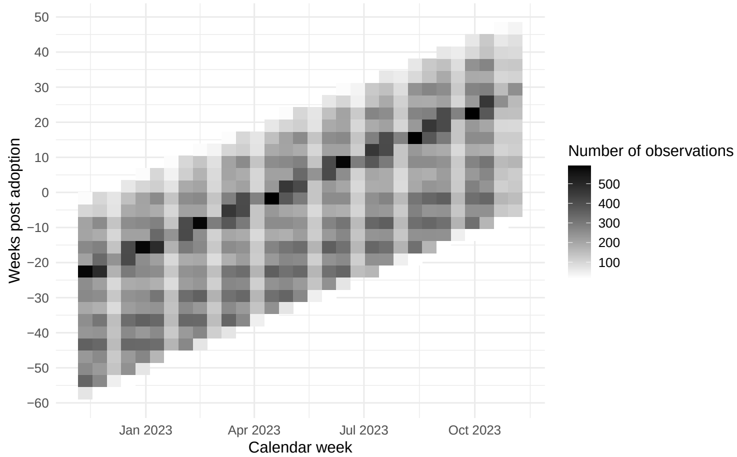

Figure 4: Number of observations per calendar week and post-adoption week.

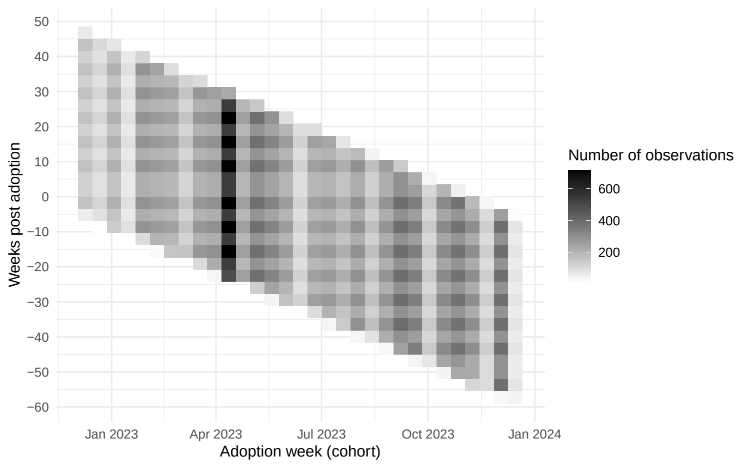

Figure 5: Number of observations per adoption cohort and post-adoption week.

3.3.2 Search Queries, Website Browsing, and Ad Exposures

Table 2 summarizes the data descriptives of our user-level weekly panel based on users' pre-adoption periods. Users made 33 search queries per week, 18 of which were through Google, other search engines include Bing and Yahoo. Among those, four searches used at least one of the following question terms: how, what, where, when, which, who, why.

To identify navigational searches (Blake et al., 2015), we identify the top 2000 websites that users navigated in our data. We then parse out Google, Bing, and Yahoo searches that use as a search term only the name of one of those websites, e.g., "reddit".[13] On

[13] We eliminate ambiguous terms such as "weather".

average, users made one such navigational search a week and two searches a week that use a navigational term and one or more other terms. Further, users make about five short searches (two or fewer words) and about seven long searches (six or more words). These correspond, respectively, to the first and fourth quartiles of word counts in search queries.

To quantify web browsing, we exclude searches, URL calls to email, retail and news sites that will serve as a control for baseline activity, URL calls that we identify as advertising impressions as well as background traffic.[14] In total, we find 4325.1 URL calls per user and week of which 2928.0 can be attributed to the 500 websites that receive the most amount of traffic, whereas 1397.1 come from all remaining websites (analogously, 3219.9 for the 1000 websites with the most traffic and 1105.2 for all others). We rank order websites in terms of the amount of traffic they receive and subgroup them into quartiles so each quartile accounts for approximately the same amount of traffic.[15] We report in the table the mean number of URL calls for each quartile.

For any website in our sample, we identify the traffic referred to from a search engine. A consumer visiting a search engine and clicking on a link counts as one referral instance, independently of the number of pageviews, and thus URL calls, on that website. On average, our data record 21.7 referrals per user and week. We rank websites in terms of the number of referrals they received and report the number of referrals per quartile.[16]

We identify advertising exposure through URL calls associated with the loading and serving of ads. The domains we can identify as associated with ad loads in our data are adservice.google.com, imasdk.googleapis.com, and ads.yieldmo.com. We are unable to capture exposure to other ads, such as on social media. In our data, users are exposed to 212 ads per week, of which the majority are Google display ads (161), a much smaller portion are

[14] We classify each URL host (e.g., "mail.google.com") as foreground or background based on manual classification of the top 500 sites and the ratio of search-referred traffic to total traffic for sites outside the top 500. Hosts with extremely low search-referral ratios — typically system or service infrastructure domains — are identified as background and removed (see Web Appendix C.1 for details)

[15] This rank order and the subsequent stratification in sub-groups is based on the entire time period for all 1,179,088 Comscore panelists in order to avoid any potential selection issues from focusing on adopters only.

[16] Again, this rank order and the subsequent stratification in sub-groups is based on the entire time period for all 1,179,088

Category	Variable	Mean	SD	5%	25%	50%	75%	95%
Main								
Search queries	All search	32.7	84.1	0	0	9	34.0	134.0
	Google	17.7	41.2	0	0	1	17.0	88.0
	Questions	3.8	12.0	0	0	0	3.0	18.0
	Navig. only	0.8	3.0	0	0	0	1.0	4.0
	Navig. + other	1.5	6.6	0	0	0	1.0	7.0
	Long	6.6	20.5	0	0	1	6.0	30.0
	Short	5.0	16.4	0	0	1	5.0	20.0
Website browsing[a]	All websites	4,325.1	14,576.4	0	402	1,495	4,078.0	16,037.8
	Top 500 websites	2,928.0	13,119.8	0	241	900	2,572.5	10,503.9
	non-Top 500 websites	1,397.1	4,246.1	0	64	368	1,246.5	5,764.5
	Top 1,000 websites	3,219.9	13,477.5	0	267	985	2,832.0	11,811.6
	non-Top 1,000 websites	1,105.2	3,369.4	0	47	297	1,025.0	4,652.9
	Top 25%	831.6	4,396.4	0	48	249	697.0	3,201.8
	25%-50%	943.5	3,981.3	0	46	222	783.0	3,677.0
	50%-75%	1,395.7	11,417.0	0	45	223	775.0	4,767.9
	Bottom 25%	1,154.3	3,435.1	0	49	309	1,074.0	4,866.9
Referred traffic	All referred	21.7	101.3	0	0	6	22.0	81.0
	Top 25%	4.7	28.6	0	0	0	3.0	18.0
	25%-50%	6.9	80.2	0	0	1	4.0	20.0
	50%-75%	5.2	31.7	0	0	1	5.0	21.0
	Bottom 25%	4.9	10.5	0	0	1	5.0	23.0
Ad exposures	All ads [Google, Yieldmo]	211.6	1,516.8	0	1	21	103.0	726.9
	Google ads: display	161.1	1,399.7	0	0	8	58.0	533.0
	Google ads: search	14.7	35.2	0	0	0	14.0	74.0
	Google ads: video	4.0	41.0	0	0	0	0.0	9.0
	Yieldmo	31.9	375.0	0	0	0	6.0	103.0
Control	All control	648.9	2,310.4	0	2	73	459.0	2,968.0
	Email	378.4	1,273.9	0	0	20	257.0	1,822.9
	Retail	208.1	1,515.4	0	0	3	53.0	723.0
	News	62.5	818.6	0	0	0	4.0	138.0
Education								
All education	URL Calls	169.7	712.7	0	0	0	44.8	950.8
	Variety of URL calls	1.3	3.3	0	0	0	2.0	6.0
By subcategory	Learning management system	62.2	282.6	0	0	0	0.0	357.0
	Online learning platform	55.8	550.5	0	0	0	0.0	177.0
User-Generated Content Platforms								
Knowledge-sharing platforms	Wikipedia	13.9	322.7	0	0	0	0.0	23.0
	Stack Overflow	0.9	5.1	0	0	0	0.0	5.0
	Reddit	14.4	126.5	0	0	0	1.0	38.0

[a] We exclude background URL traffic. For more details, see Web Appendix C.1. We also exclude search traffic as well as URL calls to email, retail and news domains which are listed separately in this table.

Table 2: Summary statistics of dependent variables and controls based on users' pre-adoption period

Google search (15) or video (4) ads. On average, a user is served 32 ads from Yieldmo.

To identify email, retail, and news activity, which later serves as a control in our estimation, we rely on the corresponding Comscore classifications.[17] On average, a user made 649 weekly URL calls related to email, retail, and news, most of which relate to email and retail.

[17] We follow Comscore's classification but remove subcategories defined by Comscore as "Other".

We turn to user activity on education-related websites and user-generated content platforms. The 1886 users who visit an education-related website, as defined by Comscore, at least 10 times during our observation period make, on average during a week, 170 URL calls to these domains and visit about one domain per week. We identify among the top 107 domains in terms of their share in URL calls (i.e. those that each have at least 0.1% of URL calls in the education category) learning management systems and online learning platforms. Learning management systems provide digital infrastructure such as Canvas or Blackboard that is populated by educational institutions or teachers. Online learning platforms provide content and target educational institutions who purchase access for their learners or learners directly.[18]

Table 2 further summarizes average behavior on user-generated content platforms, including knowledge-sharing platforms such as Wikipedia, Stack Overflow, and Reddit. These summaries reflect the pre-adoption behavior of the subset panelists who visit each respective platform at least 10 times.

4 Estimation and Identification

In choosing our estimation approach, we aim to address key challenges in identifying the effect of LLM adoption on browsing behavior in a staggered setting where different cohorts of users adopt at different points in time. First, we need to account for the general level of users' online activity so we do not find a spurious correlation between LLM adoption and other types of activity of interest. We use a user's level of email, retail, and news activity – which is not affected by LLM adoption – as an activity control. Second, to account for week-specific shocks to user behavior, our estimation requires a control group. For each cohort, we use as controls other cohorts that have not yet adopted. Because never-treated units may be systematically different, we restrict our sample to eventually treated individuals and exploit

[18]They include Edgenuity which offers B2B subscriptions to educational institutions as well as Duolingo, which markets to learners directly, and Cengage, which has both an offering for learners and for institutions.

plausibly exogenous variation in their week of adoption.

4.1 Estimation Approach

Standard two-way fixed-effects (TWFE) estimators face well-known problems in staggered adoption settings with heterogeneous treatment effects (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Borusyak et al., 2024). Specifically, TWFE estimators can yield biased dynamic effects in the presence of treatment-effect heterogeneity across cohorts and over time because they use already-treated units as controls for newly-treated units, creating comparisons that contaminate the estimates when treatment effects vary across cohorts and time.

We follow Callaway and Sant'Anna (2021) to estimate cohort- and time-specific average treatment effects on the treated ($\mathrm{ATT}_{g,t}$), where $t = 1, \ldots, T$ is the calendar week and $g \in \{1, \ldots, T\}$ is the adoption week, using not-yet-treated units as controls. This staggered difference-in-difference estimation approach explicitly accounts for treatment effect heterogeneity across time and cohorts while avoiding bias from comparisons against already-treated units.

Let $Y_{i,t}$ denote individual i's outcome in week t, $D_{it} = 1$ if i has adopted by week t and cohort indicator $G_{i,g} = 1$ if i adopted in week g. We also denote $Y_{i,t}(g)$ as the potential outcomes for individual i at week t if they adopted at period g. The cohort- and time-specific average treatment effect on the treated is defined by

$$
\begin{aligned}
\mathrm{ATT}_{g,t} &= \mathbb{E}[Y_{i,t}(g) - Y_{i,t}(0) \mid G_{i,g} = 1] \\
&= \mathbb{E}[Y_{i,t} - Y_{i,g-1} \mid G_{i,g} = 1] - \mathbb{E}[Y_{i,t} - Y_{i,g-1} \mid D_{i,t} = 0, G_{i,g} = 0].
\end{aligned}
$$

This captures the causal effect at calendar week t for cohort g, by comparing the treated cohort to not-yet-treated units over the same time period. Following Callaway and Sant'Anna (2021) we use the pre-adoption week $(g-1)$ as the pre-adoption period. We show robustness to using other pre-adoption weeks in Web Appendix D.1.

Our main outcome of interest is the average treatment effect by event time, denoted ATT(e), where $e = t - g$ indexes weeks relative to the adoption week. This parameter aggregates the cohort- and time-specific effects $\text{ATT}_{g,t}$ across all adopting cohorts, aligning units by their relative time since adoption rather than by calendar time. Formally, ATT(e), defined by

$$\text{ATT}(e) = \sum_{g} \mathbf{1}\{g + e \leq T\} \cdot P(G = g | G + e \leq T) \cdot \text{ATT}_{g,g+e},$$

is obtained by averaging $\text{ATT}_{g,g+e}$ across all cohorts g for which the event time e is observed, weighting each cohort by its share among the cohorts observed at that event time. This event-study estimand provides the dynamic ATT while respecting potential heterogeneity in treatment effects across adoption cohorts and over time.

4.2 Identification

To ensure our results are not biased by users' overall online activity, we control for URL calls to email, retail, and news websites, categories we expect to be unaffected by LLM adoption. Absent these controls, we might find a spurious positive correlation between LLM adoption and the outcome of interest due to fluctuations in internet activity unrelated to LLM adoption. Figure 6 and Table 3 show that LLM adoption had no effect on the covariates, validating their use as controls. The brief increase in weeks 0-2 reflects our adoption definition requiring three consecutive weeks of LLM activity, which is related to general desktop usage.

To account for week-specific activity shocks common to all users, we include not-yet-treated units as a control group. Our data satisfy the standard overlap condition requiring both treated and untreated units for each calendar time and relative period around treatment.

Further, our identification strategy relies on two key assumptions: conditional parallel trends and no anticipation.

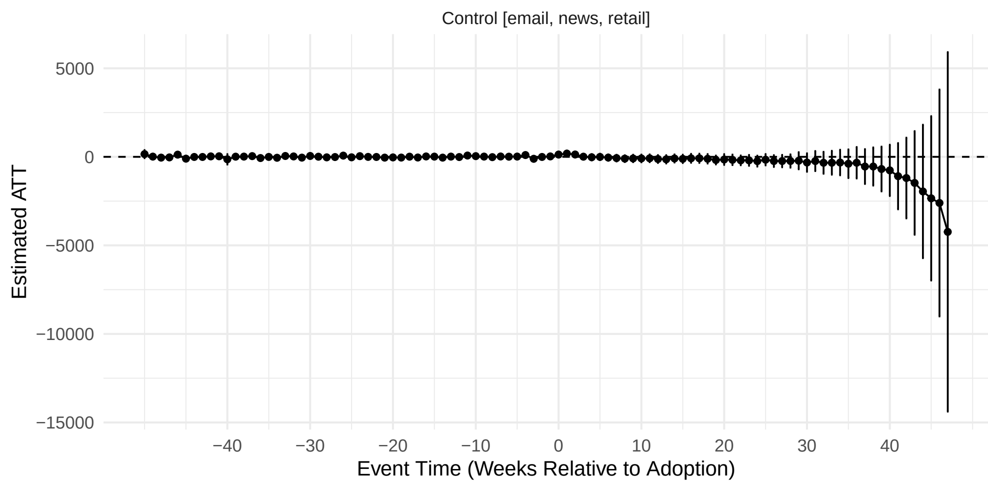

Figure 6: Staggered DiD ATT on the weekly number of URL calls in the control websites (email, news, and retail).

