# Beyond Search: LLM Adoption and Web Traffic Concentration[*]

Samira Gholami[†] Cristiana Firullo[‡] Cristobal Cheyre[§] Alessandro Acquisti[¶]

Preliminary draft: February 14, 2026.

**Abstract**

Search engines have long served as the primary gateway to the web, dispersing user attention across a wide range of external websites. The rapid adoption of large language models (LLMs), which provide answers directly within platform interfaces, raises concerns that online activity may become more centralized. We study whether LLM adoption displaces traditional web navigation or instead restructures how users discover and consume information. Using nationally representative Comscore panel data from 2019–2024 and session-level evidence from Internet Behavior Experiment, we examine how search activity, web exploration, and downstream navigation evolve with LLM use. Exploiting staggered adoption across users, we find that LLM adoption coincides with a sustained increase in the number of unique websites visited, alongside a short-lived rise in traditional search activity, suggesting that LLMs can complement rather than replace search. Session-level analyses show that most LLM use occurs alongside search, while LLM-only sessions remain rare. Moreover, when LLMs do route users to external websites, the resulting traffic is no more concentrated than traffic generated by traditional search.

[*]We thank Ashutosh Sahu and Yucheng Li for the excellent research assistance. This work was supported by the Alfred P. Sloan Foundation, Carnegie Mellon University Block Center for Technology and Society, Carnegie Mellon University CyLab Security and Privacy Institute, the Institute of Humane Studies and the Program on the Economics Privacy at George Mason University, the University of Pennsylvania Center for Technology, Innovation and Competition and the Warren Center for Network Data Sciences, the Pittsburgh Supercomputing Center, the NET Institute, the Cornell Center for Social Sciences, and the National Science Foundation (awards 2237327/2237328/2237329)

[†]Stanford University, Department of Economics.

[‡]Cornell University, Department of Information Science.

[§]Cornell University, Department of Information Science.

[¶]MIT Sloan School of Management.

# 1 Introduction

Information consumption is undergoing a substantial shift. For decades, search engines have served as the primary gateway to the web, directing users to a wide range of external websites. This intermediary role underpins the modern web's economic structure: search engines earn revenue by monetizing user traffic to third-party websites through online advertising (Decarolis and Rovigatti, 2021). Recent evidence suggests that this model is changing. A growing share of searches (approximately 60% for Google) now end without a click to an external website, coinciding with the rapid diffusion of generative artificial intelligence tools.[1] This rise in "zero-click" search coincides with the rapid diffusion of generative artificial intelligence tools. By generating answers and explanations directly within platform interfaces, large language models (LLMs) increasingly allow users to satisfy information needs without navigating across the web.

The scale of LLMs adoption is unprecedented. By the end of 2024, nearly 40% of U.S. adults had used generative AI tools (Bick et al., 2024), and by mid-2025 ChatGPT alone reached roughly 700 million weekly active users worldwide (Chatterji et al., 2025). In parallel, traditional search activity has shown early signs of decline, and incumbent platforms have begun integrating generative AI directly into the search experience, most notably through AI-generated summaries embedded alongside conventional results.

Despite these developments, systematic evidence on how LLMs reshape individual navigation remains limited. While the rise of "zero-click" search suggests a simple narrative of substitution and web contraction, the actual effects on how users navigate, discover, and allocate attention across the web remain an open empirical question. We study whether the adoption of large language models (LLMs) displaces traditional web navigation or instead restructures how users discover and consume information online. Using large-scale individual-level browsing data from Comscore, combined with session-level evidence from the Internet Behavior Experiment (IBE), we analyze how search activity, web exploration, and downstream navigation evolve as users adopt LLMs. Our empirical approach combines a staggered difference-in-differences design that exploits variation in adoption timing with detailed session-level analyses that characterize how LLMs are integrated into information-seeking

[1] See Tor Constantino, "The 60% Problem: How AI Search Is Draining Your Traffic," *Forbes*, April 14, 2025, https://www.forbes.com/sites/torconstantino/2025/04/14/the-60-problem---how-ai-search-is-draining-your-traffic/.

workflows.

