The Impact of LLM Adoption on Online User Behavior*

Nicolas Padilla, H. Tai Lam, Anja Lambrecht, Brett Hollenbeck

August 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 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). The first two authors contributed equally to this work and are ordered reverse alphabetically. We thank Rafael Greminger and Yufeng Huang 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.

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, its creation and monetization by platforms and content creators. If users shift their online queries from visiting traditional search engines that funnel them to a variety of websites, toward LLMs that may be less likely to direct users to original content sites, the revenue model of such content creators may be jeopardized.

At the same time, LLMs may serve as a complementary tool, whereby the aggregation of information fulfills a distinct function that is supplementary to traditional search and online content consumption. For example, LLMs excel at understanding complex queries and pre-structuring 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. In addition, LLMs often lack the ability to facilitate authentic social interaction, but may ultimately lead to more demand for these.

Understanding whether LLMs act as substitutes or complements to traditional online search and content consumption is important for evaluating their impact on online content providers and their role in the broader web ecosystem.

In this paper, we first investigate how LLM adoption affects search behavior, a typical entry point for online activity. LLMs' ability to answer questions suggests potential substitution, but concerns around response quality may constrain this role, and potentially increase overall search volume. Second, to assess whether LLM use meaningfully substitutes for broader online activity, we evaluate its impact on adopters' online traffic. Third, we evaluate whether LLM adoption affects adopters' advertising exposure which would imply a reduced ability for websites to monetize content through advertising. Fourth, we narrow down to two sets of domain types and evaluate how LLM usage affects content consumption within those areas. The first of these is an analysis of educational websites, which is motivated by the fact that, like web users more generally, visitors to educational sites are often searching for answers, learning something new, or verifying facts, tasks which can potentially be substituted for by LLM usage (Humlum and Vestergaard, 2025). Then, in contrast, we examine user-generated content platforms to understand whether, and under what conditions, LLMs substitute for social interaction.

Our analysis is based on a comprehensive panel data set that tracks users' online browsing behavior over a period of up to two years in 2022 and 2023. This data set allows us to identify an individual's adoption of LLMs based on their usage patterns and to estimate the impact of LLM adoption on a range of online activities.

Our identification relies on a staggered difference-in-difference estimation where users who have not yet adopted LLMs (but do so later in our data period) serve as a control group to adopters. Further, we demonstrate that during the period of our data, visits to domains that offer email, retail or news are not affected by the adoption of LLMs and in our specification use those as controls for a user's overall online activity level.[1]

Overall, our results suggest that LLM usage substitutes for some – but not all – online activities but we do not find evidence for any complementary relationship. We document a significant decrease in online search activity. Online searches drop slowly, suggesting a period during which users learn about the use of LLMs, but eventually the level of online search is more than 20% lower than in the pre-adoption period. We differentiate between different types of searches. We observe a significant drop in searches using question words, and for both shorter and longer queries in terms of the number of words used in the query. However, LLM adoption does not negatively affect searches that consist of only a single branded term and so appear to be largely navigational (Blake et al., 2015). This pattern is consistent with LLMs not offering outlinks during the time period of our data, making it difficult for users to rely on LLMs as navigational tools.

Reduced search activity could indicate that users are turning to LLMs as a substitute

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

for broader web browsing. We evaluate how LLM adoption affects the total number of URL calls to websites. We find no effect on the most frequently used websites but a decrease in visits to websites that receive less traffic. This decrease in traffic to the long tail of websites suggests the smallest content providers are those most impacted by LLM adoption.

LLMs substituting for direct visits to websites, may significantly impact these websites' ability to monetize content through advertising. Indeed, we find that following LLM adoption, ad exposures drop significantly, especially for display ads. This effect is particularly pronounced for users with a high level of retail activity who may be particularly important targets for advertisers. Interestingly, we do not find a significant effect on search ads, suggesting that the reduction in search we observe may be related to searches that Google is less likely to monetize. This pattern aligns with our finding that navigational searches, where advertisers may be likely to place ads against searches for their brand names, do not drop significantly.

The degree to which LLMs substitute for existing websites likely depends on the nature of the content users seek. Some platforms primarily provide factual or objective information, while others offer complex or semi-structured knowledge, or support more subjective, experience-based, and social interactions. These distinctions shape the extent to which LLMs can replicate the core functionality of different sites and, in turn, displace user engagement. Accordingly, we analyze two distinct types of websites to evaluate how the impact of LLM adoption on browsing behavior depends on the ability of LLMs to substitute for a site's core function.

We first study educational websites and find that the number of URL calls to these sites, as well as the variety of domains users access, both decrease significantly. We then subcategorize education websites based on functional type and monetization method. We find that the drop in the number of URL calls holds for both learning management systems and online learning platforms. We find especially large negative effects on education-related websites that monetize through individual subscriptions, suggesting that LLM adopters find

LLMs to be a a close substitute for these services.

Next, we examine user-generated content platforms and find heterogeneous effects, depending on the type of information and interaction each site facilitates. Wikipedia visits remain unaffected, possibly because it already offers comprehensive and often moderated content, which may lead users to see little added value from LLMs, and provides factual information that could potentially be hallucinated by an LLM. In contrast, we observe a decline in visits to Stack Overflow, where users typically submit well-defined, complex questions and seek precise, internally consistent answers that LLMs can plausibly generate. However, Reddit — where queries are more ambiguous and responses may be shaped by personal or subjective perspectives — shows no significant change, possibly because these features make substitution with LLMs less effective.[2] We also do not find an effect on social media usage, including sites such as Facebook, Instagram, and TikTok, suggesting that LLMs cannot effectively substitute for human social interaction and entertainment.

Across our analyses, we find evidence that the adoption of LLMs can significantly alter users' online behavior. At one end of the spectrum, when online tools fulfill a purely logistical role and provide no additional informational value, such as when users rely on search engines as a navigational tool, we find no evidence that LLMs act as effective substitutes. In contrast, our results suggest that LLMs do, however, substitute for traditional online tools when users collect objective information on or seek answers to clearly specified queries. At the other end of the spectrum, when users seek answers to ambiguous problems, potentially characterized by subjective preferences, or engage in social interactions, we find no evidence that LLMs meaningfully replace traditional means of interacting online.

Our findings have implications for online content providers, for GenAI firms, and for policy makers. For online content providers, they highlight the risk that AI platforms might use content to train their models but generate little traffic, harming online content providers'

[2] Our findings related to Stack Overflow and Reddit align with prior research by Burtch et al. (2024).

ability to monetize content through either subscriptions or advertising.[3] Our results suggest that LLMs substitute for some traditional online activities and suggest that online content providers may benefit from considering alternative revenue models, such as charging LLMs for access to content.

At the same time, our results matter for GenAI firms that currently benefit from accessing and training their models on a wide range of content, often for free. Ultimately, online content providers will only have an incentive to continue contributing high quality content if they can continue monetizing their insights – which may include charging for access to content.

In consequence, our results have implications for the structure of content creation on the World Wide Web and, in turn, for policy makers. A large part of the digital economy relies on the creation of online content that enables small creators to thrive and allows for a diversity of viewpoints to emerge. Without being appropriately remunerated for content production, such sites may no longer be incentivized to generate content and contribute new insights. As such, the question for policymakers becomes which level of future content production is desired and whether regulatory intervention will be required to ensure such outcomes.

Our results further contribute to the debate around copyright and fair use policy. One critical dimension to determine fair use is the degree to which usage of the original work substitutes for and threatens the income of the original creator. Our results suggest that, at least during the period of our data, LLM usage replaces online content consumption elsewhere in some but not in all areas of the World Wide Web with evidence suggesting that such effects are most pronounced for search and smaller websites and do not exist for simply navigational queries or in the context of social interactions. While during our data period, search engines themselves had largely not yet incorporated LLMs as part of the search process and LLMs did not yet provide outlinks, our results still speak to the degree to

[3] Such concerns have led Chegg, the website providing homework answers and inexpensive textbooks to students, to sue Google over scraping its answers for Google's AI overview tool. See `https://futurism.com/the-byte/google-lawsuit-ai-overview`, accessed April 9, 2025.

which integrated LLMs would cannibalize traditional search offered by the same platform. At the same time, integration of LLM tools into traditional search engines may possibly lead to more pronounced effects on publishers as even users who do not actively navigate to an LLM will easily access AI-generated results at their traditional search engine.

2 Related Literature

Our results contribute to five 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).[4] Together, these results demonstrate that snippets of content – news or music – can plausibly serve as advertising, thereby constituting a complementary good (Becker and Murphy, 1993). By contrast, Seamans and Zhu (2014) demonstrate that the entry of Craigslist lead 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 can 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)

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

presents descriptive evidence for the demand for LLMs, suggesting that new models experienced rapid initial adoption that stabilized within weeks. 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 on 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 find (see also Burtch et al. (2024)). The drop in display ad exposures we document suggests that such reduced visits can have direct economic consequences for online content providers.

Third, our research relates to the literature on the usage and effects of LLMs in occupational settings where the use of ChatGPT is widespread, especially among younger and less-experienced workers (Humlum and Vestergaard, 2025). However, the degree to which ChatGPT substitutes for jobs varies across tasks with stronger effects on writing and coding relative to manual-intensive skills (Demirci et al., 2025). More generally, results support that AI can increase work-force productivity (Brynjolfsson et al., 2025; Cui et al., 2024) though its usage and ability to increase productivity varies across types of tasks (Handa et al., 2025; Eloundou et al., 2024; Dell'Acqua et al., 2023) and total productivity gains could be modest (Acemoglu, 2025). In contrast to this literature focused on productivity in occupational settings, our research focuses on the adoption of LLMs in a broad range of consumer contexts and emphasizes how the adoption of LLMs affected third-party service and content providers.