	Dependent variable:
	Control [email, news, retail]
ATT (weeks 0-2)	148.064*
	(81.733)
ATT (weeks 3-19)	−83.958
	(114.006)
ATT (weeks 20-47)	−777.914
	(825.146)
Pre-adoption avg.	648.945
Panelists	2041
Weeks	52
Observations	106,132
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3: Effects on activity on control websites (email, news, and retail).

Conditional Parallel Trends: For all $t \geq g$ and all groups g and g' with $g \neq g'$,

$$\mathbb{E}[Y_{i,t}(0) - Y_{i,g-1}(0)|G_{i,g} = 1, X_i] = \mathbb{E}[Y_{i,t}(0) - Y_{i,g-1}(0)|G_{i,g'} = 1, X_i].$$

This assumption requires that, in the absence of treatment, all cohorts would have experienced parallel trends in outcomes, conditional on observables X_i. We find no evidence for violations of parallel trends, which under the staggered difference-in-difference settings would manifest as systematically positive or systematically negative pre-period effects (Roth, 2024).

No Anticipation: For all $t < g$ and all $g' \geq g$,

$$\mathbb{E}[Y_{i,t}(g)] = \mathbb{E}[Y_{i,t}(0)]$$

This assumption requires that units do not change their behavior in anticipation of future treatment. Pre-treatment outcomes must be unaffected by knowledge of future adoption timing. The assumption is violated if agents can foresee treatment and adjust their behavior beforehand, such as if individuals who have accepted a new job alter their spending before their job officially begins. In our setting, this type of anticipatory behavior is unlikely. LLM adoption does not generally involve a clear future event that would prompt users to modify their browsing patterns weeks in advance. Although we cannot directly test the no-anticipation assumption, our empirical estimates show a delayed rather than immediate response in online activity after LLM adoption, which is inconsistent with meaningful pre-treatment adjustments. Further, our results are robust to varying the length of the anticipation window—that is, the number of weeks before treatment during which individuals may change their behavior (see Web Appendix A).

More generally, our approach is robust to several potential threats to causal identification:

- **Calendar Time Shocks:** Any time-specific unobservables (e.g., holidays, seasonality) are differenced out when computing each cohort- and time-specific $\text{ATT}_{g,t}$.

- **Level Differences Across Cohorts:** If early adopters have systematically higher or lower outcome levels due to unobservables (e.g., more tech-savvy users adopt earlier), these differences are differenced out as long as the unobservables do not affect outcome trends because each $\text{ATT}_{g,t}$ is estimated using a two-by-two difference-in-difference setup. The parallel trends assumption accommodates level differences across units while requiring similar growth patterns across cohorts.

- **Heterogeneous Treatment Effects:** Our method explicitly accommodates users who adopt earlier experiencing different treatment effects than later adopters by non-parametrically estimating separate $\text{ATT}_{g,t}$ parameters for each cohort. Unlike TWFE, we do not impose homogeneous treatment effects across cohorts (or time). We later show that cohorts do not meaningfully differ in their treatment effects over time (Web Appendix E).

- **Declining Activity Over Time:** A concern might be that panelists in our sample become less active online over time, confounding our estimates. However, we provide evidence that users remain active throughout our observation window (Figure 1). Additionally, we control for plausibly unaffected behaviors (retail, news, email) to account for general fluctuations in baseline internet activity. For example, a user who starts a new job may simultaneously adopt a new technology like LLMs and change their baseline search and internet activity. Our controls would account for such variation. Moreover, the patterns in Figure 6 and Table 3 confirm that these behaviors remain largely unaffected by LLM adoption.

In sum, our identification is robust to unobservables that correlate with adoption timing as long as they do not generate differential trends in the absence of treatment. In addition, given the magnitude and consistency of the estimated effects, together with the range of potential threats to identification that our design explicitly addresses and the robustness checks we have performed, it appears unlikely that unobserved confounders that are both time-varying and cohort-specific could account for the patterns we document.

4.3 Inference

Our estimation approach achieves robustness to heterogeneous effects across cohorts and time by estimating effects non-parametrically through comparisons of each focal cohort to not-yet-treated cohorts. However, this additional robustness comes at the cost of statistical power relative to TWFE methods. Each $\text{ATT}_{g,t}$ estimate uses only a subset of the data—individuals in cohort g and not-yet-treated individuals from other cohorts, rather than pooling all observations. This means that sample size interpretation differs fundamentally from TWFE regressions. While our full dataset contains over 100,000 observations across 52 calendar weeks and 48 adoption cohorts, these are distributed across 2448 individual $\text{ATT}_{g,t}$ parameters. Consequently, statistical power is lower than in pooled regressions, particularly for long-term effects (40+ weeks post-adoption), where fewer observations are available per estimate. To partially mitigate this loss of precision, we include users' baseline traffic to email, retail, and news websites as covariates, which, in addition to controlling for potential activity bias, partially improve estimation efficiency.

Estimation proceeds via the standard approach recommended by Callaway and Sant'Anna (2021), nonparametric residualization of outcomes on covariates, aggregation of group-time ATTs, and cluster-robust inference. Standard errors are computed via influence-function linearization, clustered at the individual level. Across several specifications, we use the inverse hyperbolic sine transformation (Burbidge et al., 1988) to handle skewness in the control covariates.

5 The Effect of LLM Adoption on Online Search, Website Browsing and Advertising Exposures

To understand how LLM adoption reshapes the way users acquire information online, we first focus on its impact on search behavior, often a user's first step in seeking information that may affect subsequent online activities. Understanding the effect of LLM adoption on traditional online search is also informative on the degree of competition between LLMs and traditional search engines.

We then assess how LLM adoption shifts browsing patterns across the internet, distinguishing between effects on large and small websites and traffic that originates from search engines. We direct special attention to two categories that have received significant attention: education-related websites and user-generated content platforms. Together, these analyses provide insight into how LLM adoption may influence the economic sustainability of online platforms and the concentration of user attention across the web.

Finally, we examine whether shifts in information acquisition lead to corresponding changes in advertising exposure. Such changes would affect publishers' and search engines' ability to monetize content, retailers' ability to reach consumers, and ways in which users obtain product information across the web.

Throughout this section, we plot ATT(e), the average treatment effect on the treated e weeks since adoption, to show the dynamic response over event time. For statistical inference, we aggregate these effects and focus on the longer-run impact measured 20 or more weeks after adoption.

5.1 Online Search

Figure 7 and Table 4 summarize the effect on overall online search activity. Figure 7 (a) plots weekly estimates relative to adoption, controlling for users' email, retail, and news activity. Because activity is measured from November 2022 and adoption begins in December 2022, we observe treatment effects for up to 47 weeks post-adoption. The figure shows that search activity briefly spikes after LLM adoption before later declining slowly but steadily. The initial spike aligns with our earlier findings on control activity and likely reflects heightened digital activity stemming from our definition of adoption – since adoption requires LLM use in three consecutive weeks, users must by definition be digitally active. In addition, our controls may not fully account for short-term deviations from typical behavior, such as on days when activity shifts to tasks not captured by control measures. This pattern could imply our estimates reflect an upper bound of the true effect. Importantly, however, Figure 7 (a) shows that search activity begins to decline only around 20 weeks post-adoption, not immediately. Because the effect of interest occurs well after the adoption period, elevated activity at adoption is unlikely to inflate the estimated effect. More broadly, the temporal pattern in browsing activity suggests that users gradually learn to substitute online search with LLMs over time.

	Dependent variable:				
	All search				
	(1)	(2)	(3)	(4)	(5)
ATT (weeks 00-02)	18.374***	17.585***	18.284***	18.408***	18.148***
	(1.301)	(1.264)	(1.270)	(1.299)	(1.264)
ATT (weeks 03-19)	−0.269	−1.943	−0.289	−0.319	−0.792
	(1.834)	(1.840)	(1.796)	(1.843)	(1.784)
ATT (weeks 20-47)	−7.079**	−9.735***	−7.174**	−7.395**	−7.601**
	(3.517)	(3.732)	(3.509)	(3.764)	(3.664)
Pre-adoption avg.	32.669	32.669	32.669	32.669	32.669
Controls					
Retail	Y		Y		
News	Y			Y	
Email	Y				Y
Panelists	2041	2041	2041	2041	2041
Weeks	52	52	52	52	52
Observations	106,132	106,132	106,132	106,132	106,132
Note:			*p<0.1; **p<0.05; ***p<0.01		

Table 4: Effect on all search

Column (1) of Table 4 reports the estimates for the ATT binned into three time windows, weeks 0 to 2, weeks 3 to 19 and weeks 20 to 47. Consistent with Figure 7 (a), we see a positive effect for weeks 0 to 2. We also see an insignificant effect for weeks 3 to 19. Our main estimate of interest is the significant drop that we find for weeks 20 to 47 post adoption which mirrors Figure 7.[19] Here, the estimated drop of 7.1 searches represents a 21.7% decrease relative to the pre-adoption average number of searches of 32.7 and suggests that, overall, users substitute LLMs for traditional search engines. While the estimation in Column (1) includes our full set of controls to account for potential activity bias, we show in Column (2) results without controls. We find a similar pattern, though a somewhat lower point estimate. As further robustness checks, we account in each of Columns (3) to (5) for only one of category control activity (retail, news or email). The results continue to hold.

The fact that the decline in search behavior emerges only around 20 weeks after adoption provides an additional robustness argument. Any unobserved factor capable of generating

[19] For convenience, we focus in all future tables on the ATT estimate for weeks 20-47 and drop the remaining estimates from all other tables.

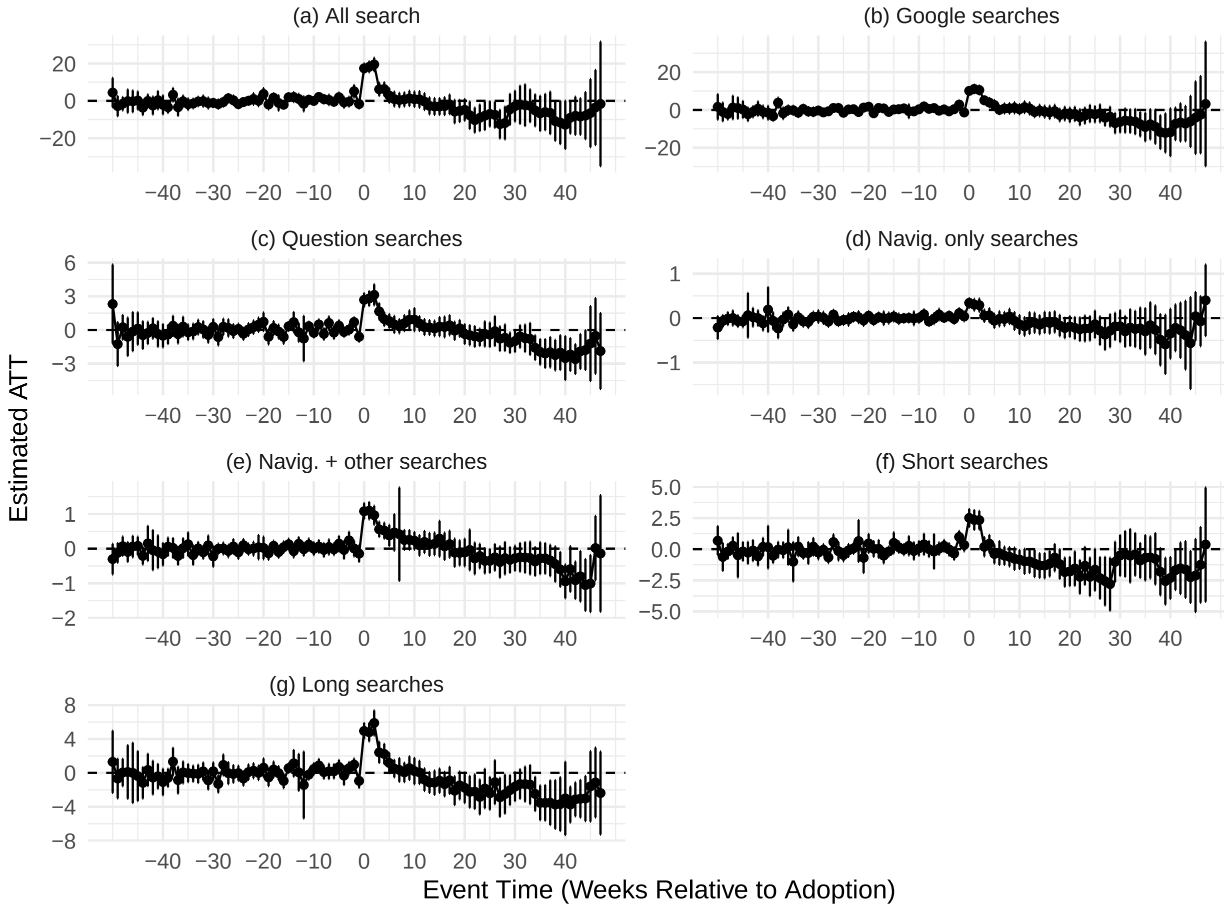

Figure 7: Staggered DiD ATT on the weekly number of total searches and the number of searches by category.

the pattern we document would need to not only correlate with adoption timing but also leave behavior unchanged for approximately five months before producing a sizable drop. Such delayed-onset confounders are difficult to reconcile with typical self-selection concerns. This interpretation aligns with our identification discussion, where we emphasize that unobservables must generate cohort-specific differential trends in the absence of treatment; the timing and consistency of our estimates make such an explanation unlikely.

Figure 7 (a) also shows an increase in the mean estimate and an increase in noise towards the end of our data period, a pattern we consistently find throughout our estimations. We attribute these patterns to two reasons. First, users for whom we observe 35 or more weeks following adoption are by definition early adopters who adopted in late 2022 (see also

Figure 5). There are few early adopters and at the same time there is a smaller number of later adopters we can match them to (see also our discussion in Section 3.3 about the lack of precision for ATTs after 35 weeks since adoption). Second, since we are focusing on early adopters, these late observations capture browsing behavior only towards the very end of our data period, around late November 2023 (see also Figure 4). As such, while estimates for other data points are averaged across many different instances of calendar time, these estimates are more likely to reflect a change in browsing behavior due to seasonalities. For example, if both treated and not-yet-treated individuals decrease activity during Thanksgiving and we only observed the Thanksgiving week for that particular post-adoption period, the estimated effects may capture no differences between treated and control units. Note that this is not a general concern but only applies to effects measured 35 weeks or more after adoption, where we have less variation on adoption timing for both control and treated units.

We next analyze different types of search. Focusing only on searches conducted using Google, Figure 7 (b) and Column (2) of Table 5 show similar patterns as for overall search. Here, we find a 31.5% drop in searches, again suggesting a large degree of substitution. (For convenience, Column (1), Table 5, reiterates the previous results for all searches.)

	Dependent variable:						
	All search	Google	Questions	Navig. only	Navig. + other	Short	Long
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ATT (weeks: 20-47)	-7.079^{**}	-5.568^{*}	-1.256^{***}	-0.242	-0.407^{**}	-1.467^{**}	-2.466^{***}
	(3.517)	(3.224)	(0.469)	(0.160)	(0.170)	(0.668)	(0.795)
Pre-adoption avg.	32.669	17.680	3.779	0.800	1.500	5.030	6.630
Panelists	2041	2041	2041	2041	2041	2041	2041
Weeks	52	52	52	52	52	52	52
Observations	106,132	106,132	106,132	106,132	106,132	106,132	106,132
Note:							*p<0.1; **p<0.05; ***p<0.01

Table 5: Effect broken down by type of search

We then explore whether different types of searches may be differentially affected. Online searches include both informational queries and simple navigational searches (Blake et al., 2015). Figure 7 (c) and Column (3) of Table 5 focus on searches using a question word. We again find a strong decline for these more complex queries, suggesting that users find LLMs a useful substitute. By contrast, Figure 7 (d) and Column (4) of Table 5, demonstrate that searches that use only a navigational term (e.g., 'reddit') do not drop – a result which is plausible given that at the time of our data LLMs did not provide outlinks so would be unlikely to substitute for navigational queries. Interestingly, Figure 7 (e) and Column (5) of Table 5 demonstrate that LLM adoption leads to a drop in searches that use both a navigational and one or more other terms (e.g., 'reddit best guitar'). We further differentiate based on query length as a proxy for the complexity of searches. Figures 7 (f) and (g) as well as Columns (6) and (7) of Table 5 demonstrate a clear drop in both short searches that include two or less words and long searches that consist of seven or more words.