Exploiting staggered adoption across users, we first examine the dynamic effects of LLM adoption on weekly browsing outcomes. Contrary to concerns that LLMs reduce engagement with the open web, adoption is associated with a sustained increase in the number of unique websites visited, alongside a short-lived rise in traditional search activity around the time of adoption. These patterns suggest that LLMs are not *yet* primarily substitutes for search, but could be adopted in the context of more complex information-seeking tasks that involve broader exploration, or enable new functionalities that expand users' digital footprint rather than consolidating it. Consistent with this interpretation, session-level evidence shows that workflows combining LLMs and traditional search generate substantially greater downstream exploration, while LLM-only sessions involve more limited navigation. Additionally, even in cases where LLMs may be substituting for search and resolving users' queries more rapidly, they may not lead to a reduction in browsing, as users instead use the time saved to browse across different domains.

We then characterize how LLMs differ from search engines in routing attention across the web. We find that both platforms frequently lead users to external websites, but traffic following LLM queries is systematically less concentrated across destinations. This is reflected in lower Herfindahl–Hirschman indices and higher entropy levels relative to traffic following search queries. While both platforms externalize users at comparable rates, they do so through distinct mechanisms: search concentrates traffic toward a narrow set of destinations, whereas LLMs disperse users across a wider range of websites. As LLM adoption grows, this diversification effect has the potential to reshape aggregate traffic patterns across the web.

Taken together, our results point to a nuanced reorganization of online navigation following LLM adoption. At the individual level, users become more active and wide-ranging explorers, visiting a broader set of distinct domains and increasingly combining LLMs with traditional search in multi-step workflows. At the aggregate level, LLM-mediated discovery redistributes attention away from a small set of dominant destinations and toward a more diffuse set of websites. These findings are consistent with LLMs being associated with greater individual exploration, alongside a shift in how attention is allocated across the open web.

Our results extend prior work by documenting how LLM use interacts with search and downstream web navigation in naturalistic settings. Prior research has examined large language models primarily through controlled experiments and platform-specific usage data. Experimental studies compare task performance with and without LLM assistance, documenting perceived improvements in convenience or quality alongside risks of over-reliance and shallow learning (Xu et al., 2023; Spatharioti et al., 2025; Melumad and Yun, 2025). Complementary work using conversation logs from platforms such as ChatGPT and Claude documents rapid adoption and widespread use, particularly in writing, programming, and decision-support tasks, with diffusion across a broad range of occupations (Zhang et al., 2023; Chatterji et al., 2025; Handa et al., 2025). While informative about what users do within LLM interfaces, these approaches do not observe how LLM use interacts with search engines or downstream web navigation.

A separate literature studies the consequences of LLM adoption for online platforms and content providers. Evidence shows declines in activity on platforms such as Stack Overflow following the release of ChatGPT (Vasconcelos et al., 2023; Burtch et al., 2024), and recent work using Comscore data documents reductions in traditional search activity and traffic to certain categories of websites after LLM adoption (Padilla et al., 2025). Related work argues that generative AI transforms search engines from traffic-routing intermediaries into providers of synthesized answers, with implications for content discovery and market power (Chen et al., 2025). Despite these advances, existing studies typically observe LLM interactions, search behavior, or platform outcomes in isolation. To our knowledge, no prior work jointly observes LLM use, traditional search, along with upstream and downstream browsing behaviors at the individual level. This paper contributes to filling this gap by tracing how users navigate across these tools and how attention is distributed across the web using real-world observational data.

The remainder of the paper proceeds as follows. Section 2 describes the Comscore and the Internet Behavior Experiment datasets and outlines the construction of our key measures. Section 3 presents descriptive empirical patterns on LLM adoption, browsing behavior, and session dynamics. Section 4 examines the causal effects of LLM adoption using a staggered difference-in-differences framework and analyzes how LLMs and search engines differ in routing attention across the web. Finally, Section 5 discusses the implications of our findings for online information markets and concludes.