Fourth, our findings are relevant for the evaluation of "fair use" under U.S. copyright law. In particular for the fourth factor U.S. Copyright Act (nd), which asks whether a use affects the market for the original work. Gans (2024) argues that, for large AI models where upfront licensing is impractical, an ex-post fair use regime—with compensation paid after training—can improve social welfare compared to rigid copyright rules. 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. A key dimension in assessing fair use is substitution versus complementarity. Our finding—that LLM adoption significantly reduces online search and traffic to niche sites—therefore offers evidence that, absent appropriate licensing, the substitutions towards LLMs could weigh against a finding of fair use.

Fifth, 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 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). 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 has 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 live updating or external links.[5] These models were employed across a wide range of personal and professional applications. Typical use cases included 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 debugging as well as data analysis. Beyond these tasks, academics used LLMs to support their research, including for literature review, improving writing and summarizing insights.

A further context where LLMs provided 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 or grammar. In language learning, LLMs can support translations or practice conversations. Further, students using LLMs for academic cheating such as for essay writing and take-home exams became a significant concern. One survey suggests that in the academic year 2022-2023, 30% of US college students used ChatGPT for schoolwork while 46% of this group

[5] 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.

says they somewhat or very frequently used the tool.[6]

The broad range of LLM use cases suggests that LLM use can potentially substitute for traditional web browsing which can have negative implications for online content providers. This, in turn, has lead to significant concerns among publishers about their continued ability to monetize online content.[7]

3.2 Data

We rely on 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,[8] 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 the 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 until the end of 2023 but in our analysis focus on their activity between November 2022 and October 2023 and drop any traffic occurring

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

[7] See, for example, `https://www.fastcompany.com/90941730/the-rise-of-chatgpt-can-we-save-digital-publishers-from-the-age-of-generative-ai` and `https://apnews.com/article/nyt-new-york-times-openai-microsoft-6ea53a8ad3efa06ee4643b697df0ba57`, accessed May 8 2025.

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

outside that one-year period.[9] We then construct a balanced weekly panel of users by subsetting the Comscore data on a number of criteria. First, we filter to users who had at least four days with any web traffic in each month from November 2022 to October 2023 (our sample period).[10] This removes users who have effectively left the sample but also 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.[11] This leaves us with a sample of 2,041 households that adopted LLMs during the relevant time period, with households adopting LLMs between December 5th, 2022, and December 17th, 2023. Third, for our analyses that are focused on groups of websites (e.g., education-related) or specific websites (e.g., Stack Overflow), we restrict attention to users who have at least 10 URL calls on the focal groups of websites over the entire period.

3.3 Descriptive Statistics

We aggregate our data to a weekly level. Table 1 summarizes the number of observations in our sample. In total, our sample includes 2,041 panelists that we use for our main analyses. We observe each panelist for 52 two weeks which gives us a total of 106,132 week-user observations. Users adopt LLMs at different points during our data period. As mentioned previously, we define as a user's adoption date the first of the three consecutive weeks of at least one instance of LLM usage. For our main sample, 64,543 observations fall into the

[9]We later in this section explain in detail how we define adoption. The inclusion of users who adopted after our observation period allows us to have a control group of no-yet-adopters for those who adopted towards the end of that period.

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

[11]We also explore robustness of our results to different numbers of consecutive weeks of LLM usage when defining our sample in Web Appendix A. As expected, the more consecutive weeks we require to define adoption, the smaller the sample size, leading to wider confidence intervals.

pre-adoption period and 41,589 observations fall into the post-adoption period.

Further, in Table 1, we display the corresponding summary statistics for the sub-samples we use in our analyses. We restrict these sub-samples to panelists who have at least ten URL calls in the entire observation window on the relevant main dependent variable. We find that 1886 panelists had at least 10 URL calls to education-related websites during our sample period.[12] 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 ten times during our observation period. 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.[13]

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

Variable	Main	Education	Wikipedia	Stack Overflow	Reddit	Social media
Unique panelists	2,041	1,886	1,634	287	1,488	2,034
Weeks in data	52	52	52	52	52	52
Total observations	106,132	98,072	84,968	14,924	77,376	105,768
Pre-adoption	64,543	59,706	51,120	8,246	46,245	64,282
Post-adoption	41,589	38,366	33,848	6,678	31,131	41,486

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

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.

[12]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

[13]In addition to those sites listed in Table 2, we include Tumblr, Pinterest, Nextdoor, Pixiv, Onlyfans.

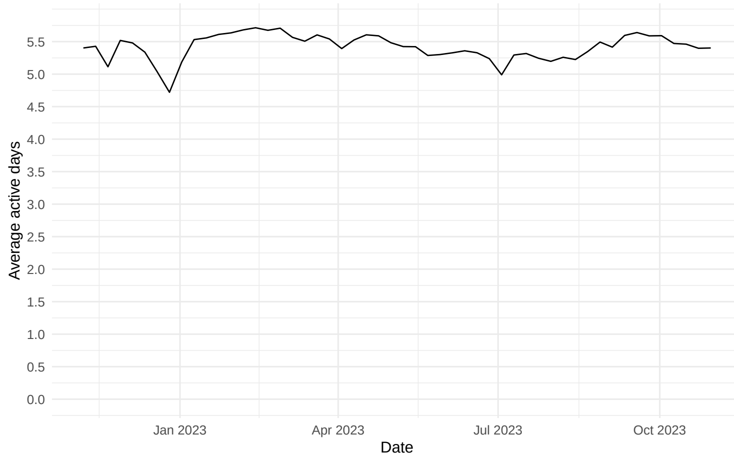

Figure 1: Average number of active days per week.

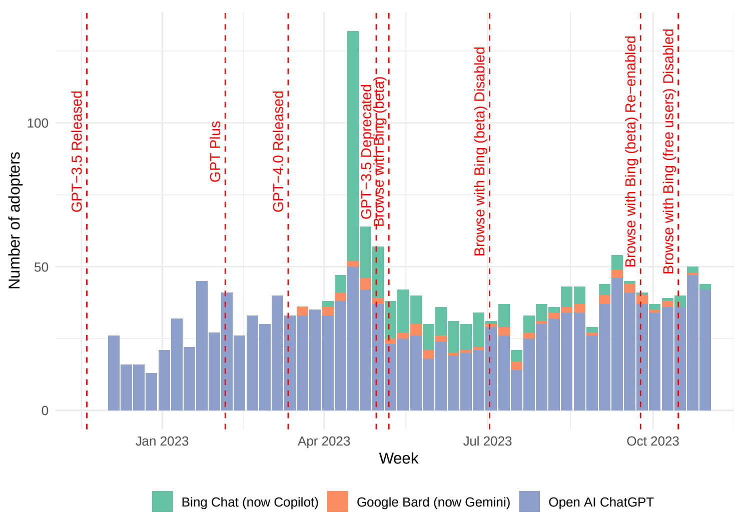

Figure 2: Adoption per week (first week of three consecutive weeks of usage).

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 a significant variation in the adoption of LLMs over time.[14] 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

[14]There is one spike in adoption in the week of April 17, 2023, which from our data, corresponds to users adopting Microsoft's Bing Chat (www.bing.com/chat, now called Copilot). It was reported that Microsoft had 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.

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.1).

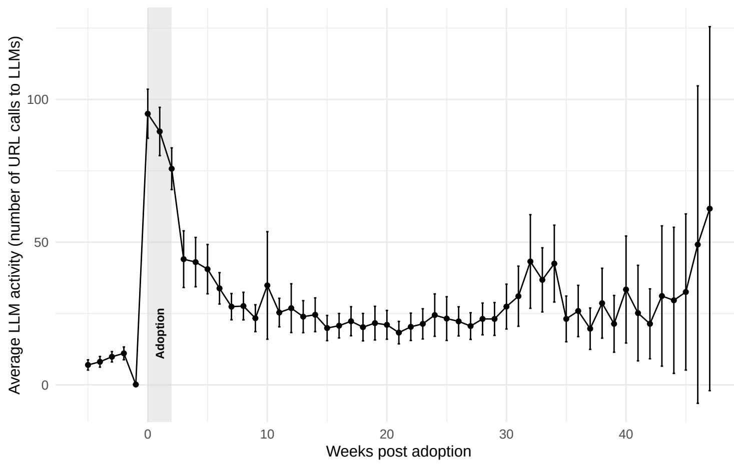

Figure 3: Average number of URL calls to LLMs per post-adoption week. Bars indicate standard errors.

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 help illustrate the amount of data available to estimate average treatment effects on the treated at different points relative to adoption (we elaborate more on our identification discussion in 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 also fewer not-yet-treated units contributing to comparisons (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.

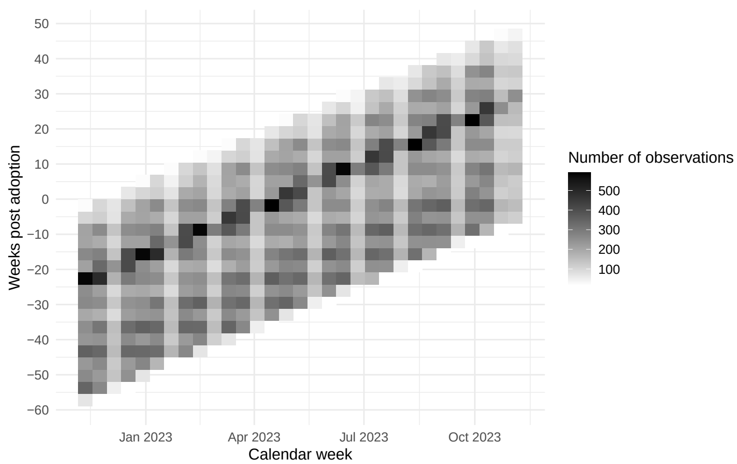

Figure 4: Number of observations per calendar week and post-adoption week.