Robustness We conduct a series of robustness checks. First, as mentioned in Section 3.2 (see also Web Appendix A), our results are largely robust to other sample criteria and definitions of LLM adoption with slight deviations in the expected direction: smaller samples lead to wider confidence intervals. Second, our estimates are largely consistent with those obtained from other model specifications, including TWFE and Poisson Pseudo-Maximum Likelihood regression (see Web Appendix D.2). Third, although our estimator is robust to cohort heterogeneity, one might be concerned that early adopters differ systematically from other users, raising the possibility that the ATT estimates partly reflect the behavior of this specific group rather than the effect of LLM adoption itself. We address this by re–estimating the model after excluding early adopters and show in Web Appendix D.3 estimates that are not statistically different from our main results, indicating that the findings are not driven by this group. Finally, one might be concerned that rather than capturing a causal effect of LLM adoption, our results are an artifact of calendar-specific trends that are not fully differenced out. It is not clear why such a pattern should occur because of our difference-in-differences estimation, the use of controls, and because the effect of interest sets in only about 20 weeks after adoption. Nonetheless, we conduct a placebo test in which we iteratively randomize adoption weeks across users, re-estimate the ATT for each replication, and plot the resulting distribution of placebo estimates (see Web Appendix D.4). The placebo treatment shows no effect on any outcome, reinforcing that our results are not driven by by such evolving patterns.

For brevity and clarity of exposition, we conduct analogous robustness checks for the remaining outcomes presented in subsequent sections but refrain from discussing, with the exception of an additional robustness exercise for the heterogeneous effects on ad impressions related to our classification of retail activity. The full set of robustness analyses is documented in Web Appendices A and D.

5.2 Website Browsing

We turn to estimate the overall effect of LLM adoption on website traffic outside of search. We first explore the effects on total traffic and traffic from search engines, focusing specifically on traffic received by large versus small websites. We then narrow in on two categories that have received particular media (The Economist, 2024) and academic attention: education where content explanation and summary matter, tasks performed easily by LLMs (Chatterji et al., 2025, show that 10% of ChatGPT messages relate to tutoring or teaching), as well as user-generated content, where there is conflicting evidence on the role of LLMs (Burtch et al., 2024).

5.2.1 Total and Referred Traffic

We estimate the overall effect of LLM adoption on the total number of URL calls.

Figure 8 (a) plots the individual weekly estimates over time. The corresponding estimate in Table 6, Column (1), shows no significant decline in URL calls. We then differentiate between the 500 websites that received the largest amount of traffic and all other websites over the entire observation period using all adopter and non-adopter panelists. We do not

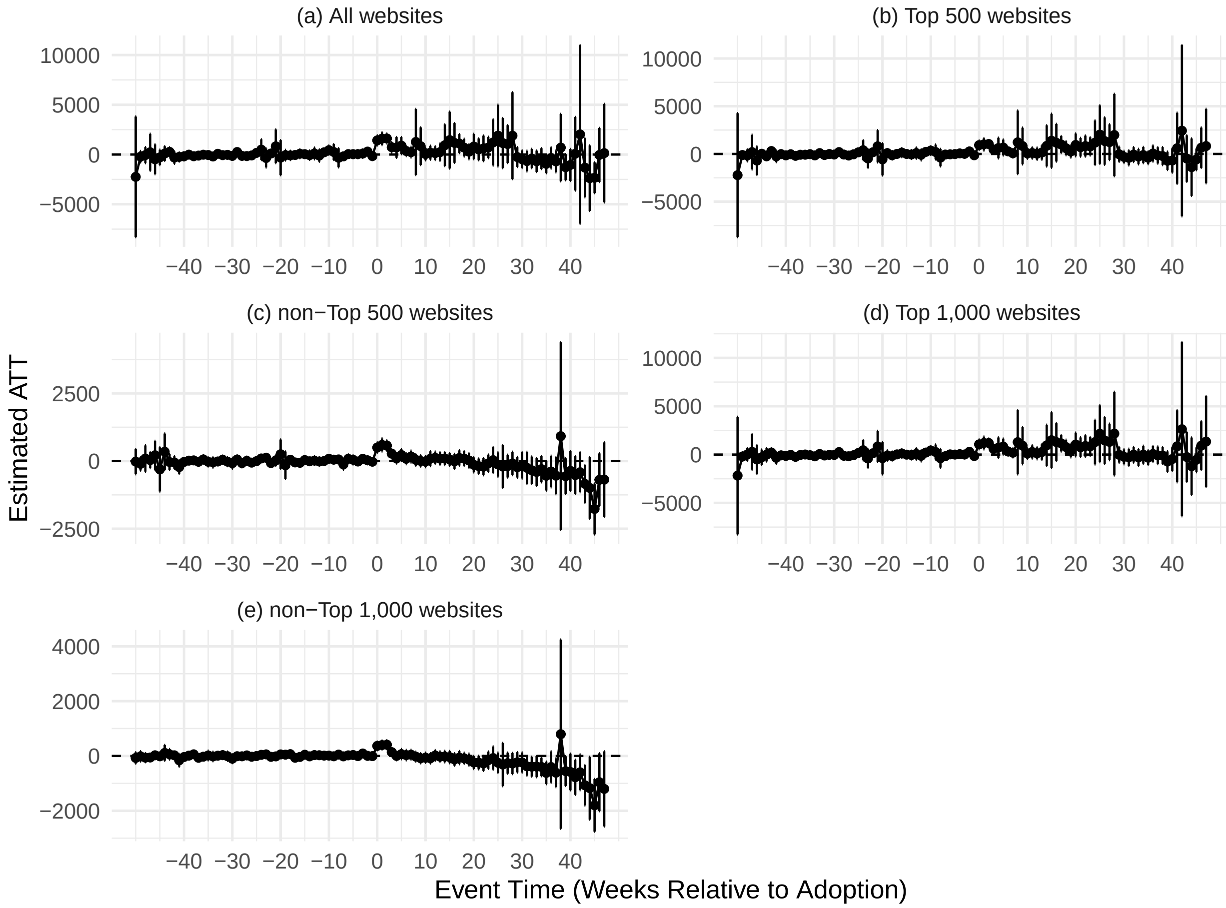

Figure 8: Staggered DiD ATT on weekly number of URL calls for top websites and all other websites.

	Dependent variable:				
		Top 500 websites		Top 1,000 websites	
	All websites	Top	non-Top	Top	non-Top
	(1)	(2)	(3)	(4)	(5)
ATT (weeks: 20-47)	−17.752	349.046	−366.798*	470.449	−488.201**
	(561.734)	(477.852)	(217.689)	(490.040)	(212.767)
Pre-adoption avg.	4,325.075	2,927.971	1,397.104	3,219.854	1,105.221
Panelists	2041	2041	2041	2041	2041
Weeks	52	52	52	52	52
Observations	106,132	106,132	106,132	106,132	106,132
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 6: Effect on all traffic: number of URL calls for top websites and all other websites

find a drop in URL calls for the largest 500 websites (see Figure 8 (b) and Table 6, Column (2)). By contrast, URL calls to websites not among the top 500 drop significantly after

	Dependent variable:			
	Top 25%	25%-50%	50%-75%	Bottom 25%
	(1)	(2)	(3)	(4)
ATT (weeks: 20-47)	102.276	−63.976	426.258	−482.310**
	(103.300)	(177.949)	(395.310)	(211.224)
Pre-adoption avg.	831.576	943.488	1,395.746	1,154.265
Websites	7	39	821	4,532,215
Panelists	2041	2041	2041	2041
Weeks	52	52	52	52
Observations	106,132	106,132	106,132	106,132
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 7: Effect on all traffic: number of URL calls by quartile of traffic

around week 20 (see Figure 8 (c) and Table 6, Column (3)). The estimated coefficient of -366.8 for the drop in activity for weeks 20 to 47, relative to the pre-adoption mean of 1397.1, implies that at the mean, visits to smaller websites dropped by 26%. To ensure the patterns are not sensitive to the specific cutoff we chose, we conduct the same analysis for the top 1000 websites by traffic, relative to all others. We find similar results (see Figures 8 (d) and (e) and Table 6, Columns (4) and (5)).

To further validate the differential effects on large versus small websites, we rank-order websites based on the number of URL calls they received and estimate separate regressions for each quartile of traffic. Columns (1) to (3) in Table 7 show no significant effect of LLM adoption on traffic for the first, second, and third quartile, which account for 7, 39 and 821 websites, respectively, mirroring Table 6, Column (2). In contrast, Column (4) demonstrates a significant drop in traffic for the remaining websites that together account for the bottom quartile of traffic, reflecting Table 6, Column (3). We find that, collectively, these websites suffer a drop of 41.8% of total traffic.

Taken together, these results indicate that large websites remain largely insulated from LLM-induced changes in user behavior, whereas smaller websites experience notable declines in visits. Smaller sites often depend more heavily on discovery through search engines, rather than on direct navigation, and these patterns suggest that LLMs may disproportionately reduce such traffic. We note that we measure an average effect and are unable to pin down whether high-quality sites offering differentiated niche content and lower-quality sites with little original information are impacted to the same degree.

	Dependent variable:				
	All referred	Top 25%	25%-50%	50%-75%	Bottom 25%
	(1)	(2)	(3)	(4)	(5)
ATT (weeks: 20-47)	−0.078	0.142	2.204	−1.403	−1.021*
	(5.513)	(0.904)	(5.348)	(0.983)	(0.566)
Pre-adoption avg.	21.723	4.749	6.869	5.158	4.947
Panelists	2041	2041	2041	2041	2041
Weeks	52	52	52	52	52
Observations	106,132	106,132	106,132	106,132	106,132
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 8: Effect on website traffic referred from search engines. All traffic vs. traffic by website popularity.

To further examine this mechanism, we focus on URL calls generated directly by clicks on search results. In Table 8, we show the effects on traffic referred from search engines. Recall that in referred traffic, we only account for the first URL call to a website during a sequence of page views on that website. Column (1) shows there is no significant drop in the total amount of traffic referred to websites from search engines. We then rank order the data by the number of referred URL calls to each website and then stratify our estimation by quartiles of referred URL calls. Columns (2) to (4) indicate no significant drop in referrals for websites that receive the top 75% of search engine referrals. By contrast, Column (5) indicates a significant drop in referrals for the websites that receive the lowest number of referrals. At the mean, this amounts to a drop of 20.6%.

Together, the results in Tables 6, 7 and 8 demonstrate that while URL calls to very large websites do not drop significantly, there is a large and statistically significant drop in URL calls to smaller websites, suggesting that LLMs can serve as a substitute. This result may have important implications to how these small websites may sustainably monetize the content they generate. We further explore these implications in our conclusions.

5.2.2 Effect on Specific Domain Types

We next examine the effect of LLM adoption on two particular kinds of online content that have drawn disproportionate attention by media and academics alike: education–related websites and user–generated content (UGC) platforms.

Education-Related Websites Education is a natural test case for substitution. Students were among the earliest adopters of LLMs,[20] and many educational tasks (summarization, drafting, practice problems) are directly automatable. We therefore evaluate effects among the 1886 panelists who visited at least 10 times education–related sites during our sample period.

	Dependent variable:			
	All education	Variety of URL calls	Learning management system	Online learning platform
	(1)	(2)	(3)	(4)
ATT (weeks: 20-47)	−151.532**	−0.520***	−23.124*	−123.794**
	(65.729)	(0.166)	(14.049)	(54.269)
Pre-adoption avg.	169.688	1.290	62.199	55.785
Panelists	1886	1886	1886	1886
Weeks	52	52	52	52
Observations	98,072	98,072	98,072	98,072
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 9: Effect on Education-related websites traffic

Column (1) uses traffic to education-related websites as the dependent variable and shows a decline in overall category traffic. Column (2) examines the number of unique domains visited within the education category by a user and indicates a reduction in the variety of educational sites users access. Columns (3) and (4) show a decrease in activity on learning management systems[21] and, respectively, online learning platforms.

Our findings indicate that LLMs serve as a strong substitute for a broad range of education-related activities and as such have the potential to negatively impact a wide range

[20] `https://www.intelligent.com/one-third-of-college-students-used-chatgpt-for-schoolwork-during-the-2022-23-academic-year/`, accessed May 12, 2025.

[21] While we cannot pin down the exact source of this drop, this could be a result of decreased engagement with the platform in general through fewer practice quizzes, less discussion in boards, faster completion of assignments, or decreased downloads of course notes.

of education-related content providers. These patterns mirror findings by Chatterji et al. (2025) who document that about 10% of all messages on ChatGPT are requests for tutoring or teaching.

We also examine in Web Appendix F.1 the effects by monetization strategies used by education-related websites. Although we do not detect a significant impact on sites that monetize through advertising – likely due to the small number of such sites in our sample – we show that websites relying on subscriptions or business-to-business sales experience a decline in traffic.

User-Generated Content Platforms We next ask whether LLMs substitute for platforms where users seek answers and advice from other individuals. We consider three knowledge–sharing sites, Wikipedia, Stack Overflow, and Reddit. These platforms span a spectrum from factual and highly structured content (Wikipedia), through technical Q&A (Stack Overflow), to subjective, experience–based discussion (Reddit and social media).

Table 10 shows heterogeneous effects across knowledge-sharing platforms. Column (1) shows no significant effect of LLM adoption on URL calls to Wikipedia. In contrast, in Column (2) we observe a decline in visits to Stack Overflow. However, in Column (3) we find no significant change in URL calls to Reddit. Figure F.11 in Web Appendix F.2 mirrors these patterns. Our findings related to Stack Overflow and Reddit align with research by Burtch et al. (2024) and Quinn and Gutt (2025).

A factor that could explain the differences in effects is the type of information users seek on each platform. LLMs can substitute for traditional search engines when users look for direct answers, but their usefulness depends on the nature of the query. For factual questions, such as "How many albums did Ozzy Osbourne record?," LLMs may hallucinate or provide unreliable details (Huang et al., 2025; Wiggers, 2023), limiting their ability to replace authoritative sources like Wikipedia, particularly during our observation window (2022-2023). In contrast, many technical questions on Stack Overflow have well-defined structures and multiple valid solutions, making them well-suited to LLM-generated responses. At the other end of the spectrum, questions rooted in personal experience or social context, which are common on Reddit and social media, are less amenable to LLM substitution because users often value diverse human perspectives. Consistent with this interpretation, we find no measurable effect on visits to major social media platforms (see Web Appendix F.3).

	Dependent variable:		
	Wikipedia	Stack Overflow	Reddit
	(1)	(2)	(3)
ATT (weeks: 20-47)	−1.104	−0.649*	12.736
	(5.668)	(0.345)	(12.172)
Pre-adoption avg.	13.878	0.944	14.384
Panelists	1634	287	1488
Weeks	52	52	52
Observations	84,968	14,924	77,376
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 10: Effect on traffic on knowledge-sharing platforms: Wikipedia, Stack Overflow and Reddit

5.3 Advertising Exposures

Since our results demonstrate that LLMs substitute for at least some online activities, an important question is to what extent they reduce websites' ability to monetize through advertising. We next examine how LLM adoption affects users' advertising exposures.