## 2 Data

We use two complementary datasets. Our main dataset comes from the Internet Behavior Experiment (IBE), a longitudinal, IRB-approved field study of U.S.-based Chrome users. Launched in May 2025, the study recruits participants online, screens them for eligibility (age 18+, Windows OS, Chrome browser, recent online shopping activity, no ad-blocking tools), and invites them to install custom browser software.[2] Participants keep the software installed for three months and are compensated up to $245. The extension records page visits (URLs and timestamps), browser context (tab and window events), and full-page content, allowing us to reconstruct queries, results, and LLM interactions. This rich instrumentation makes it possible to directly compare downstream traffic following LLM prompts versus search queries. The current analysis draws on data from 308 participants. This dataset is part of a broader project that will reach 1200 participants in which participants are randomly assigned to advertising conditions (control, anti-tracking, ad-blocking).[3] For this paper, however, we treat IBE as an observational longitudinal panel and do not exploit this randomization.

Some limitations of the current IBE sample are important to note. First, because recruitment is ongoing, participants have contributed different amounts of data. To account for this, for some measures we compute participant-level averages and then average across participants. Second, because recruitment has primarily occurred via Facebook ads and is subject to our inclusion criteria, the IBE sample may not be fully representative of the broader online population. For this reason, we use a second data source to benchmark our IBE findings and assess whether the patterns we observe are consistent with broader population-level behavior.

Specifically, we rely on Comscore's U.S. panel (Version 5.5, curated by the Stanford GSB Library on Redivis), which provides individual-level browsing data from 2019 to 2024. The panel contains detailed information on URL visits, sessions, search activity, clicks, and user demographics for thousands of desktop and mobile users. A key advantage of this dataset is its scale and long time horizon, which allow us to benchmark aggregate browsing patterns—particularly around the release of ChatGPT in November 2022—and to evaluate whether the IBE-based results align with population-representative trends.

A comparison of demographic characteristics across Comscore, the full IBE sample, and IBE

[2]An example recruitment advertisement is provided in Appendix Figure 10.

[3]See Cheyre et al. (2024)

participants who use LLMs shows that the samples are broadly comparable, with only modest differences. Relative to Comscore, IBE participants are slightly older on average, include a somewhat higher share of women, and report marginally higher household incomes. Within the IBE sample, LLM users closely resemble the full participant pool, though they are on average somewhat younger and include a slightly higher share of students and unemployed individuals. Overall, these differences are small, suggesting that LLM adoption in IBE is not driven by sharp demographic selection. Taken together, the two datasets pair the fine-grained behavioral detail of IBE with the scale and representativeness of Comscore, providing complementary and largely consistent perspectives on LLM adoption and online browsing behavior.[4]

## 3 Descriptive Evidence

In this section, we present a set of empirical patterns that characterize how users adopt and interact with large language models (LLMs) and how this adoption relates to traditional search engines. The goal is to document descriptive patterns in usage, query intent, and downstream traffic that provide a foundation for assessing whether LLM adoption might determine changes in web traffic.