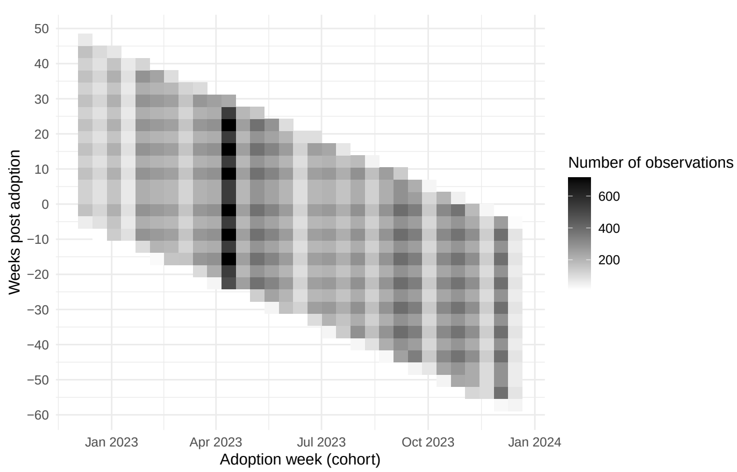

Figure 5: Number of observations per adoption cohort and post-adoption week.

3.3.2 Website traffic

Table 2 summarizes the data descriptives of our user-level weekly panel based on users' pre-adoption period. We first identify in our data search activity. This includes searches through Google, Bing, and Yahoo. On average, users made 33 online searches in a week, 18 of which were using Google. Among those, four searches used at least one of the following question terms: how, what, where, when, which, who, why.

We identify searches that were likely to be navigational (Blake et al., 2015). To do so, we identify the top 2,000 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 as the search term and no other terms. For example, if a user uses the search term 'reddit', that would be classified as a navigational search. On average, users made one navigational search a week. We likewise identify searches that use a navigational term as well as one or more other terms, such as a user searching for 'reddit buy guitar.' Users made, on average, two searches a week that used a navigational term as well as another term. We also differentiate between long and short searches. Short searches consist of two or fewer words, and long searches consist of six or more words. These correspond, respectively, to the first and fourth quartiles of word counts in search queries. We find that users make about seven long and about five short searches during a week.

We evaluate the degree to which users browse the broader web. We exclude from this analysis searches, URL calls of email, retail, and news sites since these categories will serve as a control as well as URL calls that we identify as advertising impressions. Further, we exclude background traffic by classifying each URL host as either 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). In total, we find 4325.1 URL calls per user and week. Out of this total number, 2928.0 can be attributed to the 500 websites that receive the most amount of traffic, whereas 1397.1 comes 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 then identify the traffic that was referred to from a search engine. We refer to this as referred traffic. Specifically, if a consumer visits a search engine and clicks on a link, this 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].

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

As we explain in more detail in Section 4, we control for users' baseline internet usage by accounting in our estimation for their activity levels in email, retail, and news. In order to identify email, retail, and news activity, we rely on the corresponding Comscore classifications.[17] On average, a user made 649 URL calls related to email, retail and news in a week with the majority of this traffic coming from email and retail activity.

We turn to user activity on education-related websites. The 1886 users who visit an

[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

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

Category	Variable	Mean	SD	5%	25%	50%	75%	95%
Main								
Search	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 traffic[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
Ads	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
Category	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
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
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
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

[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

education-related website at least ten times during our observation period make, on average during a week, 170 URL calls to these domains and visit, on average, about one domain in a 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 as two important categories. Learning management systems are platforms that provide digital infrastructure that is populated by educational institutions or teachers, such as Canvas and Blackboard. On average, a user makes 62 calls a week to a learning management system. Online learning platforms provide content and target either educational institutions who then purchase access for their learners or learners directly. This latter group includes 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. On average, our data records 56 URL calls to online learning platforms per user and week.

We further classify educational websites by their monetization methods. For each website, we identify whether it relies on advertising revenues, on subscriptions by learners, on purchases by learners (e.g., of digital textbooks), or on B2B revenues such as when online learning management systems contract with educational institutions.[18] We note that individual websites may monetize their offering in more than one way. In our data, we document user-level weekly average of URL calls to educational websites monetizing through ads of 20, to websites offering subscriptions of 54, to those that sell services or products for purchase of 50, and to those relying on business-to-business revenues of 126.

We then turn to user-generated content platforms where we differentiate between knowledge-sharing platforms such as Wikipedia, Stack Overflow and Reddit, and social media platforms. Users who visited Wikipedia at least 10 times during our data period made on average about 14 URL calls to the site a week. Users in our Stack Overflow sample had one weekly URL call to the site whereas users in our Reddit sample had 14 weekly URL calls to the site. By

[18]Classifications are based on browsing the website, ChatGPT queries and Google searches.

contrast, users accessing social media platforms, made 180 URL calls on average, though there is significant variation across the individual platforms with users on Facebook being most active and TikTok users being least active, potentially because TikTok users are more likely to instead use mobile devices.

4 Estimation and Identification

Our aim is to estimate the causal effect of LLM adoption on search, web-browsing and ad exposure over time using our data consisting of weekly activity, for individuals over one year.

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 online activity of a user 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 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 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.

To address these challenges, we follow Callaway and Sant'Anna (2021) to estimate cohort- and time-specific average treatment effects on the treated ($\text{ATT}_{g,t}$), 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 the 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 parameter of interest is

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

This captures the causal effect $t - g$ weeks after adoption 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.

4.2 Identification

Two features of our approach aid in identification: To ensure our results are not biased by users' overall online activity, we control for URL calls to email, retail, and news websites. These are categories we expect to be unaffected by LLM adoption. Absent these controls, we might find a spurious positive correlation between LLM adoption and other types of activity of interest, such as search behavior, due to fluctuations in internet activity that are unrelated to LLM adoption. We show in Figure 6 and Table 3 that LLM adoption had no effect on our 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 may correlate

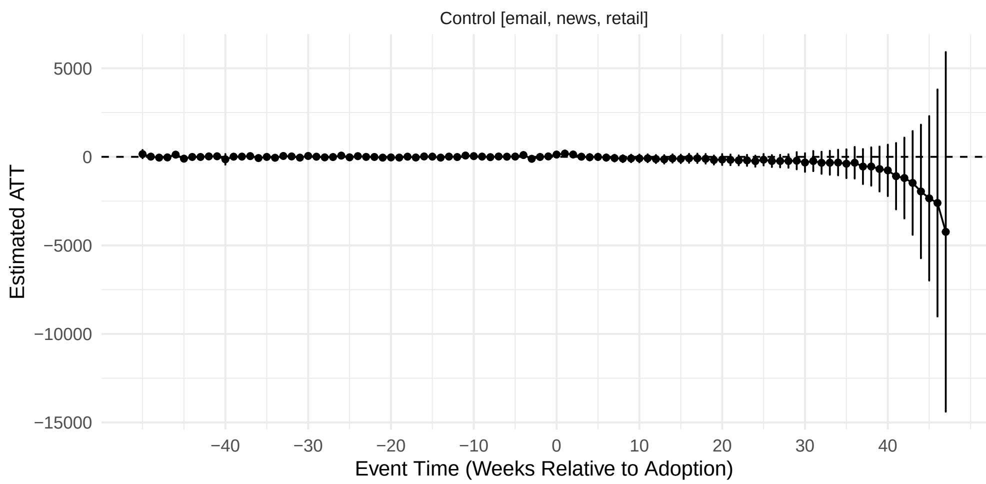

Figure 6: Staggered DiD ATT on the weekly number of URL calls in the control websites (email, news, and retail).

with general desktop usage during those weeks.

	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).

To account for week-specific activity shocks common to all users, we include a control group. Our data satisfy the standard overlap condition, that is that at each calendar time and for each relative period around treatment there are both treated and untreated units to compare.

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

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 of violations of parallel trends, which under the staggered difference-in-difference settings manifests as pre-period effects that are systematically positive or systematically negative (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. While we are unable to test for the presence of no anticipation directly, our later estimation results document a delayed and not an immediate effect of LLM adoption on online activities, suggesting that this is not a concern in our setting. Further, our results are robust to varying the number of weeks before treatment that individuals may anticipate treatment and change their behavior beforehand (see Web Appendix A).

Our approach is robust to several potential threats to causal identification:

- **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 in our estimation as long as the unobservables do not affect outcome trends. This is because each $\text{ATT}_{g,t}$ is estimated using a two-by-two difference-in-difference setup. The parallel trends assumption accommodates such level differences while requiring similar growth patterns across cohorts.

- **Heterogeneous Treatment Effects:** If users who adopt earlier experience different treatment effects than late adopters, our method explicitly accommodates this 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 show how cohorts differ in their treatment effects over time in Web Appendix F.

- **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.

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.

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 among 2,448 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 for each estimate.

The inclusion of users' levels of traffic to email, retail, and news websites as covariates improves estimation efficiency, in addition to controlling for potential activity bias.

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 compute 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, Browsing Activity and Advertising Exposure

5.1 Online Search

We first explore the effect on overall online search activity. Figure 7 and Table 4 summarize the results.

Figure 7 (a) plots the individual weekly estimates over time since adoption when controlling for a user's email, retail and news activity. We note that since we measure activity starting from November 2022, but we only have adopters since December 2022, we can only measure treatment effects for up to 47 weeks post adoption. The figure demonstrates that following LLM adoption, activity briefly spikes before exhibiting a slow but steady decline. The initial spike is consistent with our previous results on control activity and likely reflects increased activity related to the timing of LLM adoption. Importantly, part of this spike stems from our definition of adoption, which requires LLM usage in three consecutive weeks —implying that during adoption, individuals have to be digitally active. While our activity controls absorb part of this variation, they may not fully account for instances when users temporarily deviate from typical patterns —for example, days where users concentrate activity on tasks not captured by controls like email, news, or retail browsing. As such, the initial spike likely reflects a combination of true adoption-related activity and spurious positive correlations between all activity types. Consequently, this pattern suggests that our estimates could be an upper bound of the true effect. Notably, the effect of LLM adoption on browsing behavior appears to set in slowly over time, rather than lead to an abrupt reduction, suggesting that users learn over time how to substitute activity with LLMs.