	Dependent variable:				
	All ads [Google, Yieldmo]	Google ads: display	Google ads: search	Google ads: video	Yieldmo
	(1)	(2)	(3)	(4)	(5)
ATT (weeks: 20-47)	−154.213**	−122.217*	−1.914	−5.798	−24.284
	(73.299)	(73.479)	(1.978)	(6.485)	(32.310)
Pre-adoption avg.	211.606	161.092	14.686	3.952	31.876
Panelists	2041	2041	2041	2041	2041
Weeks	52	52	52	52	52
Observations	106,132	106,132	106,132	106,132	106,132
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 11: Effects on advertising exposure

Figure 9 (a) and Column (1) in Table 11 demonstrate a significant drop in ad exposure across all ads that we identified in our data. Again, this drop sets in approximately 20 weeks

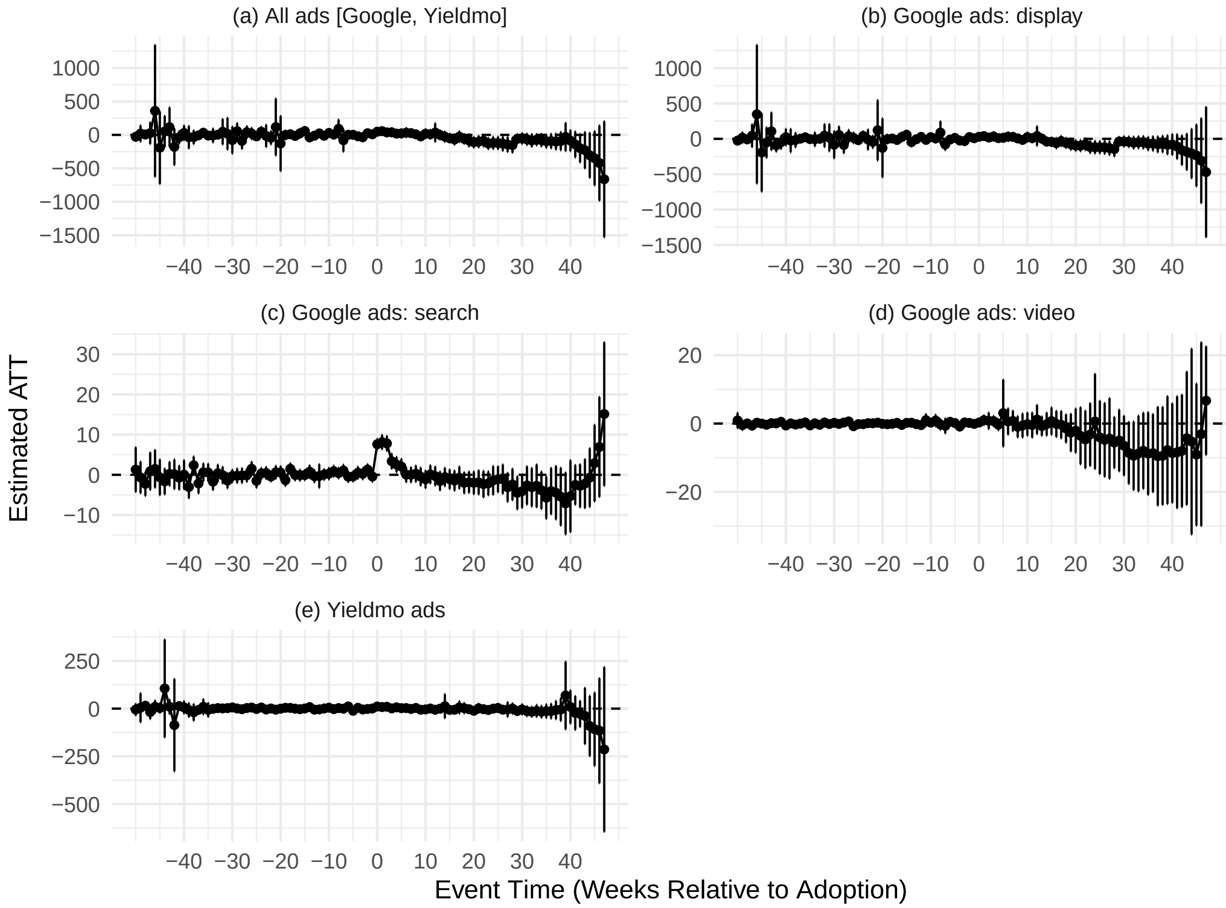

Figure 9: Staggered DiD ATT on the weekly number of URL calls to ad servers.

post adoption. Figure 9 (b) and Column (2) in Table 11 demonstrate that this pattern holds when we narrow down the sample to impressions of Google display ads. In contrast, Panel (c) in Figure 9 and Column (3) in Table 11 suggest no significant drop in the number of advertising impressions for Google search ads. These results may reflect the fact that search ads are more often shown for queries less affected by LLM adoption, such as navigational searches, and are less likely to appear after queries that experienced a sharper decline.

Figure 9, Panel (d), and Table 11, Column (4) show no significant negative effects for Google video ads, likely because LLMs are not a good substitute for video viewing. Column (5) shows no significant effect for Yieldmo ads.

Our finding that LLM adoption affects display but not search or video advertising is consistent with Alphabet's Q3 2025 report (Alphabet Inc., 2025), which shows a year-on-year decline in revenue from only "Google Network", referring to ads on third-party websites and mobile apps. While Alphabet reports increases in "Google Search & other" and "YouTube ads," neither their figures nor our estimates indicate any decline in these categories.

5.3.1 Heterogeneity by Retail Activity

Not all ad impressions are equally valuable to content creators, as bid prices in programmatic advertising depend heavily on the characteristics of the users who visit ad-supported websites. A key determinant of display ad prices is the amount of user information available to advertisers at the time of bidding. Users with richer browsing histories, especially those active on retail websites, are more likely to be retargeted by retailers. They also carry more behavioral data in their cookies, making them more attractive targets for advertisers. In turn, they become more valuable to publishers. We therefore use users' retail activity as a proxy for the value of their ad impressions and examine whether the effects of LLM adoption differ across users with varying levels of retail activity.

We classify users into three terciles based on their pre-adoption average weekly number of URL calls to retail domains: Low ($[0, 13.5]$), Mid ($[13.5, 75.1]$), and High ($[75.1, \infty]$).[22] and explore the effect of LLM adoption on ad exposures separately for each level of retail activity. Columns (1) to (3) in Table 12 demonstrate a significant drop in the level of exposure to ads overall only for the group of consumers with the highest level of retail activity. Columns (4) to (6) show that this pattern also holds for Google display ads. These results suggest that those consumers who are likely to be most valuable to retailers that advertise and to online content publishers suffer the greatest drop in advertising exposure.

Because our measure of retail activity is computed over different calendar periods and time windows depending on when users adopt, the results might partly reflect cohort-specific differences in retail activity rather than intrinsic differences across users. To address this concern, we additionally classify users based on (i) their retail activity during the same first

[22] Averages are computed using only pre-adoption activity so that post-adoption behavior does not affect user classification.

four weeks of observation and (ii) the four weeks immediately preceding adoption. The results are robust to both alternative measures (Web Appendix D.5). We consistently find a significant drop in advertising impressions for those users that have the highest level of online retail activity and, as such, are most valuable.

	Dependent variable:					
	All ads [Google, Yieldmo]			Google ads: display		
	Low	Mid	High	Low	Mid	High
	(1)	(2)	(3)	(4)	(5)	(6)
ATT (weeks: 20-47)	34.650	−9.113	−561.285**	22.493	12.079	−448.582*
	(66.737)	(34.557)	(234.000)	(61.866)	(32.213)	(243.936)
Pre-adoption avg.	106.885	166.582	364.512	85.641	124.363	275.721
Panelists	680	680	681	680	680	681
Weeks	52	52	52	52	52	52
Observations	35,360	35,360	35,412	35,360	35,360	35,412
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 12: Effects on ads by pre-adoption traffic on retail websites.

6 Discussion

We document that the adoption of LLMs leads to a significant decline in traditional online search, reduced browsing of smaller websites, and a drop in advertising impressions. These effects materialize gradually as users gain experience with LLMs. Our findings highlight a shift in how consumers engage with online information and, importantly, suggest that LLMs increasingly compete with traditional intermediaries for user attention, with implications for the monetization of online content by web publishers.

Our data capture the early phase of LLM adoption from late 2022 through late 2023 when most LLMs relied on static pre-training data . Since then, three developments have started to alter the competitive landscape. First, Google introduced AI Overviews, embedding LLM-generated answers directly into search results. Second, LLMs have increasingly incorporated live retrieval and selective outlinking. Third, some platforms have started to experiment with LLM-embedded advertising. While these changes modify the user experience and cross-platform dynamics, we believe that our core findings describe behavioral and economic forces that continue to shift user online behavior.

First, while Google's AI Overviews represent an important strategic response to LLM adoption,[23] it does not shift the dynamics leading to the decline of website traffic. During our study period, users were already shifting search activity toward stand-alone LLMs. Indeed, by providing synthesized answers directly on the results page, AI Overviews replicate one of the core mechanisms underlying the decline in website traffic: users obtain satisfactory information without clicking through to underlying sources. The implications for browsing patterns are similar to those we observe: smaller websites, which disproportionately rely on traffic referred to from search, are likely to be most affected. More broadly, the introduction of AI Overviews may intensify competition between AI-based search engines and publishers for user attention.

Second, the introduction of retrieval-augmented LLMs sharpens competition between AI platforms and search engines. The ability to access up-to-date information and provide citations makes LLMs more viable substitutes for traditional search. These improvements may increase the share of queries for which users rely primarily on LLMs. Although the inclusion of selective outlinks could mitigate some of the decline in browsing, LLMs typically surface only a small number of curated links, which may increase concentration of traffic and reinforce the competitive advantages of already well-established sources. Smaller publishers may continue to suffer a greater decline in traffic and, by implication, in advertising exposures. Indeed, recent industry reports suggest a significant drop in the amount of traffic going to publishers' sites for such newer models (Tollbit, 2024). Further, for the large number of users relying on free versions of LLMs, which only supply limited outlinking and often restrict retrieval, the dynamics we document remain unchanged.

Third, the evolving landscape raises the prospect of new advertising markets emerging within LLM ecosystems. Although LLM-embedded advertising is not yet widespread, it rep-

[23] See `https://www.semrush.com/blog/semrush-ai-overviews-study/`

resents a natural monetization strategy for AI platforms. If LLMs become central information intermediaries, advertising funds may shift toward these platforms, further intensifying competition with traditional web publishers and existing digital advertising platforms such as display ad networks and search ads. Importantly, even if LLMs create new opportunities for advertisers, such developments are unlikely to compensate publishers for lost traffic or lost advertising impressions, particularly if LLMs themselves become the primary point of contact between consumers and information.

As the boundaries between search engines and LLMs continue to blur, both incumbents and AI entrants face strategic pressure to differentiate on features, accuracy, live updates, and user experience. The resulting competitive dynamics are likely to shape how information is accessed, how content is monetized, and how value is distributed across the digital ecosystem. Although our data do not allow us to observe these later developments directly, the mechanisms we document – substitution away from search, reductions in traffic to smaller sites, and declines in advertising impressions – provide early evidence of behavioral and economic forces that will continue to influence the evolving competitive landscape.

7 Conclusion

The broad adoption of AI tools has the potential to reshape how consumers acquire information online, as these tools may serve either as substitutes for or complements to existing digital resources. In turn, these shifts may fundamentally alter the economics of online platforms and more broadly online content production.

In this paper, we use a large panel dataset of detailed browsing behavior during 2022 and 2023 to provide initial evidence for how users' adoption of LLMs affects their online behavior.

Our primary results suggest that concerns about LLMs substituting for web browsing may be well-founded, at least for a subset of online content providers. After adopting LLMs, users make fewer searches in traditional search engines, including question queries as well as both short and long searches. This pattern indicates that for many types of queries, LLMs act as a substitute rather than expanding the demand for online search. As more users adopt LLMs, the volume of traditional online search is likely to fall further. We also find that while visits to high-traffic websites remain largely unchanged, visits to smaller websites decline significantly, which suggests that such websites are particularly vulnerable to substitution by LLM usage. Consistent with users obtaining more information directly within LLMs, ad impressions fall for LLM adopters, potentially threatening the viability of at least some types of publishers and, in the longer run, diminishing incentives to produce new content. We note, though, that we are unable to pin down the quality of content produced by the websites that experience more pronounced drops in visits.

Notably, the decline in search and browsing activity does not occur immediately but emerges roughly 20 weeks after adoption, suggesting that users learn about effective LLM usage and that substitution is contingent on this learning experience. This lag implies that economic effects for online content providers may materialize only gradually, relative to the timing of users' initial adoption.

Our analysis has several limitations to our analysis. First, our data provides no visibility on how users interact with LLMs. That is, we do not know what queries are being made inside these tools and hence the specific behaviors that are substituting for the reduction in online activity. Contemporary work by Handa et al. (2025) provides some evidence here. Second, we classify a user as having adopted LLMs after using them for at least three consecutive weeks. We highlight that different definitions of adoption yield different effect size estimates. Third, our sample ends at the end of 2023. Since then, LLM adoption and usage intensity have grown rapidly, potentially changing the magnitude of the effects we measure. Our results should be seen as initial evidence of what we expect to become longer-term trends. Fourth, while we demonstrate a reduction in traffic to smaller websites, this is an average effect, and we are unable to pin down whether high-quality sites offering differentiated niche content and lower-quality sites with little original information are impacted to the same degree. Last, our data period covers LLMs that are pre-trained but did not provide live retrieval or outlinks and when Google AI Overviews were not yet available. However, as we explain, while users are likely to further adjust their behavior as models change over time, we expect the fundamental behavioral and economic forces and their effect to remain similar.

References

Alphabet Inc. (2025). Alphabet announces third quarter 2025 results. Press release; quarter ended September 30, 2025. Accessed on 2025-12-08.

Athey, S., Mobius, M., and Pal, J. (2021). The impact of aggregators on internet news consumption. Technical report, National Bureau of Economic Research.

Bairathi, M., Lambrecht, A., and Rao, A. (2024). Social media, music consumption, and cross-platform spillover effects. Working paper.

Becker, G. S. and Murphy, K. M. (1993). A simple theory of advertising as a good or bad. *The Quarterly Journal of Economics*, 108(4):941–964.

Blake, T., Nosko, C., and Tadelis, S. (2015). Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica*, 83(1):155–174.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, 91(6):3253–3285.

Burbidge, J. B., Magee, L., and Robb, A. L. (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American statistical Association*, 83(401):123–127.

Burtch, G., Lee, D., and Chen, Z. (2024). The consequences of generative AI for online knowledge communities. *Scientific Reports*, 14(1):10413.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2):200–230.

Calzada, J. and Gil, R. (2020). What do news aggregators do? Evidence from Google News in Spain and Germany. *Marketing Science*, 39(1):134–167.

Chatterji, A., Cunningham, T., Deming, D. J., Hitzig, Z., Ong, C., Shan, C. Y., and Wadman, K. (2025). How people use ChatGPT. Technical report, National Bureau of Economic Research.

Cheng, M. M., Ofek, E., and Yoganarasimhan, H. (2024). The value of silence: The effect of UMG's licensing dispute with TikTok on music demand. *working paper*.

Chiou, L. and Tucker, C. (2017). Content aggregation by platforms: The case of the news media. *Journal of Economics & Management Strategy*, 26(4):782–805.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American economic review*, 110(9):2964–2996.

del Rio-Chanona, R. M., Laurentsyeva, N., and Wachs, J. (2024). Large language models reduce public knowledge sharing on online Q&A platforms. *PNAS nexus*, 3(9):pgae400.