### 3.1 LLM adoption

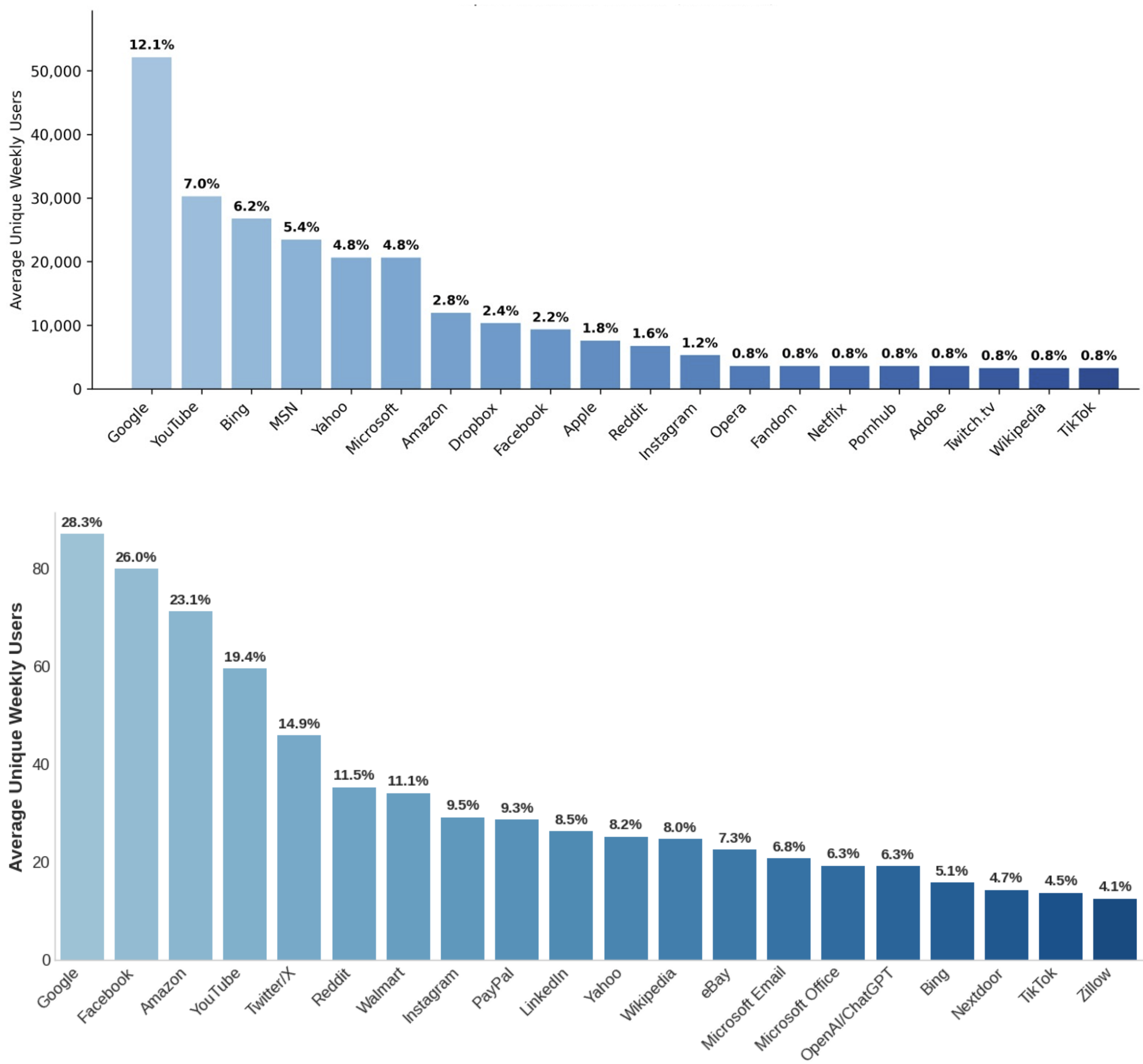
LLM adoption has been rapid. Recent estimates suggest that by mid-2025, ChatGPT alone reached roughly 700 million weekly active users worldwide (Chatterji et al., 2025). Consistent with this broader trend, adoption is clearly visible in both the Comscore and IBE data, though at different scales and levels of aggregation.

LLM adoption in the Comscore panel was staggered over 2023–2024, rather than concentrated around a single adoption event. As adoption expanded, LLMs became an increasingly salient component of users' overall browsing activity. Figure 1 compares the 20 most visited websites in the 2024 Comscore data and the 2025 IBE data. In 2024, the top-ranked sites are dominated by search engines, social media platforms, and large content hubs, with no LLM platform appearing among the most visited domains. By contrast, in the IBE sample, ChatGPT appears among the top 20 websites by average weekly active users, indicating that LLMs have become integrated into regular

[4] Refer to Appendix Table 5 for descriptive statistics of the two samples.

online browsing for a nontrivial share of users. While LLM use has become widespread, activity

Figure 1: Top 20 Websites in 2024 (Comscore) and 2025 (IBE)




remains concentrated across providers. Table 1 reports the share of unique LLM users visiting each platform in the Comscore and IBE samples. ChatGPT and Gemini account for the majority of observed usage, while other providers - including Claude, Copilot, Perplexity, and Grok - appear with smaller but nonzero shares. This pattern indicates that a few platforms capture most LLM usage - likely reflecting early entry and integration into existing products - while multiple alternative providers are also used by consumers.