	Dependent variable:				
	search_all_n				
	(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

In Figure 7 (a), we note 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. This is consistent with our discussion in Section 3.3 about the lack of precision for ATTs after 35 weeks or more 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, that is, around late November 2023 (see also

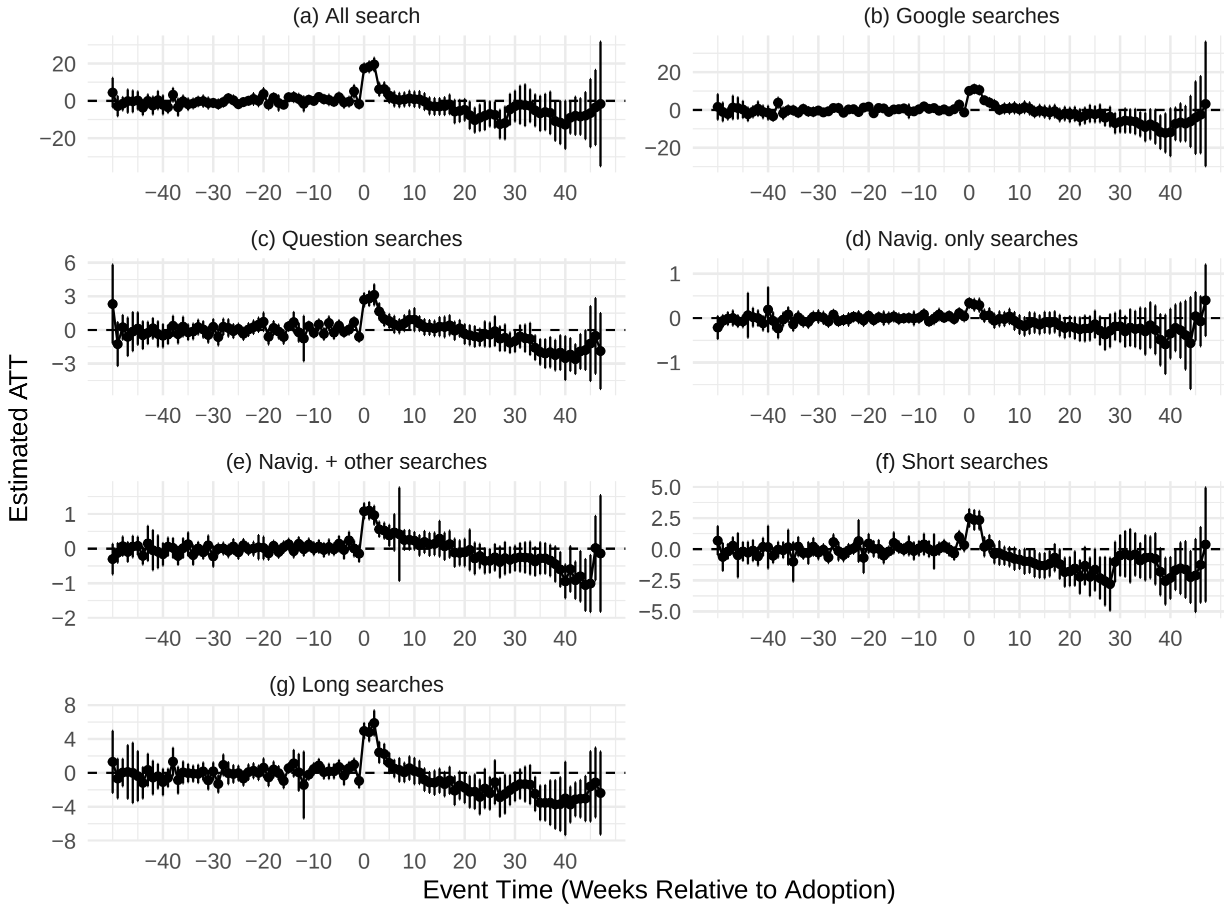

Figure 7: Staggered DiD ATT on the weekly number of total searches and the number of searches by category.

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 time post adoption, the estimated effects may capture no differences between treated and control units. Note that this effect only occurs for effects measured 35 weeks after adoption, where we have less variation on adoption timing for both control and treated units.

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 likewise 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 category control activity (retail, news or email). The results continue to hold.

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 where a search engine facilitates browsing (Blake et al., 2015). Figure 7 (c) and Column (3) of Table 5 focus on searches using a question words. 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

[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.

Column (4) of Table 5, demonstrate that searches that at 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. 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.

We conduct a series of robustness checks on these estimates. We find that our estimates are largely consistent with those obtained from other model specifications, including TWFE and Poisson Pseudo-Maximum Likelihood regression (see Web Appendix E). We also find our results to be largely robust to other sample criteria and LLM adoption definition (see Web Appendix A) in the expected direction: smaller samples leading to wider confidence intervals.

5.2 Browsing Activity

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

	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

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

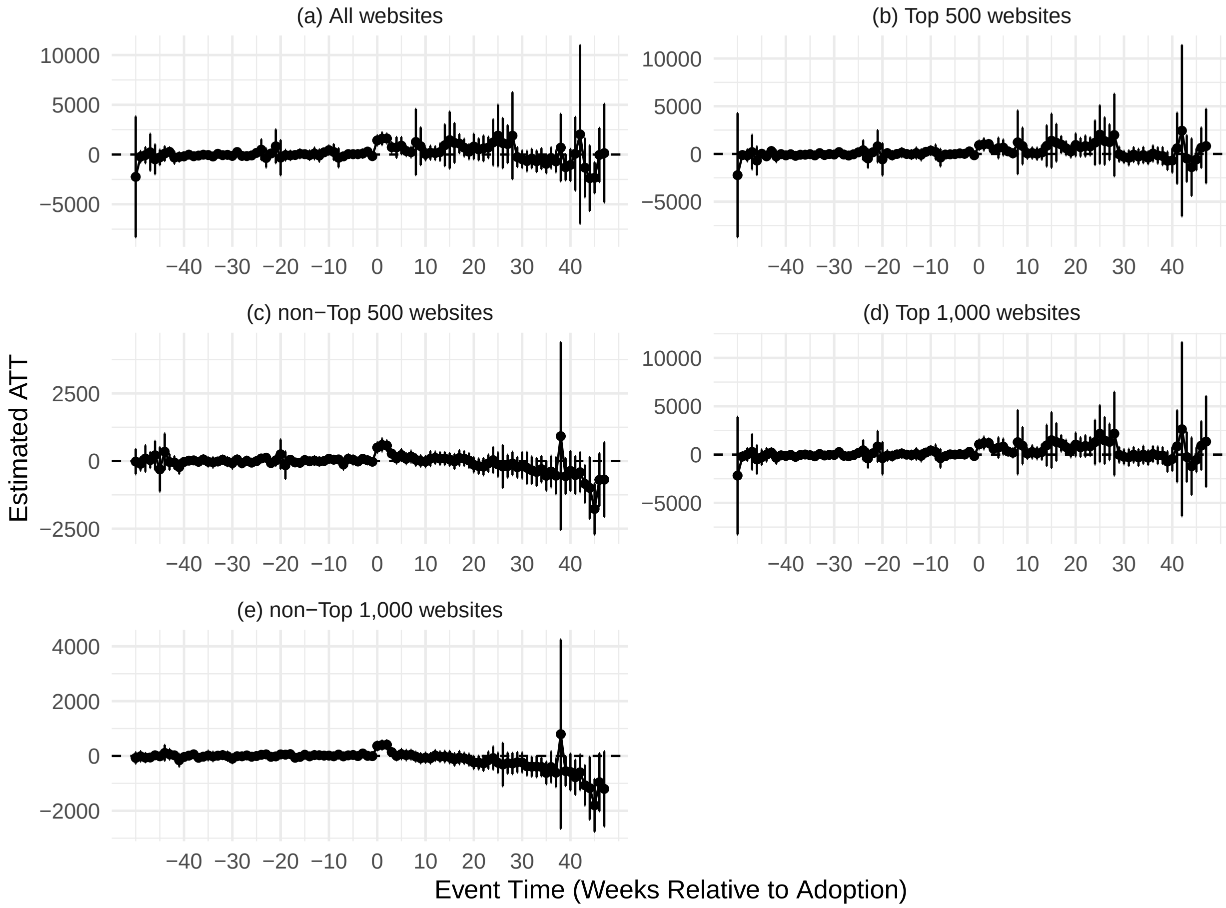

Figure 8: Staggered DiD ATT on weekly number of URL calls for top websites and all other websites.

	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

between the 500 websites that received the largest amount of traffic and all other websites. We do the same for the top 1000 websites and the rest. Figure 8 (b) again shows no evidence for a drop in URL calls for the 500 websites that received the largest amount of traffic. Figure 8 (d) mimics these findings for the top 1000 websites. Table 6, Columns (2) and (4), confirm these patterns.

Figure 8 (c) reports results for all other websites not in the top 500 websites. It demonstrates that URL calls to these smaller websites drop significantly after week 20. Table 6, Column (3) confirms the significant drop in activity for weeks 20 to 47. The estimated coefficient of -366.8 relative to the pre-adoption mean of 1397.1 implies that at the mean, visits to smaller websites dropped by 26%.

We then in Table 7 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) 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. As expected, the results mirror 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.

	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.

In Table 8, we focus on traffic referred to 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. In Columns (2) to (5), we 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 not 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 7 and 8 demonstrate that while in the aggregate 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.