Deng, Y., Lambrecht, A., and Liu, Y. (2023). Spillover effects and freemium strategy in the mobile app market. *Management Science*, 69(9):5018–5041.

Fradkin, A. (2025). Demand for llms: Descriptive evidence on substitution, market expansion, and multi-homing. *Working Paper*.

Goldberg, S. and Lam, H. T. (2025). Generative AI in equilibrium: Evidence from a creative goods marketplace. Technical Report SSRN 5152649, Stanford Graduate School of Business.

Greminger, R. P., Huang, Y., and Morozov, I. (2023). Make every second count: Time allocation in online shopping. *Available at SSRN 4643819*.

Handa, K., Tamkin, A., McCain, M., Huang, S., Durmus, E., Heck, S., Mueller, J., Hong, J., Ritchie, S., Belonax, T., et al. (2025). Which economic tasks are performed with ai? evidence from millions of claude conversations. *arXiv preprint arXiv:2503.04761*.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., et al. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2):1–55.

Humlum, A. and Vestergaard, E. (2025). The unequal adoption of ChatGPT exacerbates existing inequalities among workers. *Proceedings of the National Academy of Sciences*, 122(1):e2414972121.

Lambrecht, A. and Misra, K. (2017). Fee or free: when should firms charge for online content? *Management Science*, 63(4):1150–1165.

Larsen, P. and Proserpio, D. (2025). The impact of LLMs on sponsored search: Evidence from Google's BERT. *USC Marshall School of Business Research Paper Sponsored by iORB*.

Lee, S., Zhang, J., and Wedel, M. (2021). Managing the versioning decision over an app's lifetime. *Journal of Marketing*, 85(6):44–62.

Li, X. and Kim, K. (2024). Impacts of generative AI on user contributions: evidence from a coding Q&A platform. *Marketing Letters*, pages 1–15.

Lyu, L., Siderius, J., Li, H., Acemoglu, D., Huttenlocher, D., and Ozdaglar, A. (2025). Wikipedia contributions in the wake of ChatGPT. *arXiv preprint arXiv:2503.00757*.

Oberholzer-Gee, F. and Strumpf, K. (2007). The effect of file sharing on record sales: An empirical analysis. *Journal of political economy*, 115(1):1–42.

Pattabhiramaiah, A., Sriram, S., and Manchanda, P. (2019). Paywalls: Monetizing online content. *Journal of marketing*, 83(2):19–36.

Quinn, M. and Gutt, D. (2025). Heterogeneous effects of generative artificial intelligence (GenAI) on knowledge seeking in online communities. *Journal of Management Information Systems*, 42(2):370–399.

Rob, R. and Waldfogel, J. (2007). Piracy on the silver screen. *The Journal of Industrial Economics*, 55(3):379–395.

Roth, J. (2024). Interpreting event-studies from recent difference-in-differences methods. Papers 2401.12309, arXiv.org.

Seamans, R. and Zhu, F. (2014). Responses to entry in multi-sided markets: The impact of Craigslist on local newspapers. *Management Science*, 60(2):476–493.

Shankar, R. and Sim, J. (2024). Advances in generative AI and platform moderation: Implications for online knowledge sharing. *Available at SSRN 4867815*.

Shorakaei, H., Zhang, M., Bayley, T., and Begen, M. A. (2025). Empowering or eroding contributions? how generative AI impacts user-generated content across diverse communities. *How Generative AI Impacts User-Generated Content Across Diverse Communities (March 24, 2025)*.

Sun, M. and Zhu, F. (2013). Ad revenue and content commercialization: Evidence from blogs. *Management Science*, 59(10):2314–2331.

Telang, R. and Waldfogel, J. (2018). Piracy and new product creation: A bollywood story. *Information Economics and Policy*, 43:1–11.

The Economist (2024). What ChatGPT's corporate victims have in common. *The Economist*. Accessed on 2025-12-09.

Tollbit (2024). State of the bots - q4 2024. *Journal of the American statistical Association*, 83(401):123–127. Accessed on 2025-04-09.

Warren, T. (2023). You can play with Microsoft's Bing GPT-4 chatbot right now, no waitlist necessary. The Verge. Accessed on 2025-05-15.

Wiggers, K. (2023). Are AI models doomed to always hallucinate? TechCrunch. Accessed on 2025-08-03.

Winkler, D., Hotz-Behofsits, C., Wlomert, N., Papies, D., and Liaukonyte, J. (2024). The impact of social media on music demand: Evidence from a quasi-natural experiment. *working paper*.

Yang, S. A. and Zhang, A. H. (2024). Generative AI and copyright: A dynamic perspective. *arXiv preprint arXiv:2402.17801*.

WEB APPENDIX

A Sample Criteria Sensitivity

This appendix shows how sensitivity our main results are to our sample criteria. Note that for the main sample we have two layers of criteria: at least 4 URL calls in every month for our 1 year sample period; and adoption defined based on 3 consectuive weeks of at least one URL call to an LLM. Figures A.1 and A.2 show that our main estimates are all directionally consistent when varying one layer of the criteria (holding the other layer fixed). Generally speaking, a more stringent criteria (move to the right of the figures) leads to noisier estimates due to a smaller sample size.

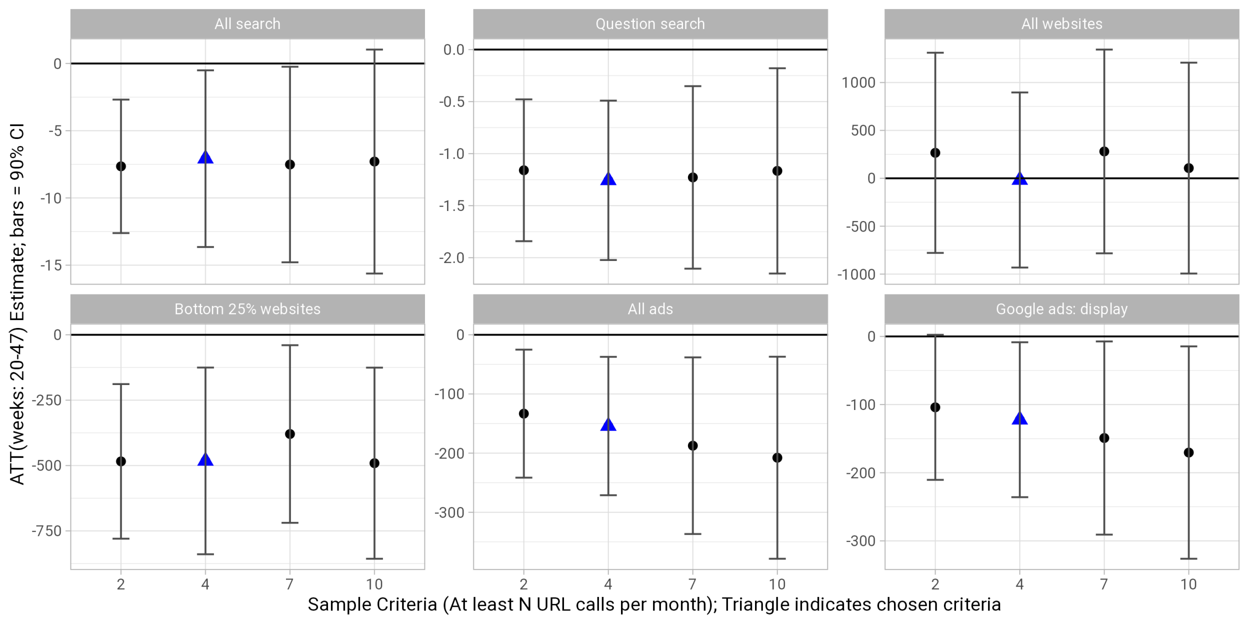

Figure A.1: Sensitivity to overall sample criteria. The number of panelists for each criterion is: (a) 2 URL calls: 2,374, (b) 4 URL calls (main sample): 2,041, (c) 7 URL calls: 1,686, and (d) 10 URL calls: 1,392.

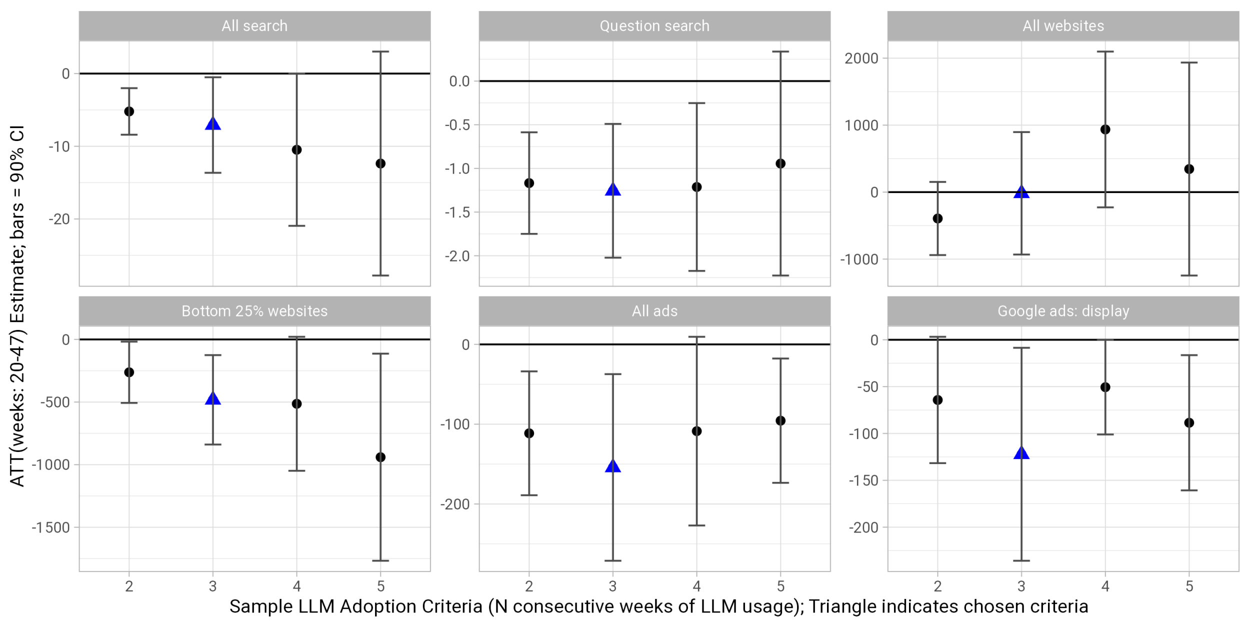

Figure A.2: Sensitivity to LLM adoption sample criteria. The number of panelists for each criterion is: (a) 2 weeks: 4,056, (b) 3 weeks (main sample): 2,041, (c) 4 weeks: 1,309, and (d) 5 weeks: 922.

B Additional Data Descriptives

B.1 Summary statistics of adopters vs. full sample

		Full Sample	LLM Adopters			
			Sample LLM Adoption Criteria			
Category	Variable		2 weeks	**3 weeks**	4 weeks	5 weeks
Search queries	All search	7.9	26.7	32.7	37.3	43.6
	Google	5.0	15.8	17.7	19.7	21.5
	Questions	0.7	3.1	3.8	4.4	4.9
	Navig. only	0.3	0.8	0.8	0.8	0.9
	Navig. + other	0.4	1.2	1.5	1.7	2.0
	Long	1.5	5.4	6.6	7.6	8.7
	Short	1.4	4.3	5.0	5.4	6.2
Website browsing	All websites	1,688.1	3,577.2	4,325.1	4,809.0	5,494.2
	Top 500 websites	1,107.1	2,361.5	2,928.0	3,294.0	3,801.5
	non-Top 500 websites	581.1	1,215.8	1,397.1	1,515.0	1,692.7
	Top 1,000 websites	1,222.4	2,613.1	3,219.9	3,594.3	4,121.0
	non-Top 1,000 websites	465.8	964.1	1,105.2	1,214.6	1,373.1
	25%-50%	399.1	763.4	943.5	1,002.0	1,140.8
	50%-75%	531.4	1,098.9	1,395.7	1,631.6	1,916.9
	Bottom 25%	489.1	1,007.0	1,154.3	1,273.2	1,438.4
	Top 25%	268.6	707.9	831.6	902.1	998.1
Referred traffic	All referred	6.8	18.9	21.7	23.1	24.0
	Top 25%	1.5	4.2	4.7	5.2	5.6
	25%-50%	1.8	5.9	6.9	6.8	6.7
	50%-75%	1.7	4.6	5.2	5.8	5.9
	Bottom 25%	1.8	4.3	4.9	5.3	5.8
Ad exposures	All ads [Google, Yieldmo]	70.3	175.9	211.6	231.0	263.6
	Google ads: display	50.4	132.4	161.1	178.5	205.9
	Google ads: search	4.4	12.7	14.7	16.1	17.5
	Google ads: video	1.2	2.8	4.0	5.3	6.6
	Yieldmo	14.4	28.0	31.9	31.1	33.6
Control	All control	354.6	558.4	648.9	742.7	787.8
	Email	133.7	324.2	378.4	442.6	488.0
	Retail	196.3	187.0	208.1	236.2	234.6
	News	24.6	47.2	62.5	63.9	65.2
Panelists		74,940	4,056	2,041	1,309	922

Table B.1: Summary statistics of dependent variables and controls of full sample vs. LLM adopters. Main sample in bold (adoption criterion of 3 weeks of LLM usage).

B.2 Cohort heterogeneity in LLM activity

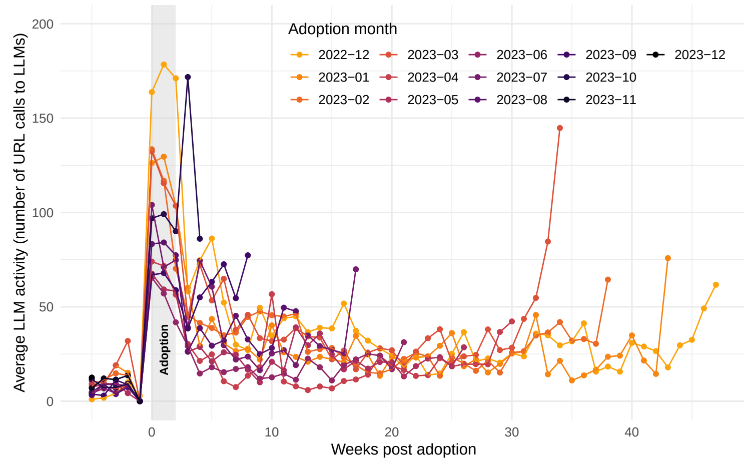

Figure B.3: Average number of URL calls to LLMs per post-adoption week by adoption month.

C Comscore Data Processing

This appendix provides additional details on key variables constructed in this paper. The primary data source is the Comscore US desktop panel, which provides detailed, event-level records of internet browsing activity for a large sample of users.

C.1 Constructing Measures of Web Activity: Foreground vs. Background Traffic

A significant challenge in using raw web traffic data is that it contains a mixture of user-initiated "foreground" activity and automated "background" activity. Foreground activity reflects a user's active browsing choices, such as navigating to a job website (e.g., `www.indeed.com`) or visiting a social media page (e.g., `www.instagram.com`). It is important to note that a single user action, like loading a webpage, can trigger multiple associated URL requests (e.g., to load images, scripts, or advertisements). Our definition of foreground activity includes these associated requests, as they are a direct consequence of the user's initial choice. In contrast, background activity consists of HTTP requests generated by software or web page components without direct user interaction. Examples include telemetry pings from browser extensions (e.g., `self.events.data.microsoft.com`), and automated checks for software updates.

To construct meaningful measures of user behavior, it is crucial to distinguish between these two types of traffic. Our primary strategy for this is to identify URLs that are likely the result of active user choice.