Table 1: Share of unique users by LLM platform

| **LLM Platform** | **Comscore 2024(%)** | **IBE 2025 (%)** |
|---|---|---|
| ChatGPT | 78.43 | 60.0 |
| Gemini | 11.1 | 20.3 |
| Perplexity | 1.11 | 4.7 |
| Claude | 1.3 | 2.8 |
| Microsoft Copilot | 3.3 | 1.5 |
| Grok | 11.3 | 10.2 |
| Poe | 1.7 | 0.3 |
| DeepSeek | – | 0.1 |
| Character.AI | 15.1 | $<0.1$ |

*Notes:* Entries report the share of LLM activity registered in the two dataset. The Comscore sample includes a smaller set of major platforms, while the IBE data capture a broader set of LLM services.

### 3.2 LLMs and browsing sessions

To understand how LLMs fit into users' online activity, we analyze the bowsing behavior at the session level. *Browsing sessions* are defined as continuous sequences of browsing activity with no gap longer than 30 minutes of inactivity.[5] A browsing session is a natural boundary for a single episode of attention allocation and information seeking: within a session, users typically pursue a coherent goal (or a small set of related goals) and make repeated navigation decisions among substitutable sources (e.g., search engines, social platforms, publishers, and LLM interfaces). In this sense, a session approximates a short-run "market" for attention and referral traffic, in which websites (and intermediaries such as LLMs) compete to receive a user's next click and, ultimately, the session's traffic share. Defining the market at the session level is also empirically useful: it reduces mechanical variation driven by time-of-day and cross-day routines, allows us to compare like-with-like across users, and aligns measurement with the behavioral margin we study—how the introduction or use of an LLM within an episode reshapes the allocation of traffic across domains and the concentration of visited sites.

Within each session, we classify activity based on whether the user interacted with a search engine, an LLM, or neither. In particular, a session is: *search only* if the user visited (among other websites) search engines but not any LLM; *LLM only* if the user visited (among other websites)

[5]This definition comes from the industry practice, and it is known as the 'Google Analytics' definition. See Google, *About Analytics Sessions*, https://support.google.com/analytics/answer/9191807?hl=en

LLM platforms, but not search engines; *mixed* if the browsing session included both search engines and LLMs; *no LLMs nor search* alternatively. Figure 2 shows the breakdown of session types among LLM and non-LLM users in IBE data. For both user groups, the majority of browsing sessions involve search engines. For LLM users, sessions that include LLM visits account for only about 8% of browsing activity.

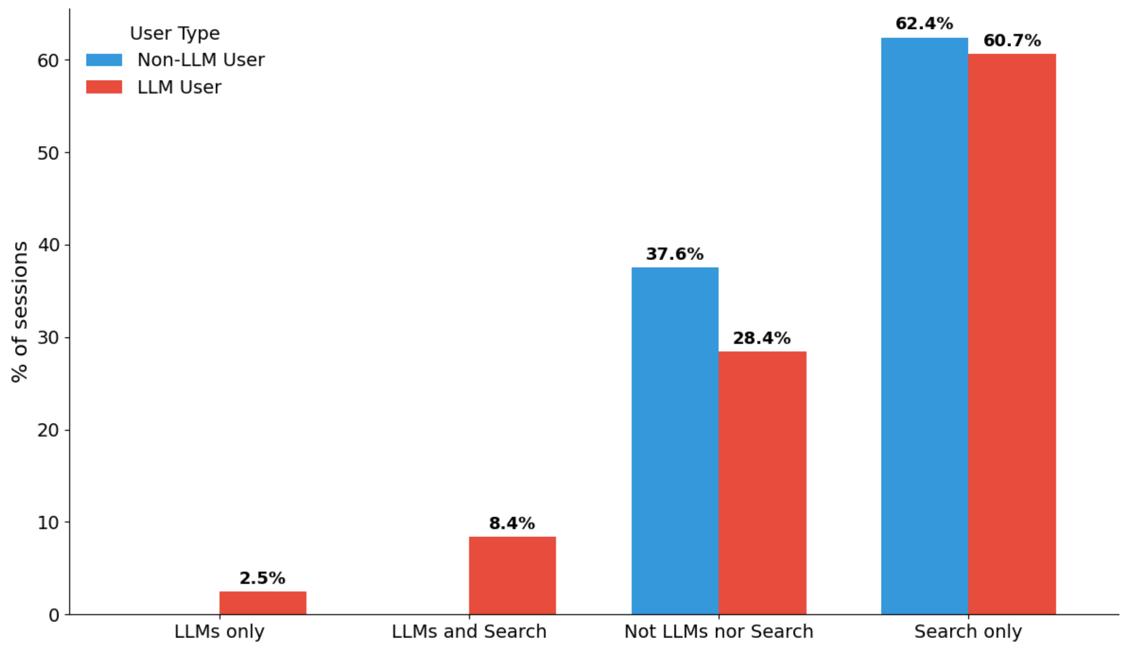



Figure 2: Distribution of search types among LLM and non-LLM users (IBE data)