5.3 Advertising Impressions

A substitution in visits to websites has the potential to significantly affect their ability to monetize through advertising. We turn to directly measuring the impact of adoption of LLMs on advertising exposures to users.

	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 9: Effects on advertising exposure

Figure 9 and Table 9 demonstrate the effect on advertising exposures. Figure 9 (a) and Column (1) in Table 9 demonstrate a significant drop in ad exposure across all ads that we identified in our data. Again, this drop sets in approximately 20 weeks post adoption. Figure 9 (b) and Column (2) in Table 9 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

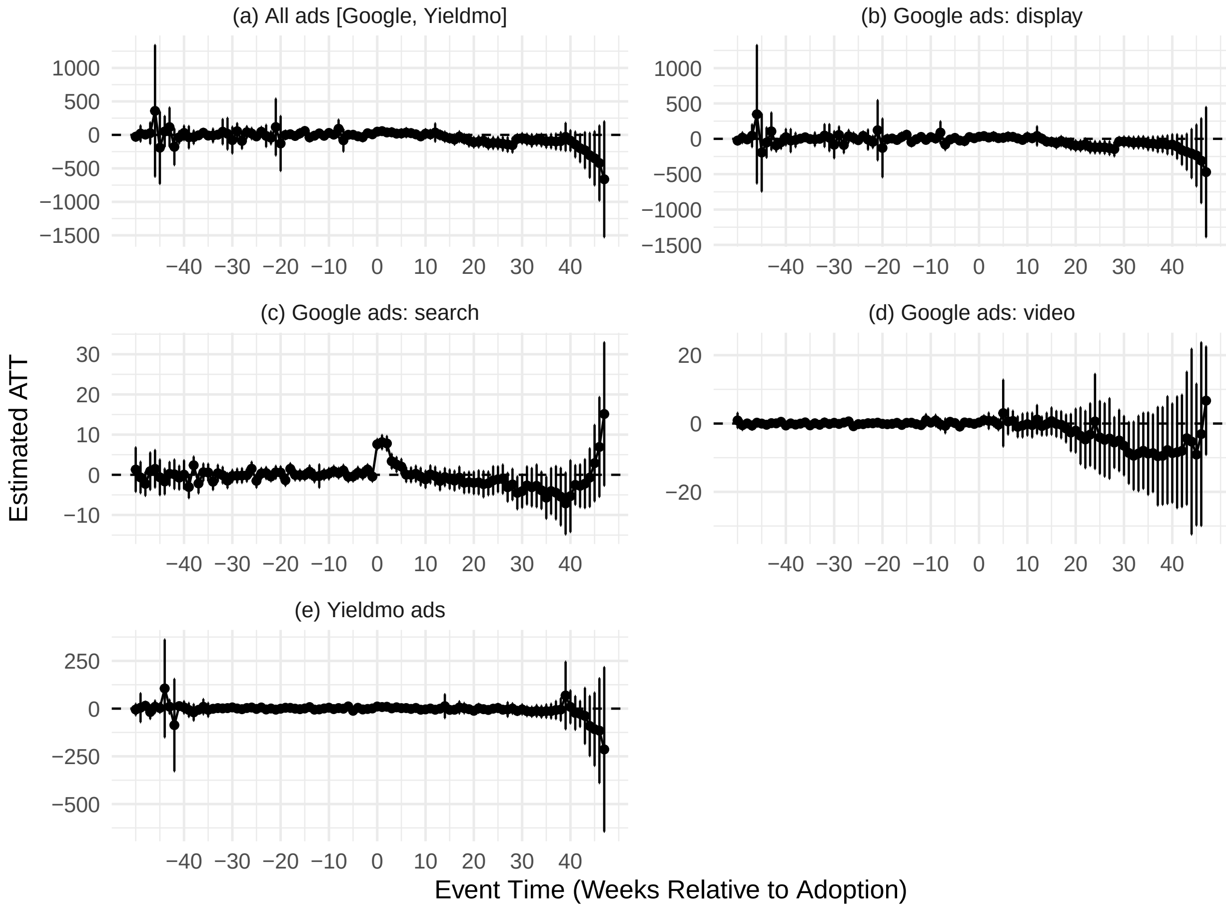

Figure 9: Staggered DiD ATT on the weekly number of URL calls to ad servers.

Column (3) in Table 9 suggest no significant drop in the number of advertising impressions for Google search ads. This may be a result of ads generally being more likely to be displayed in response to searches that are less likely to be affected by LLM adoption, such as navigational searches, and are less likely to be shown after searches that suffered a more severe drop.

Figure 9, Panels (d) and (f), and Table 9, Column (4) shows 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.

5.3.1 Heterogeneity by Retail Activity

We next explore heterogeneity in the effect of LLM adoption on advertising exposures by users' level of retail activity. To do so, we classify users based on their pre-adoption average level of retail activity, bucketed into three terciles based on their average weekly number of URL calls to retail domains: Low corresponds to $[0, 13.5]$, Mid to $[13.5, 75.1]$, and High to $[75.1, \infty]$ calls to retail domains, respectively.

We explore the effect of LLM adoption on ad exposures separately for each of these three levels of retail activity. Columns (1) to (3) in Table 10 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 and Columns (4) to (6) shows that this pattern holds for Google display ads. These results suggest that those consumers who are likely to be most valuable to both retailers that advertise and to online content publishers suffer the greatest drop in advertising exposure.

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

6 Effect on Specific Domain Types

6.1 Education-related websites

We turn to education-related websites as one category of online content providers. There are at least two reasons why education-related websites may be particularly affected by the adoption of LLMs. First, many students adopted LLMs and especially ChatGPT early.[20] Second, it is easy to see how educational tasks can be completed more quickly using LLMs.

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

This could include using LLMs to summarize insights for an essay or to complete an entire homework assignment. We, thus, explore the effect of the adoption of LLMs on educational content providers. This analysis is based on the 1886 panelists who visited any of the education-related websites at least once during our data 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 11: Effect on Education-related websites traffic

	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 12: Effect on Education-related websites broken down by monetization strategy

Table 11 demonstrates the effect on education-related websites. Table 11, Column (1), uses the traffic to those websites as the dependent variable, demonstrating a drop in traffic across the entire category. Table 11, Column (2), focuses on the number of unique domains a user visited in the education category and demonstrates a drop in the variety of educational websites users visit. Table 11, Column (3), focuses on learning management systems and Table 11, Column (4), focuses on online learning platforms. Again, both demonstrate similar patterns. Figures G.6 (a)-(d) in Web Appendix G.1 confirm these patterns by plotting individual estimates over time.

Next, we break down the drop in visits to education-related websites by the monetization method that the website uses. 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. Table 12, 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, the small size of this subset likely explains the lack of statistical significance. Table 12, Column (2), demonstrates a significant negative effect for subscription platforms, which is sizable relative to the pre-adoption average value. Table 12, Column (3), find no significant effect for visits to educational websites that offer individual services for purchase. Table 12, 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 G.6 (e)-(h) in Web Appendix G.1.

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.

6.2 User-Generated Content Platforms

We analyze the extent to which LLMs act as substitutes for various user-generated content platforms. Specifically, we consider three prominent knowledge-sharing platforms — Wikipedia, Stack Overflow, and Reddit — as well as social media more broadly. These platforms span a continuum in the type of information they provide: from factual and objective (Wikipedia), to complex and semi-structured knowledge (e.g., programming advice on Stack Overflow), to more subjective and experience-based content (e.g., discussions on Reddit). At the far end of this spectrum lies social media, which primarily facilitates the sharing of personal opinions, experiences, and social interactions.

LLMs have the potential to serve as an alternative to traditional search engines when users seek answers to questions online, particularly given the substitution effects shown in Section 5.1. However, the degree to which LLMs can effectively replace these platforms likely depends on the nature of the information being sought. For example, factual queries such as "How many albums did Ozzy Osbourne record?" are susceptible to hallucination by LLMs (Huang et al., 2025; Wiggers, 2023), raising concerns about their reliability for such tasks. Conversely, questions grounded in personal experience or social context — such as those typically addressed through social media — are less amenable to LLM-based responses, as users may be explicitly seeking diverse human perspectives. In between these extremes lie more complex, open-ended questions, such as those found on Stack Overflow, where multiple technically correct solutions may exist. These types of queries may be particularly well-suited to LLM-generated content, provided the responses are coherent and contextually appropriate.

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

We begin by analyzing substitution effects across knowledge-sharing platforms. Table 13, Column (1), shows no significant effect of LLM adoption on URL calls to Wikipedia, possibly because it already offers comprehensive and often moderated content, which may lead users to see little added value from LLMs. In contrast, in Table 13, Column (2), we observe a decline in visits to Stack Overflow, where users typically submit well-defined questions and seek precise answers that LLMs can plausibly generate. However, in Table 13, Column

	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 14: Effect on Social Media

(3), shows no significant change in URL calls to Reddit, where queries are more ambiguous and responses may be shaped by personal or subjective perspectives, possibly because these features make substitution with LLMs less effective. These results are consistent with those from Figure G.7 in Web Appendix G.2.[21]

Table 14 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 signifcant 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 G.8 in Web Appendix G.2.

7 Conclusion

The broad adoption of AI tools has the potential to significantly reshape how consumers acquire information online as these tools may potentially serve as substitutes or complements to existing digital resources. As such, the adoption of AI tools also has the potential to reshape the economics of 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 the adoption of LLMs affects user online behavior. We focus on the set of users who we observe adopting LLMs and study how their behavior changes once they have done so.

[21] Our findings related to Stack Overflow and Reddit align with prior research by Burtch et al. (2024).