C.1.1 The `Website traffic` Variable: A Measure of General Foreground Activity

The main variable for general web browsing, `Website traffic`, is a count of visits to general-purpose "foreground" websites. The construction of this variable relies on a data-driven approach to differentiate foreground from background URLs based on their traffic sources.

We posit that websites actively visited by users receive a meaningful share of their traffic from explicit user actions, such as a search engine referral. Conversely, background services, while generating substantial traffic, are rarely reached via search engines. The construction process is as follows:

1. *Aggregate Total Traffic:* We first calculate the total number of visits for every unique URL host across the entire 2022-2023 sample period. This provides a baseline measure of overall traffic for each host.

2. *Aggregate Referred Traffic:* We then isolate traffic that is referred from major search engines (Google, Bing, and Yahoo). This gives us a count of search-referred visits for each URL host. These referrals are strong indicators of active, user-initiated navigation.

3. *Calculate Referral-to-Total Traffic Ratio:* For each URL host, we compute the ratio of its search-referred traffic to its total traffic. This ratio serves as our primary indicator for classifying a host as foreground or background.

4. *URL Classification:* A URL host is classified as "foreground" if its referral ratio exceeds a threshold of 1×10^{-6}, and "background" otherwise.

5. *Validation of Classification:* The choice of this threshold was validated by comparing the ratio-based classification against a manual classification of the top 500 URLs by traffic. This manual classification was conducted by research assistants who visited each website to determine its primary function. The automated threshold method agrees with the manual classification for 89% of these top 500 URLs, which shows that the chosen threshold effectively separates known foreground sites (like `youtube.com` or `wikipedia.org`) from known background sites (like `self.events.data.microsoft.com`).

Exclusions from General Activity The `general_fore_n` measure only includes traffic to URLs classified as foreground. This means we explicitly exclude traffic to hosts that exhibit characteristics of background services. Many of these sites, such as the example `self.events.data.microsoft.com`, have extremely low or zero referral ratios despite having high overall traffic volume. This pattern is typical of services that do not serve user-facing content and are instead part of a software's operational infrastructure.

Robustness and Justification This classification method provides a scalable and objective way to clean the raw traffic data. A potential concern is that by using a sharp threshold, we might misclassify some legitimate, albeit niche, websites as background if they receive little search traffic. However, our analysis indicates that this is not a major source of bias. The set of hosts excluded by our threshold accounts for a negligible fraction (less than 0.1%) of all search-referred traffic in our sample. Therefore, the resulting sample of foreground traffic remains representative of user-initiated browsing behavior originating from search engines.

C.2 The `Ads` Variable: A Measure of Advertising Activity

To ensure that our classification of ad exposures is accurate, we validate our classification of ad service URLs against user browsing behavior. Specifically, we classify ad requests originating from domains such as `adservice.google.com` as either "Google Search Ads" or "Google Display Ads" based on specific URL patterns.

We define Google Search Ads by URLs containing `/adsid/google/ui` or `/adsid/google/si`. We define Google Display Ads by URLs containing `/adsid/integrator.js`, `/adsid/integrator.json`, or domains like `www.googletagservices.com`.

In the below, we provide validation of these classifications by examining the user's browsing history immediately preceding the ad request. For Search Ads, we check for the presence of a Google Search URL (e.g., containing "google" and "search") within a short time window prior to the ad event. Similarly, for Display Ads, we check for visits to non-Google websites, which would serve as the publisher for the display ad. We employ two time windows: a 10-second window and a 30-second window. The "Strict" specification requires the source event to strictly precede the ad event ($t_{source} < t_{ad}$), while the "Inclusive" specification allows for simultaneous timestamps ($t_{source} \leq t_{ad}$), accounting for potential measurement resolution limits.

Our validation results, presented in Table **??**, strongly support our classification scheme. We find that approximately 70.8% of identified Display Ads are preceded by a non-Google website visit within 30 seconds, and 69.2% of Search Ads are preceded by a Google search. The existence of a delay between the source visit and the ad request is expected; it reflects the time required for page rendering, network latency, and user behavior such as scrolling down a page before ads—often loaded lazily—are triggered.

The remaining fraction of ad requests that are not matched to a specific source visit within the window can be attributed to common browsing behaviors that trigger ad loading without a new page view. Detailed examination suggests these often occur when a user returns to a previously opened browser tab, triggering a refresh of the ad slots without reloading the main page URL. Additionally, "infinite scroll" features on many modern content feeds can load new ads long after the initial page load event recorded in the data. Thus, our match rates are consistent with a highly accurate classification of ad types, limited only by the granularity of the browsing log events.

C.3 Data Validation: Video Ads

To verify the nature of the ad service URL `imasdk.googleapis.com`, commonly associated with Google Video Ads, we examine whether it is exclusively used for YouTube or delivers video ads across a range of websites. We classify these ad requests as "Google ads: video".

We validate this by calculating the share of `imasdk.googleapis.com` requests that are immediately preceded by a visit to `youtube.com` within a 30-second window (allowing for simultaneous timestamps). We find that only 4.6% of these ad requests are associated with a YouTube visit. This low match rate confirms that `imasdk.googleapis.com` is a general video ad service used by a wide variety of publishers beyond YouTube, consistent with our classification of it as a broad ad service rather than a platform-specific one.

D Robustness Checks

D.1 Anticipation Robustness

Figure D.4 shows the main results of the paper under a different control period. The blue triangle shows the results shown in the paper using the week prior to adoption as the control period, which relies on the no-anticipation assumption. That is, individuals do not anticipate they will adopt the LLM by changing their behavior in anticipation of the adoption. Results under other control periods use the assumption that there is no anticipation until that control period. For example, using a control period of $t = -2$ means that we assume individuals do not change their behavior two weeks prior to adopting, but could potentially anticipate adoption one week before adoption. Most of the estimates under different anticipation assumptions are covered by the 90% CI of our main specification.

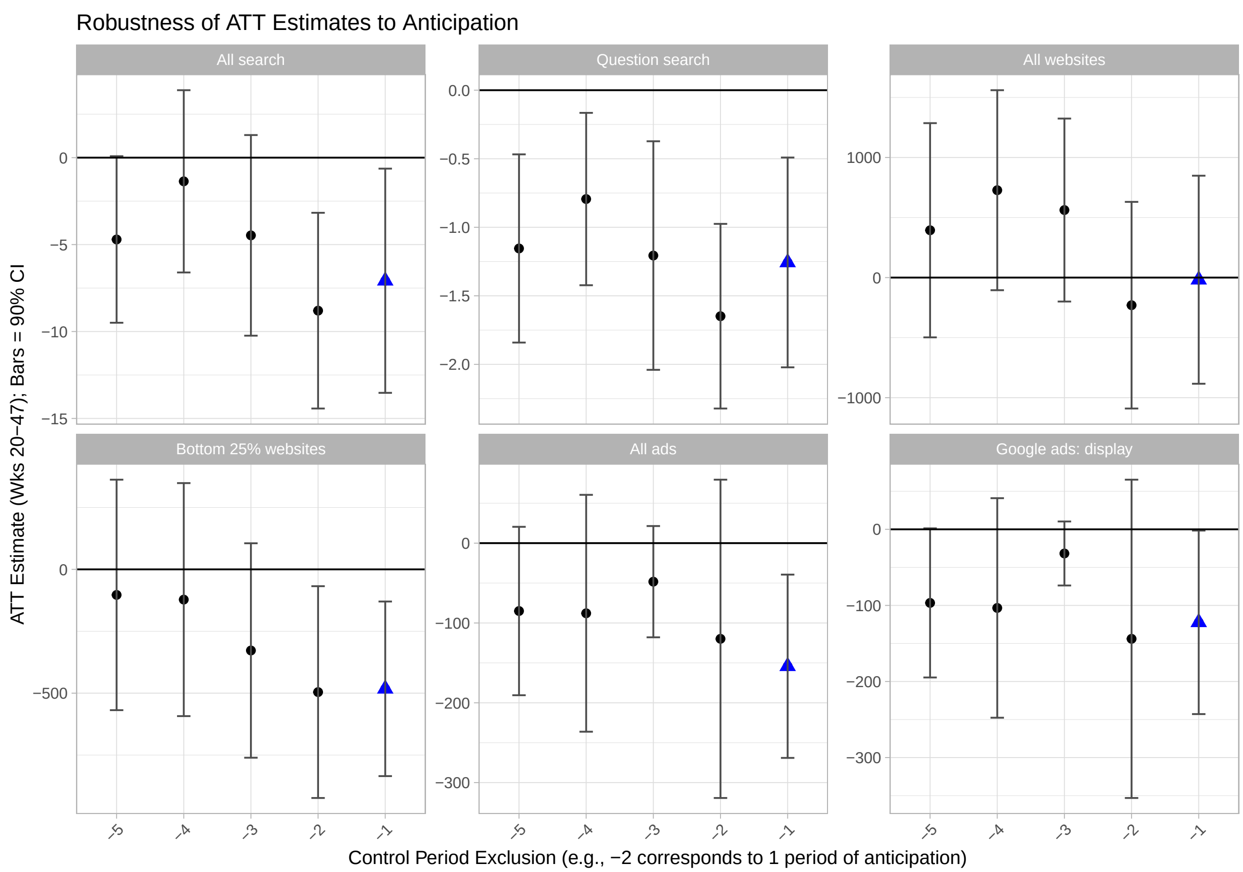

Figure D.4: Sensitivity to Control Period Exclusion

D.2 Model Specification Sensitivity

This appendix shows the effects of the main results of the paper under different model specifications. In addition to the results included in the main manuscript displayed in blue, Figure D.5 shows the effects for: Callaway and Sant'Anna (2021) using never treated units (which essentially includes only are those treated after November 2023 and before the end of our observation window), two-way fixed effects difference-in-differences regression (TWFE), and Poisson Pseudo-Maximum Likelihood Regression evaluated at the pre-adoption mean of the outcomes.

We find that our estimates are largely consistent with those by other model specifications. That is, the estimates under other specifications are mostly covered by the 90% CI of our main specification.

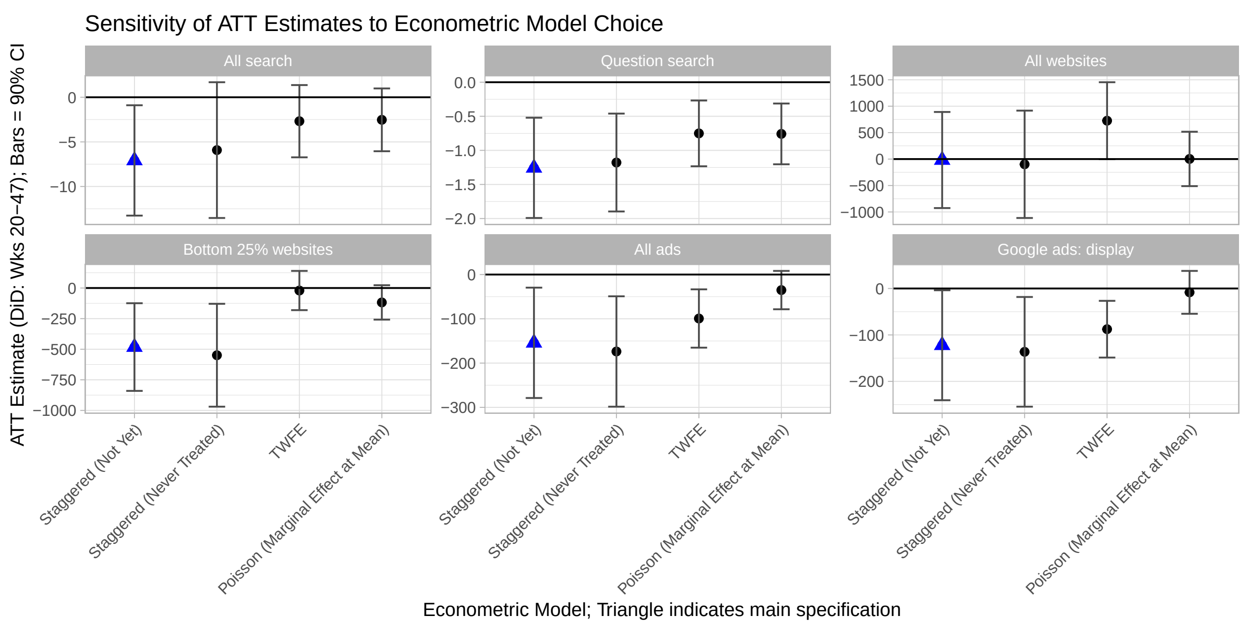

Figure D.5: Sensitivity to Model Specification

D.3 Exclusion of Early Adopters

This appendix shows the robustness of the effects of the main results of the paper at excluding early adopters. In addition to the results included in the main manuscript shown in blue, Figure D.6 shows the effects when removing the first four adopter cohorts.

We find that our estimates are largely consistent with those using all users. That is, the point estimates when excluding early adopters are mostly covered by the 90% CI of our main specification.

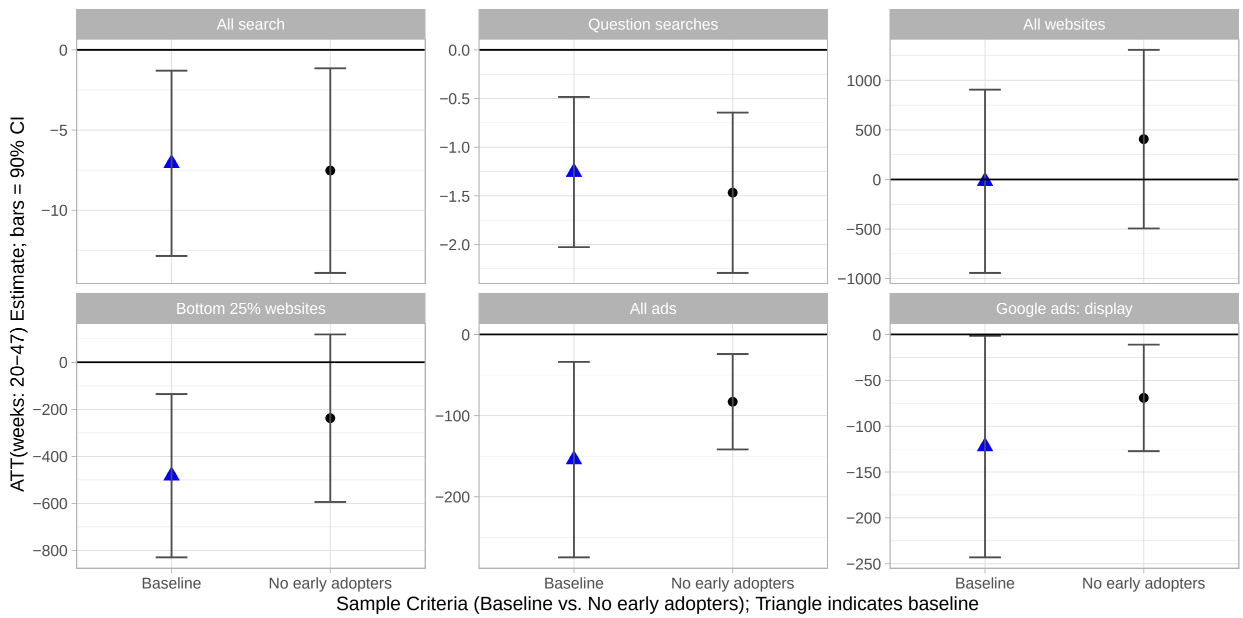

Figure D.6: Sensitivity to exclusion of early adopters

D.4 Placebo Tests

A potential concern with our empirical strategy is that, even though the staggered differences-in-differences approach accounts for calendar-specific shocks, the estimated effects of LLM adoption might still be confounded with broader behavioral trends that unfold later in the observation window. For example, if users naturally change their search, browsing, or ad-related behavior over time for reasons unrelated to LLM usage, our estimates could inadvertently capture these secular shifts. To ensure that the patterns we observe are truly associated with LLM adoption, and not with underlying temporal dynamics, we conduct a series of placebo tests.