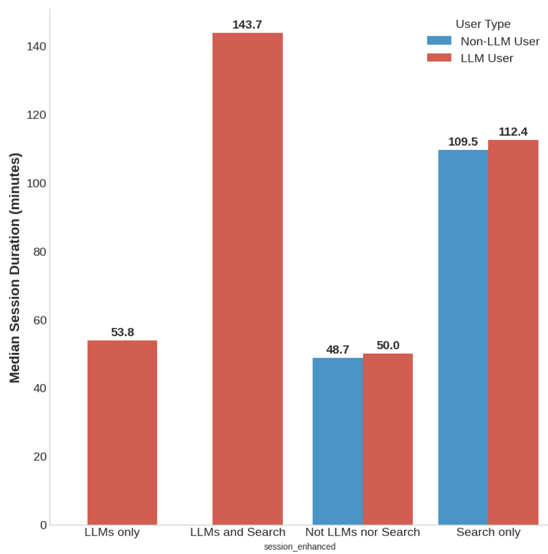
This distribution highlights two points. First, while LLM adoption is widespread at the user level, it remains embedded within a search-dominated ecosystem: most online activity continues to pass through search engines. Second, the small share of LLM-only sessions suggests that, for now, LLMs are rarely the sole entry point for browsing. Instead, they are typically used alongside search, consistent with the idea that LLMs are currently extending rather than replacing established patterns of online navigation.

### 3.2.1 Session characteristics

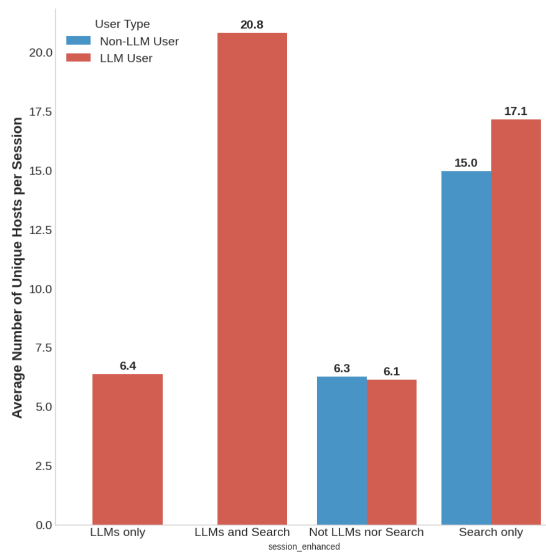
Browsing sessions capture how users navigate the web over short periods of activity. Sessions vary in their length, the number of websites visited, and the tools used during the session. Clear differences emerge across session types: sessions that combine LLMs and search engines are both longer and involve visits to more websites than sessions relying on a single tool. This pattern is consistent with users turning to multiple tools when addressing more complex or multi-step information needs.

Figure 3 compares the median session duration and the average number of unique hosts visited across session types.[6]

Figure 3: Comparison of session length and breadth across session types (IBE data).




(a) Median session duration by type (minutes).




(b) Average number of unique hosts per session by type.

The results show clear differences. Firstly, for both types of users, search-only sessions are long, with a median duration 109.5 minutes and 112.4 minutes for non-LLM and LLM users respectively. For LLM users, mixed sessions (LLMs and Search) are even longer, suggesting that the user might benefit from both tools for certain types of tasks. To capture the breadth of activity within a session, we count the number of distinct URL hosts visited (e.g., `nytimes.com`, `wikipedia.org`, `amazon.com`). A higher number indicates more exploratory browsing, while a lower number reflects narrower attention. LLM users visited an average of 20.8 hosts per session in mixed sessions, compared with 17.1 for search-only and around 6.1-6.4 for other sessions and LLM-only sessions. On the other hand, non-LLM users visit an average of 15 unique hosts in search-only sessions (2 less than LLM users) and about 6.3 in other types of sessions. Together, these results suggest that mixed sessions not only last longer but also span a wider range of destinations, consistent with