Our primary results suggest that concerns about LLMs substituting for web browsing may be well-founded, at least for a subset of online content provider. In particular, we find that after adopting LLMs, users make fewer searches in traditional search engines, including for question searches and both short and longer queries. This suggests that for these types of queries, LLMs act as a substitute rather than creating a pure expansion effect. It also suggests that as more online users adopt LLMs total searches through search engines are likely to fall significantly. We also find that while visits to websites that generally receive more traffic do not change significantly during our observation period, visits to smaller websites drop significantly which suggests that such websites are particularly vulnerable to substitution by LLM usage. In line with users making fewer visits to content producers' sites and instead obtaining the information they need within an LLM, we also find that ad impressions fall for LLM-adopters. This pattern has the potential to threaten the viability of some types of publishers and, in the long run, reduce the incentives for content to be produced at all. We find this may be a particular concern for education-related websites.

Interestingly, our results indicate throughout that the drop in online search or browsing does not materialize immediately following adoption but sets in about 20 weeks post adoption, suggesting that substitution is contingent on users' experience with LLMs and that economic effects for online content producers are likely to be somewhat delayed, relative to users' initial adoption.

There are several important limitations of our analysis. First, our sample ends at the end of 2023. LLM adoption and usage intensity has grown rapidly since that time, so the effects may be different – potentially larger and affecting a larger share of the population – than what we are able to identify. Our results should be seen as initial evidence of what may become longer-term trends. While our data is based on LLMs that are pre-trained and do not yet link out as newer models do, recent industry reports still suggest a significant drop in the amount of traffic going to publishers' sites for such newer models (Tollbit, 2024). Second, we classify a user as having adopted LLMs after using them in at least three consecutive weeks. We highlight that different definitions of adoption yield different effect size estimates. Third, 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 what types of behavior are most likely substituting for the changes we see in web browsing activity. Contemporary work by Handa et al. (2025) provides some evidence here. Lastly, our work studies a time period when LLMs did not link out to websites and did not collect and provide real-time information in their training data. At the same time, more integrated search tools such as Google's AI overview were not available during the period of study. As these capabilities expand, the ultimate impact on browsing behavior may change as well. Nevertheless, we provide a range of results supporting that LLM adoption is reshaping how consumers interact with the internet in ways that promise to have major implications going forward.

References

Acemoglu, D. (2025). The simple macroeconomics of ai. *Economic Policy*, 40(121):13–58.

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.

Brynjolfsson, E., Li, D., and Raymond, L. (2025). Generative ai at work. *The Quarterly Journal of Economics*, page qjae044.

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.

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.

Cui, Z. K., Demirer, M., Jaffe, S., Musolff, L., Peng, S., and Salz, T. (2024). The effects of generative ai on high skilled work: Evidence from three field experiments with software developers. *Available at SSRN 4945566*.

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.

Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., and Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (24-013).

Demirci, O., Hannane, J., and Zhu, X. (2025). Who is ai replacing? the impact of generative ai on online freelancing platforms. *Management Science*.

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

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2024). Gpts are gpts: Labor market impact potential of llms. *Science*, 384(6702):1306–1308.

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

Gans, J. S. (2024). Copyright policy options for generative artificial intelligence. Technical Report w32106, National Bureau of Economic Research.

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.

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.

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.

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.

Tollbit (2024). State of the bots - q4 2024. *Journal of the American statistical Association*, 83(401):123–127. (accessed on April 9, 2025).

U.S. Copyright Act (n.d.). 17 U.S.C. §107(4): Effect of the use upon the potential market for or value of the copyrighted work. Cornell Law School Legal Information Institute.

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: 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.

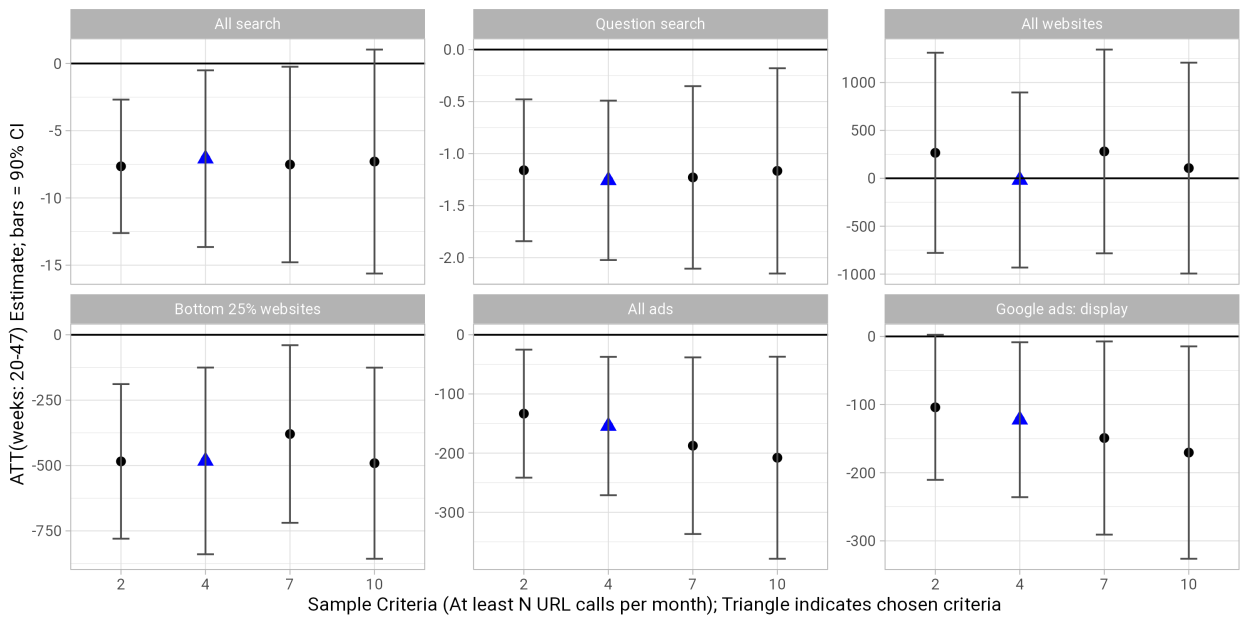

Figure A.1: Sensitivity to overall sample criteria.

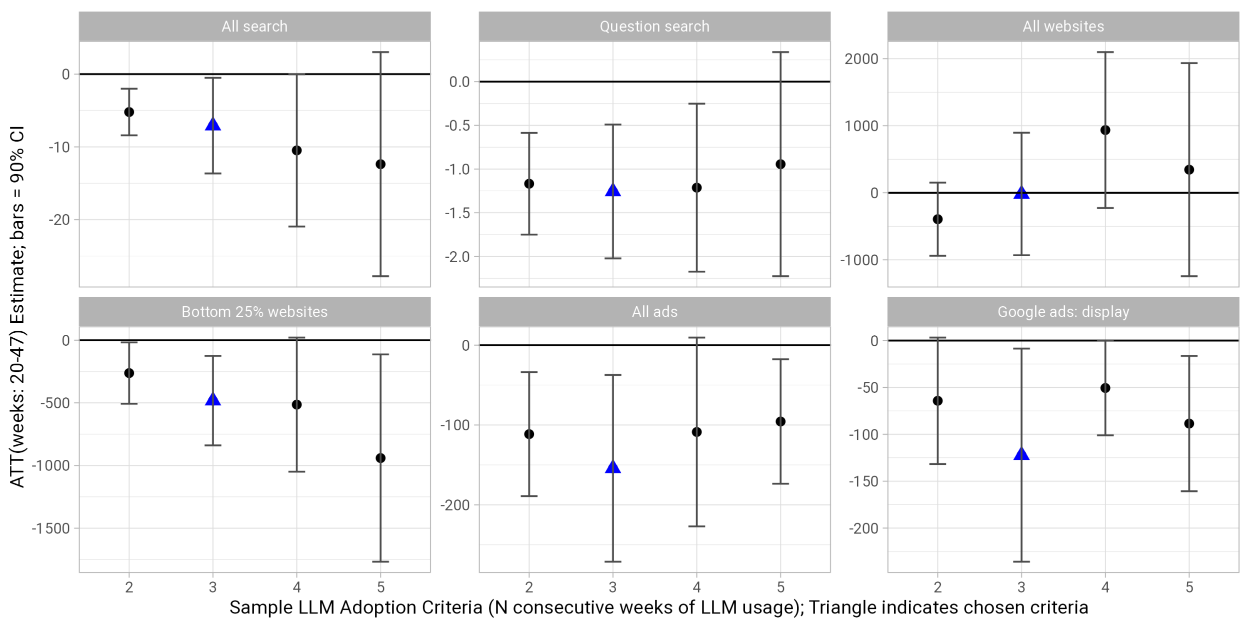

Figure A.2: Sensitivity to LLM adoption sample criteria.

B Additional Data Descriptives

B.1 Cohort heterogeneity in LLM activity

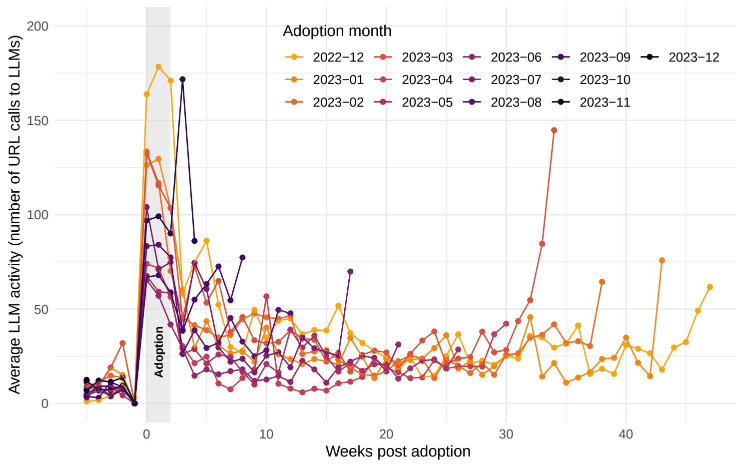

Figure B.3: Average number of URL calls to LLMs per post-adoption week by adoption month.