We implement these placebo tests by randomly assigning each user a placebo "adoption" week, drawn from the set of weeks during which actual adoption is observed. For each randomized assignment, we apply the same estimation procedure used in the main analysis and compute the estimated effects 20 weeks after placebo adoption. To assess whether any spurious effects arise purely by chance, we repeat this randomization procedure 100 times.

Figures D.7, D.8, and D.9 present histograms of the resulting ATT estimates across these replications. As expected, these placebo estimates are centered near zero and show no systematic deviations, indicating that placebo "adoption" has no detectable effect on any outcome. This provides strong evidence that the effects documented in the main paper are not artifacts of underlying temporal patterns but are instead associated with actual LLM adoption.

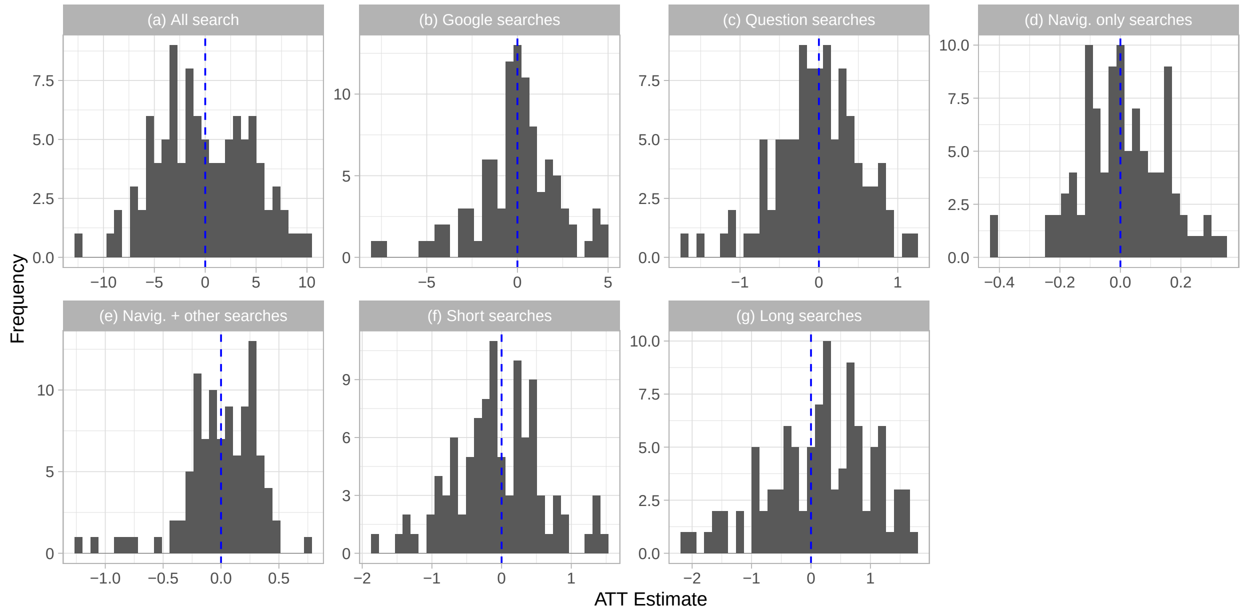

Figure D.7: Distribution of ATT on weekly searches under placebo treatment randomization.

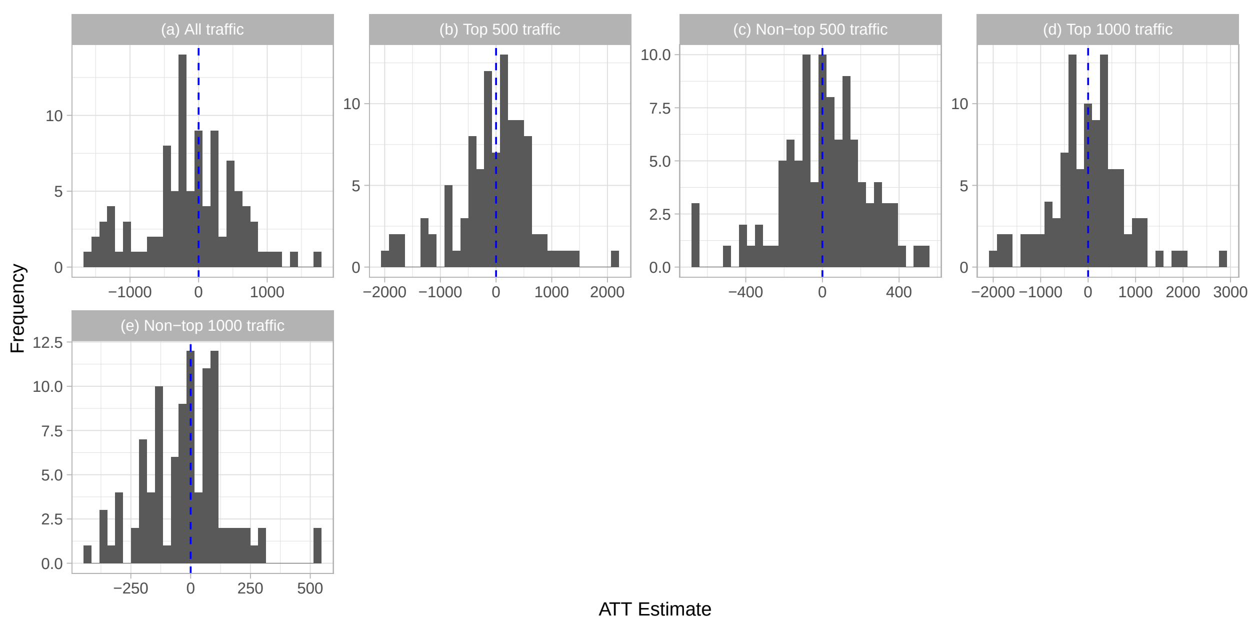

Figure D.8: Distribution of ATT on the weekly number of URL calls under placebo treatment randomization.

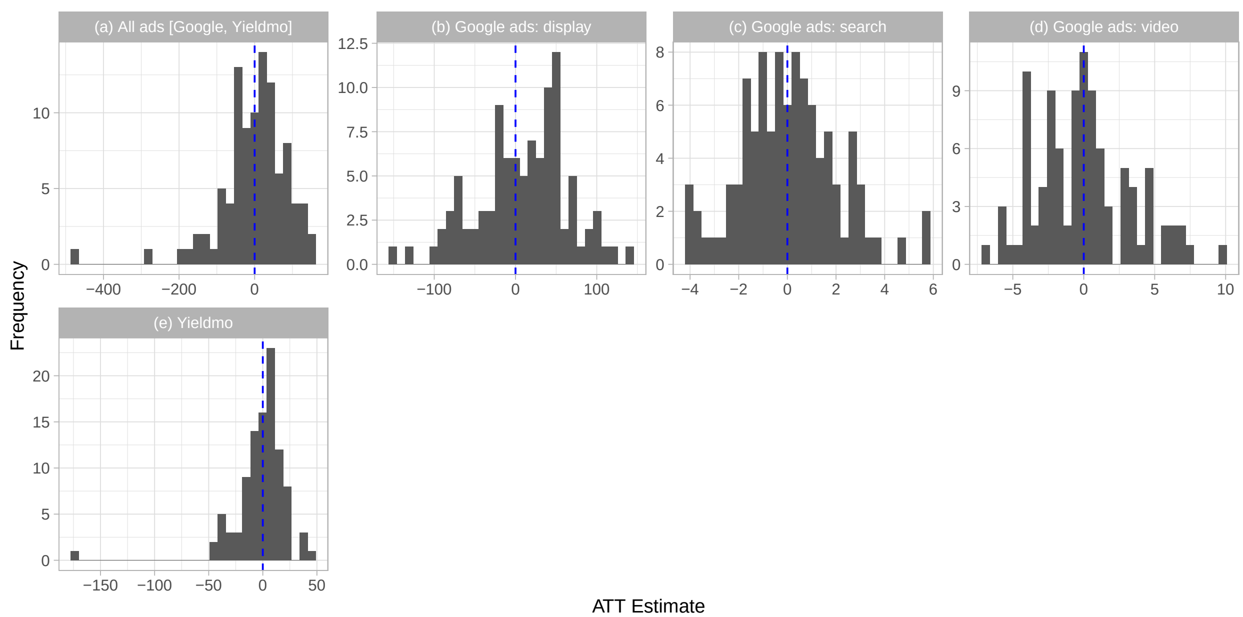

Figure D.9: Distribution of ATT on the weekly number of URL calls to ad servers under placebo treatment randomization.

D.5 Classification of Retail Activity

We verify that the results shown in Table 12 are not driven by adoption cohort differences by using two alternative definitions. First, we compute retail activity based on the same first four weeks of data across all users to ensure all averages are measured over a common calendar window. Second, we compute retail activity using only the four weeks immediately preceding adoption (and control period, i.e., periods -5 to -2 relative to adoption), to control for effects related to activity bias relative to the adoption period.. Tables D.2 and D.3 report the results for these alternative measures. In both cases, we continue to find that the reduction in ad exposures is concentrated among users with the highest levels of retail activity, confirming the robustness of our main findings.

	Dependent variable:					
	All ads [Google, Yieldmo]			Google ads: display		
	Low	Mid	High	Low	Mid	High
	(1)	(2)	(3)	(4)	(5)	(6)
ATT (weeks: 20-47)	−10.741	15.863	−472.288**	−6.835	0.195	−361.188*
	(64.427)	(27.214)	(209.222)	(62.493)	(21.239)	(216.687)
Pre-adoption avg.	118.867	136.558	389.606	93.097	98.173	299.684
Panelists	683	677	681	683	677	681
Weeks	52	52	52	52	52	52
Observations	35,516	35,204	35,412	35,516	35,204	35,412
Note:						*p<0.1; **p<0.05; ***p<0.01

Table D.2: Effects on ads by pre-adoption traffic on retail websites.

	Dependent variable:					
	All ads [Google, Yieldmo]			Google ads: display		
	Low	Mid	High	Low	Mid	High
	(1)	(2)	(3)	(4)	(5)	(6)
ATT (weeks: 20-47)	16.185	−104.754	−418.074*	1.080	−15.777	−399.203*
	(66.732)	(86.006)	(239.546)	(63.597)	(48.533)	(232.650)
Pre-adoption avg.	102.457	175.837	363.650	82.056	132.170	274.213
Panelists	680	680	681	680	680	681
Weeks	52	52	52	52	52	52
Observations	35,360	35,360	35,412	35,360	35,360	35,412
Note:						*p<0.1; **p<0.05; ***p<0.01

Table D.3: Effects on ads by pre-adoption traffic on retail websites.

E Cohort-level ATTs

We analyze heterogeneity across adoption cohorts by focusing on four groups of cohorts: (1) weeks 4-15, (2) weeks 16-21, and (3) weeks 22-28. The two groups of users adopt before the spike in adoption for Bing Chat. The final group captures those that adopted during the launch of Bing (Chat). We removed cohorts that adopted after week 28 since we only observe for a handful of weeks in the 20-48 weeks post-adoption window. Then, we aggregate $ATT_{g,t-g}$ effects across groups of cohorts (instead of all cohorts as done in the primary analysis).

	ATT (weeks 20-47)						
	All search	Google	Questions	Navig. only	Navig. + other	Short	Long
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cohort (weeks: 4-15)	−6.406	−7.004	−1.413***	−0.304	−0.419**	−1.240*	−2.257**
	(4.531)	(4.297)	(0.537)	(0.199)	(0.169)	(0.731)	(0.914)
Cohort (weeks: 16-21)	0.124	−4.751	−0.234	0.042	−0.273	0.061	−0.602
	(6.380)	(3.758)	(0.742)	(0.228)	(0.324)	(1.034)	(1.351)
Cohort (weeks: 22-28)	−18.803***	0.236	−0.721	−0.322	−0.281	−4.554**	−4.528**
	(6.139)	(2.741)	(0.765)	(0.253)	(0.338)	(1.814)	(1.829)
Pre-adoption avg.	32.669	17.680	3.779	0.800	1.500	5.030	6.630
Note:							*p<0.1; **p<0.05; ***p<0.01

Table E.4: Cohort heterogeneous effects broken down by type of search

	ATT (weeks 20-47)				
	All referred	Top 25%	25%-50%	50%-75%	Bottom 25%
	(1)	(2)	(3)	(4)	(5)
Cohort (weeks: 4-15)	−0.017	0.247	2.106	−1.249	−1.121*
	(4.913)	(0.829)	(4.248)	(1.101)	(0.630)
Cohort (weeks: 16-21)	−4.571	−0.276	−0.208	−2.765	−1.322
	(13.472)	(2.450)	(13.170)	(3.601)	(0.979)
Cohort (weeks: 22-28)	0.202	−2.143	3.466	−0.740	−0.381
	(4.386)	(2.217)	(3.065)	(0.968)	(0.657)
Pre-adoption avg.	21.723	4.749	6.869	5.158	4.947
Note:					*p<0.1; **p<0.05; ***p<0.01

Table E.5: Cohort heterogeneous effects broken down on referred traffic

	ATT (weeks 20-47)				
	All ads [Google, Yieldmo]	Google ads: display	Google ads: search	Google ads: video	Yieldmo
	(1)	(2)	(3)	(4)	(5)
Cohort (weeks: 4-15)	−133.492	−100.197	−2.574	−5.711	−25.010
	(84.338)	(82.071)	(2.339)	(4.668)	(38.611)
Cohort (weeks: 16-21)	−22.099	−7.145	−3.358	−3.960	−7.638
	(40.855)	(38.944)	(2.596)	(5.500)	(10.657)
Cohort (weeks: 22-28)	−275.374***	−264.413**	−0.906	−3.697	−6.358
	(105.341)	(105.693)	(2.465)	(7.150)	(17.651)
Pre-adoption avg.	211.606	161.092	14.686	3.952	31.876
Note:					*p<0.1; **p<0.05; ***p<0.01

Table E.6: Cohort heterogeneous effects broken down on ads

	ATT (weeks 20-47)			
	All education	Variety of URL calls	Learning management system	Online learning platform
	(1)	(2)	(3)	(4)
Cohort (weeks: 4-15)	−176.342**	−0.616***	−23.051*	−146.631**
	(84.972)	(0.195)	(13.441)	(72.547)
Cohort (weeks: 16-21)	−29.101	0.031	5.492	−40.445
	(62.685)	(0.210)	(23.213)	(49.219)
Cohort (weeks: 22-28)	−90.154	−0.377	−34.869	−58.493
	(61.597)	(0.274)	(28.419)	(40.619)
Pre-adoption avg.	169.688	1.290	62.199	55.785
Note:				*p<0.1; **p<0.05; ***p<0.01

Table E.7: Cohort heterogeneous effects broken down on education websites

F Additional Analysis on Specific Domain Types

F.1 Additional Analysis on Education-related websites

We extend the analysis on the effects of the adoption of LLMs on educational content providers. First, Figures F.10 (a)-(d) show the ATTs over time. These figures confirm the patterns shown by Table 9.