[6]To account for heterogeneity across participants, we first compute averages and median for each user and then average across users. This ensures that results reflect typical behavior rather than being dominated by a handful of heavy participants.

users combining search and LLMs to complete complex or multi-step tasks. By contrast, LLM-only sessions are both shorter and narrower, suggesting that users often remain within the model interface or circle between a small set of sites. Search-only sessions fall in between, reflecting their traditional role as outward gateways but without the sustained engagement seen when search and LLMs are used together.

Examining browsing activity on a weekly basis also suggests there are no substantial underlying differences between LLM and non-LLM users. On average, non-LLM users engage in about 18.4 weekly browsing sessions, compared to roughly 20 sessions for LLM users, a difference that is not statistically meaningful. Thus, although LLM users tend to spend more time per session and visit a larger number of websites, the overall frequency of browsing sessions does not differ significantly by LLM adoption status.

#### 3.2.2 Switching behavior across platforms

In what follows, we look at session trajectories to characterize how users navigate between LLMs, search engines, and the broader web. Instead of focusing only on duration or the number of sites visited, we classify sessions by their underlying structure and transitions. In particular, we look at:

- transitions to/from search engines and LLMs during the same browsing session
- *looping* behaviors, namely intentional re-visits of the same website within a given sessions

To account for the first metrics, we focus on *mixed* sessions (those containing both search and LLM activity), and we classify the first move from one platform type to the other that is not followed by a return within the remainder of the session as a *no-return transition*. For example, if a user begins on Google, switches to ChatGPT, and then spends the rest of the session within the LLM without going back to search, this counts as a no-return transition from *Search* → *LLM*. By contrast, if a user starts in ChatGPT, jumps to Bing, and never returns to the LLM, that counts as *LLM* → *Search*. Figure 4 shows that 60% of such transitions go from *Search* → *LLM*, while only 40% go from *LLM* → *Search*. In other words, once a session shifts from search to an LLM, it is more likely to remain on the LLM for the rest of the session than to move back. This asymmetry is descriptive rather than causal, but it is consistent with search acting as the gateway and LLMs serving more often as the terminal destination where tasks are completed. Moreover, it might suggest that we are experiencing a phase in which users are experimenting the usage of LLMs for queries and tasks that have been traditionally leveraged search engines.