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.

D 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 95% CI of our main specification.

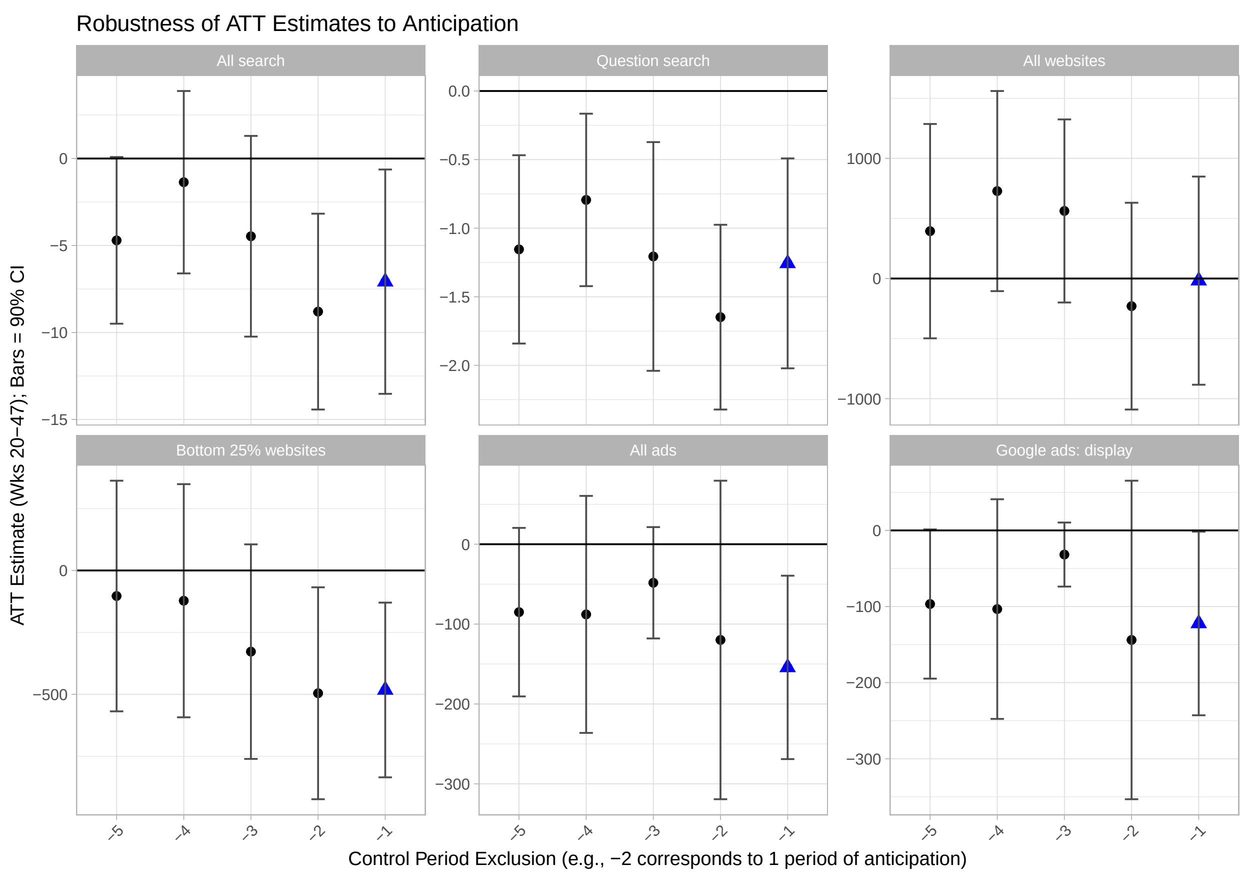

Figure D.4: Sensitivity to Control Period Exclusion

E 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 E.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 95% CI of our main specification.

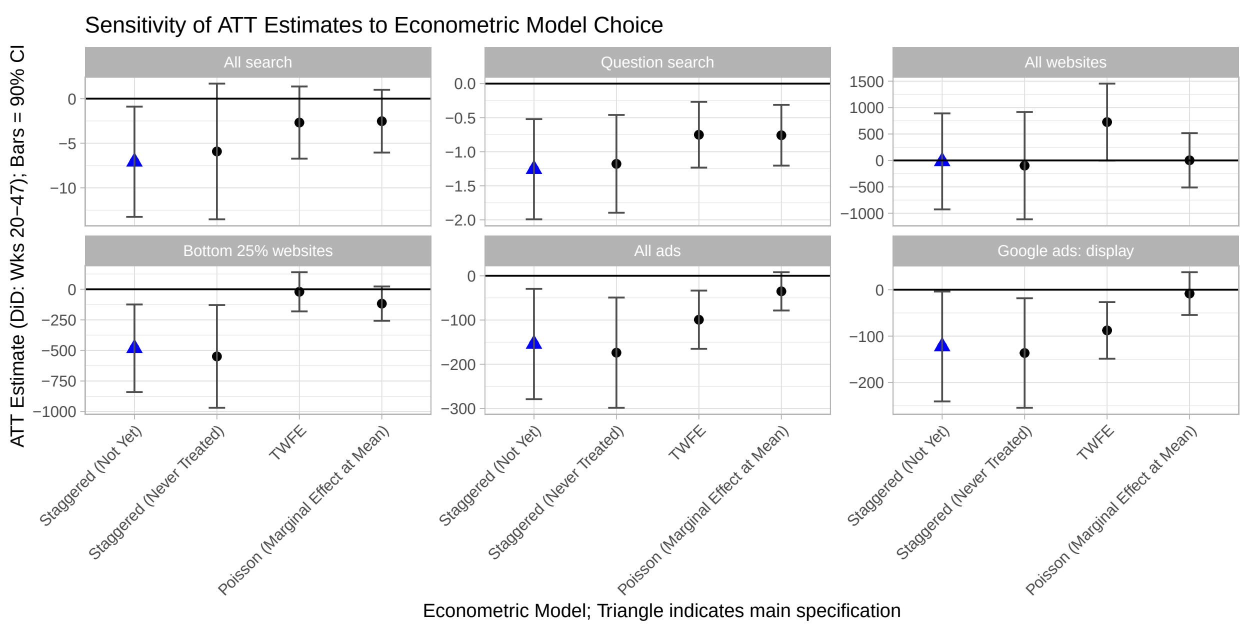

Figure E.5: Sensitivity to Model Specification

F Cohort-level ATTs

We analyze heterogeneity across adoption cohorts by focusing on four groups of cohorts: (1) weeks 4-15, (2) weeks 16-21, (3) weeks 22-28, and (4) weeks 29-48. The two groups of users adopt before the spike in adoption for Bing Chat. The third group captures those that adopted during the launch of Bing (Chat). The final group 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).

	Dependent variable:						
	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)
Cohort (weeks: 29-48)	1.770	0.419	−0.517	0.145	0.122	0.628	−0.762
	(5.033)	(4.032)	(0.763)	(0.240)	(0.308)	(0.826)	(1.288)
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 F.1: Cohort heterogeneous effects broken down by type of search

	Dependent variable:				
	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)
Cohort (weeks: 29-48)	5.136	6.817	−1.359	−0.311	−0.011
	(12.342)	(11.173)	(3.881)	(1.179)	(0.990)
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 F.2: Cohort heterogeneous effects broken down on referred traffic

	Dependent variable:				
	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)
Cohort (weeks: 29-48)	−18.435	−7.686	−2.302	−5.038	−3.410
	(34.451)	(29.204)	(3.077)	(8.042)	(24.169)
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 F.3: Cohort heterogeneous effects broken down on ads

	Dependent variable:			
	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)
Cohort (weeks: 29-48)	−91.794	−0.697**	−38.216	−33.077
	(91.110)	(0.284)	(62.419)	(44.839)
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 F.4: Cohort heterogeneous effects broken down on education websites

G Additional Results on Specific Domain Types

We show the figures displaying the average treatment effects over time since adoption for all specific domains discussed in the paper: education websites and user-generated content platforms.