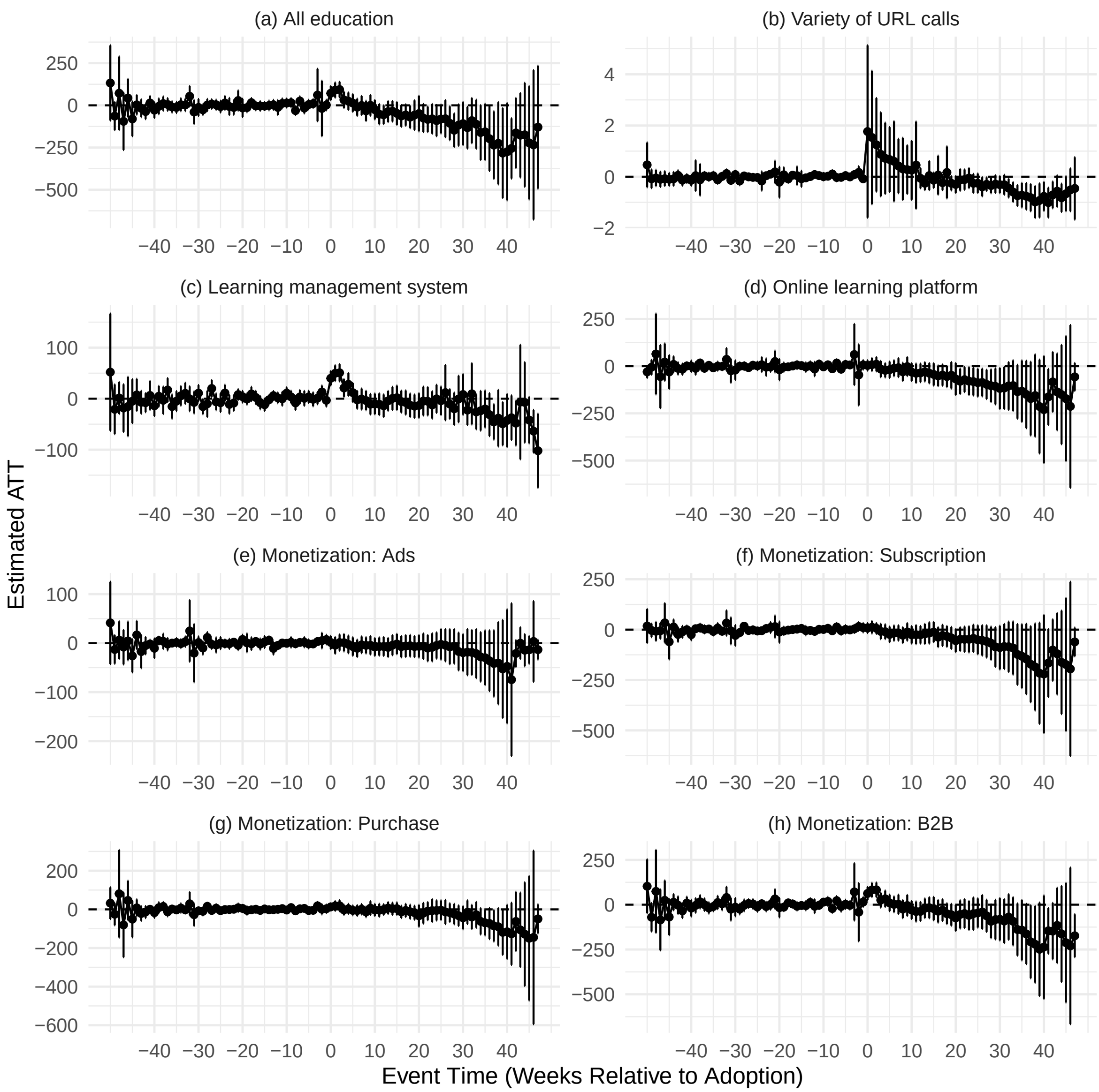

Figure F.10: Staggered DiD ATT on the weekly number of total traffic in education website, broken down by categories and monetization strategy.

Category	Variable	Mean	SD	5%	25%	50%	75%	95%
Main								
Education								
Monetization	Ads	20.1	190.9	0	0	0	0.0	48.0
	Subscription	53.6	344.5	0	0	0	0.0	215.0
	Purchase	49.8	341.8	0	0	0	0.0	187.8
	B2B	125.9	631.6	0	0	0	5.0	722.8

Table F.8: Summary statistics of education dependent variables by monetization strategy based on users' pre-adoption period

Next, we break down the drop in visits to education-related websites by the monetization method that the website uses. We classify educational websites by whether they rely on advertising revenues, subscriptions by learners, purchases by learners (e.g., of digital textbooks), or on B2B revenues (e.g., contracts with educational institutions).[24] We note that a single website can use multiple monetization methods, for example, selling advertising and offering subscriptions or offering subscriptions as well as one-time purchases. We show the pre-adoption summary statistics for the URL calls to education-related websites by monetization strategy in Table F.8.

	Dependent variable:			
	Ads	Subscription	Purchase	B2B
	(1)	(2)	(3)	(4)
ATT (weeks: 20-47)	−20.084	−109.946**	−59.846	−124.665**
	(20.363)	(55.582)	(41.389)	(58.473)
Pre-adoption avg.	20.111	53.556	49.790	125.892
Panelists	1886	1886	1886	1886
Weeks	52	52	52	52
Observations	98,072	98,072	98,072	98,072
Note:				*p<0.1; **p<0.05; ***p<0.01

Table F.9: Effect on Education-related websites broken down by monetization strategy

We show the corresponding ATT estimates for those variables in Table F.9. Column (1) shows a negative coefficient, though not significant, for advertising-funded educational websites. Since only 7 of the 107 domains we classify as educational rely on advertising,

[24]Classifications are based on browsing the website, ChatGPT queries and Google searches. We cannot rule out that our data does not fully capture all monetization methods for all websites, for example, we may not always have been able to detect a website relying on advertising if such ads are only shown after a user logged in.

the small size of this subset likely explains the lack of statistical significance. Column (2) demonstrates a significant negative effect for subscription platforms, which is sizable relative to the pre-adoption average value. Column (3) find no significant effect for visits to educational websites that offer individual services for purchase. Column (4) shows a significantly negative effect for educational websites that typically sell to businesses, as opposed to consumers. These results are consistent with those shown in Figures F.10 (e)-(h).

In sum, our results regarding education-related websites suggest that LLMs are a good substitute for a variety of education-related activities and as such have the potential to hurt a wide range of education-related content providers. They also show that different types of revenue streams may suffer. While in this specific context, we do not find a significant effect on sites that monetize through advertising, we demonstrate that sites that monetize through subscriptions and those engaging in business-to-business sales suffer a drop in traffic.

F.2 Additional Analysis on User-Generated Content Platforms

We confirm the results shown in Table 10 in Figure F.11.

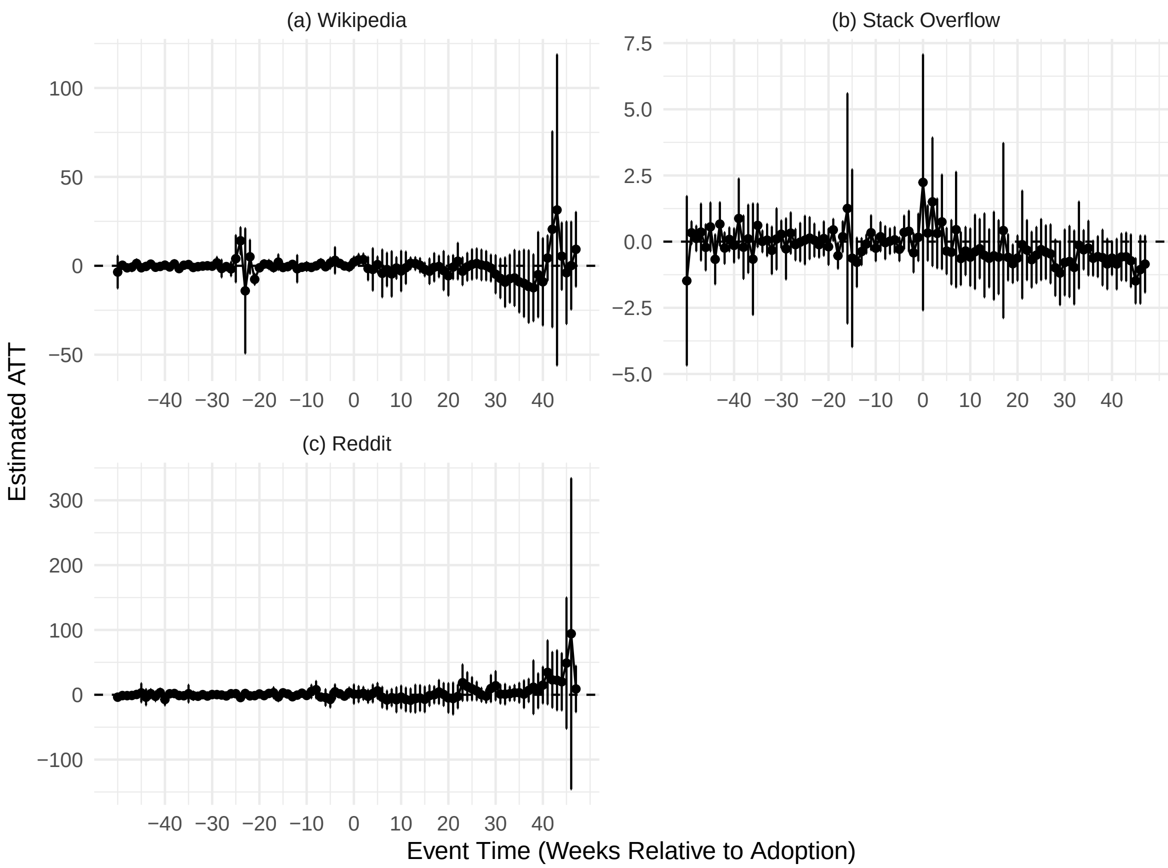

Figure F.11: Staggered DiD ATT on the weekly number of total traffic to knowledge-sharing platforms: Wikipedia, Stack Overflow and Reddit

F.3 Additional Analysis on Social Media Platforms

For our sample on social media, we rely on the 2034 panelists who had at least 10 URL calls to social media sites. We rely on Comscore's classification as social media sites and, among the top 50 URL hosts, we manually selected those that are indeed social media sites. Table F.10 shows the pre-adoption summary statistics of the number of URL calls on the most visited social media websites in the full sample. In addition to the six sites listed in Table F.10, we include Tumblr, Pinterest, Nextdoor, Pixiv, Onlyfans in the total "All social media" variable.

Category	Variable	Mean	SD	5%	25%	50%	75%	95%
Social Media								
Social media	All social media	179.7	927.4	0	0	5	64.0	914.0
	Facebook	85.8	343.0	0	0	1	19.0	472.0
	Instagram	40.3	538.8	0	0	0	0.0	88.0
	X	18.1	223.6	0	0	0	0.0	32.0
	Discord	11.0	89.3	0	0	0	0.0	15.0
	LinkedIn	9.5	179.6	0	0	0	0.0	13.0
	TikTok	7.5	520.2	0	0	0	0.0	13.0

Table F.10: Summary statistics of social media dependent variables based on users' pre-adoption period

Table F.11 reports the result aggregated for all social media sites and then separately for the six largest sites in our data. Throughout, we do not find a significant change to URL calls to these sites, suggesting that LLMs cannot effectively substitute for human social interaction and entertainment. We confirm these results visually in Figure F.12.

	Dependent variable:						
	All social media	Facebook	Instagram	X	Discord	LinkedIn	TikTok
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ATT (weeks: 20-47)	62.431	14.590	−14.437	−2.121	−3.151	22.949	45.717
	(75.321)	(18.726)	(46.566)	(18.080)	(4.206)	(30.431)	(69.720)
Pre-adoption avg.	179.690	85.804	40.296	18.088	11.004	9.465	7.518
Panelists	2034	2034	2034	2034	2034	2034	2034
Weeks	52	52	52	52	52	52	52
Observations	105,768	105,768	105,768	105,768	105,768	105,768	105,768
Note:							*p<0.1; **p<0.05; ***p<0.01

Table F.11: Effect on Social Media

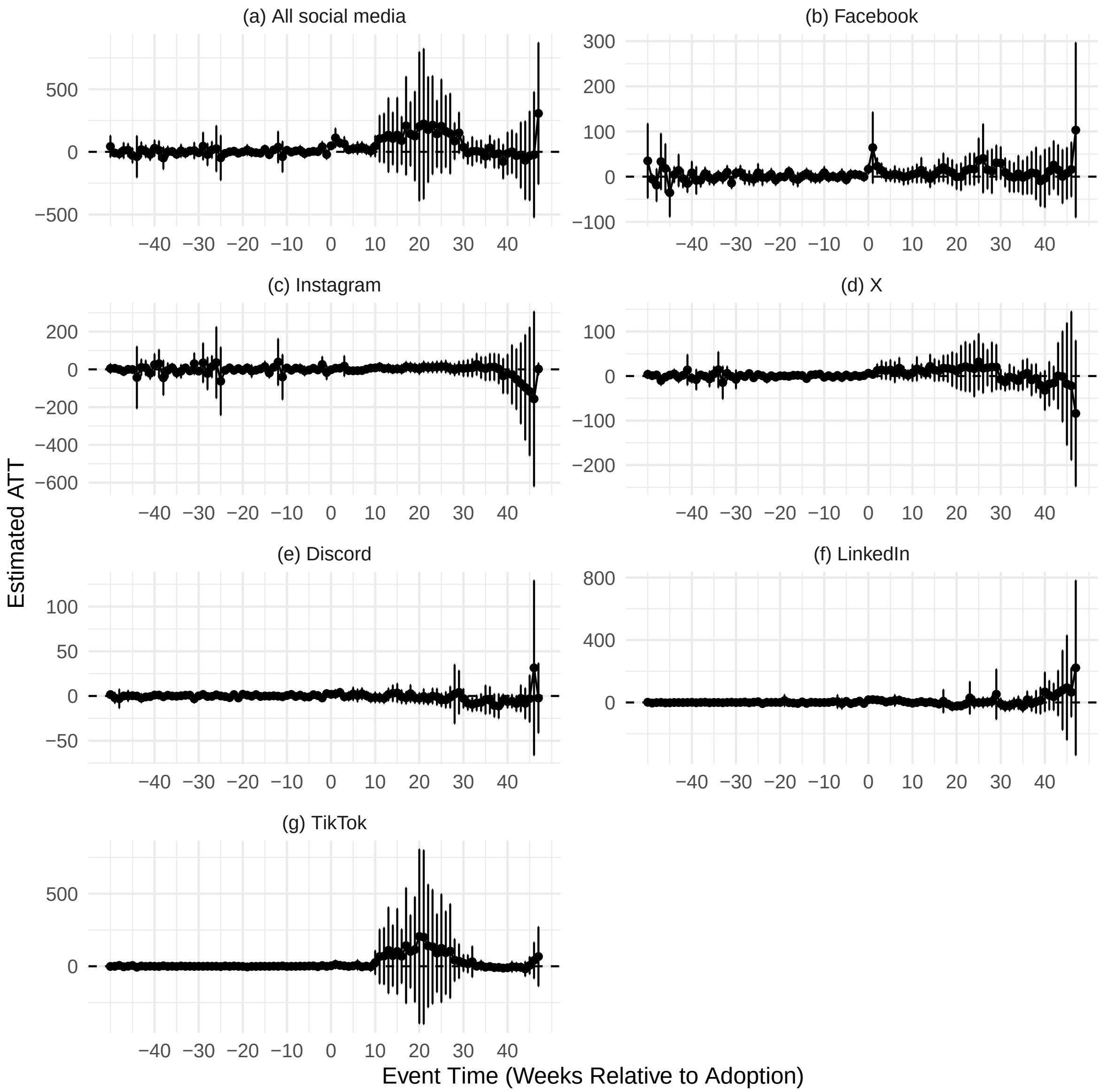

Figure F.12: Staggered DiD ATT on the weekly number of total traffic to social media platforms