G.1 Education websites

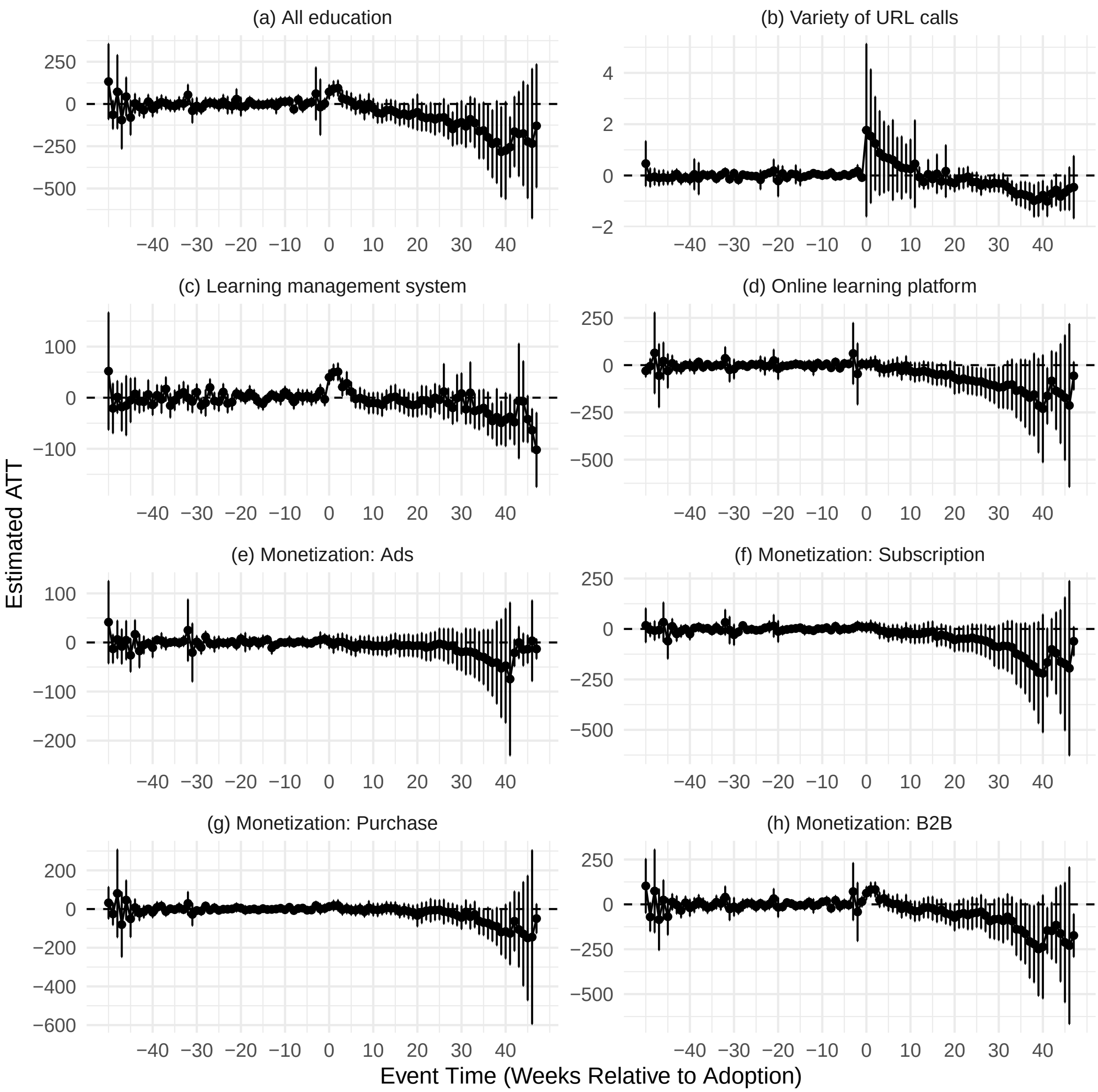

Figure G.6: Staggered DiD ATT on the weekly number of total traffic in education website, broken down by categories and monetization strategy.

G.2 User-Generated Content Platforms

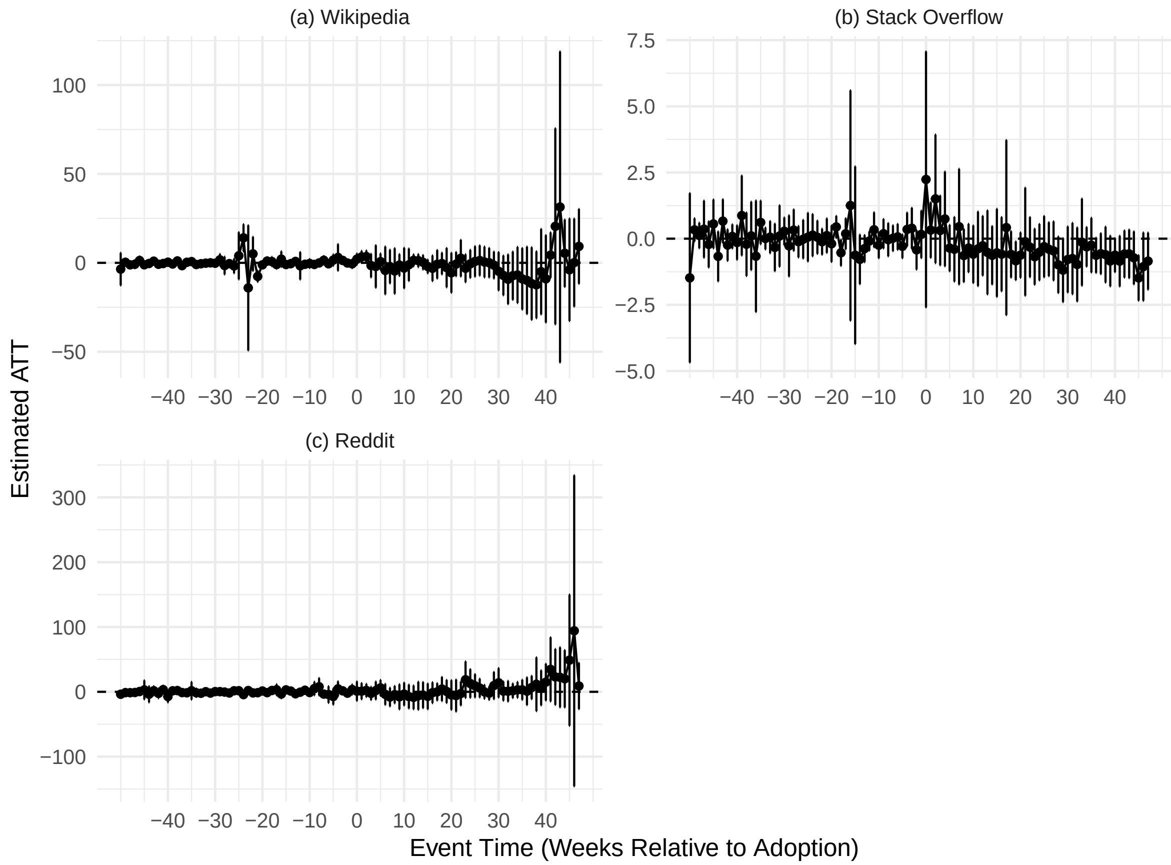

Figure G.7: Staggered DiD ATT on the weekly number of total traffic to knowledge-sharing platforms: Wikipedia, Stack Overflow and Reddit

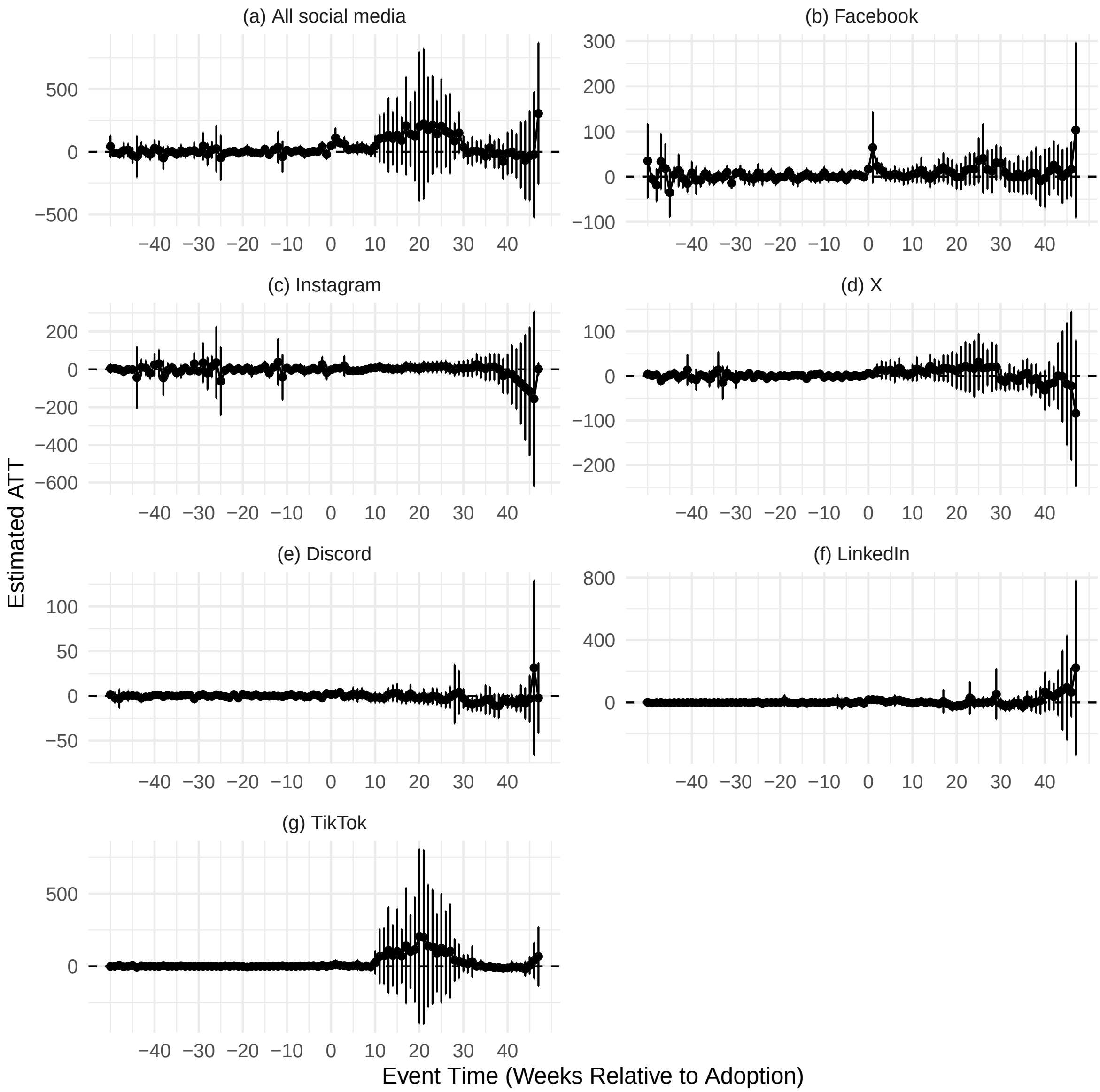

Figure G.8: Staggered DiD ATT on the weekly number of total traffic to social media platforms