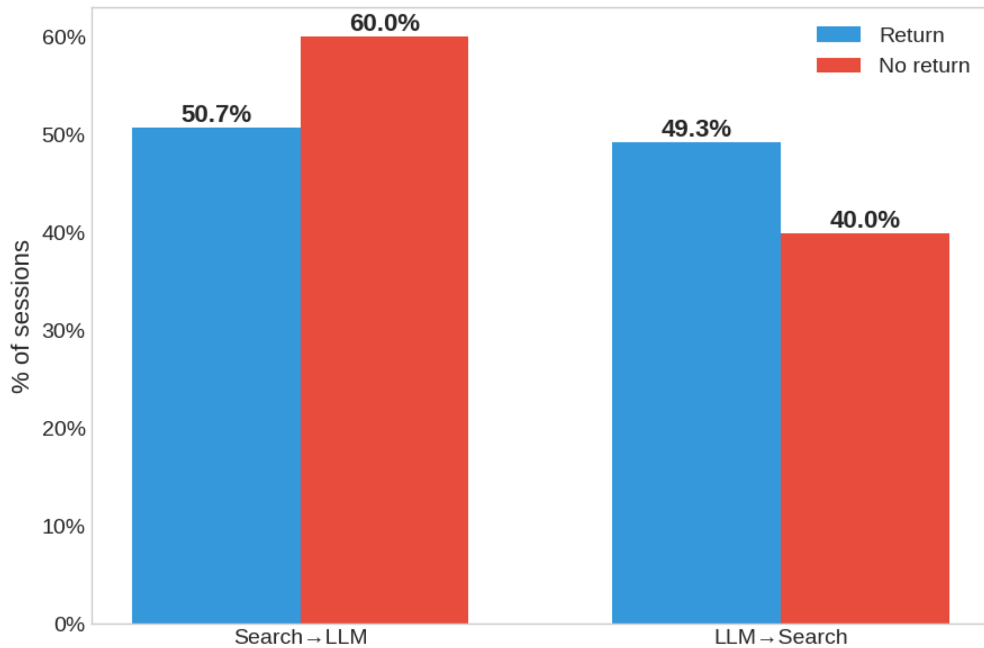
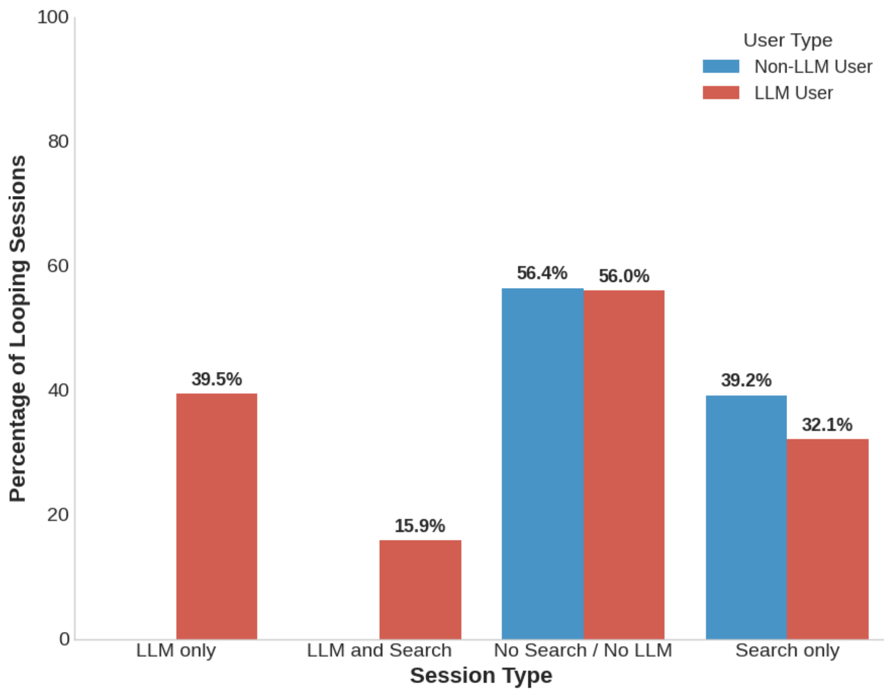



Figure 4: Comparison of transitions and no-return transitions in mixed sessions

To capture the structure of browsing within a session, we distinguish between two patterns. A session is classified as *looping* if the top domain accounts for more than half of all pageviews, meaning the user repeatedly returns to the same site or small cluster of sites. By contrast, a session is *exploratory* if activity is spread more evenly across multiple domains, suggesting the user is branching outward into new destinations. Figure 5 shows the proportion of looping browsing sessions by session type and user type. Firstly, we notice that LLM-only sessions are overwhelmingly looping (about 60%), more than search-only sessions for both type of users. These often look like short cycles anchored on the LLM itself—for example, a user asks ChatGPT a question, opens Amazon to check a product, and then returns to ChatGPT, repeating this pattern without venturing into new sites. Such behavior is consistent with treating the LLM as a hub. Search-only sessions are more balanced, with 39.2% and 32.1% looping for respectively non-LLM and LLM users. A looping search session might end up with the user landing to the final destination of their browsing session soon after having searched on a search engine. In a sense, a looping search session might reflect the quality of a search engine in minimizing the time needed for the user to achieve their browsing goals. On the other end, an exploratory search session is characterized by the visit to a diverse set of sources—Wikipedia, Reddit, YouTube—before moving on to unrelated sites. Mixed sessions are the most outward-facing: about 56% are exploratory. A typical mixed session might start with Google, move into ChatGPT to refine a question or draft text, and then branch out to several new websites that were not initially in mind—producing a much broader footprint. These sessions illustrate how combining search and LLMs can sustain more complex or multi-step tasks. This matches what we saw with unique hosts. Using both search and LLMs leads people to visit more sites, while LLM-only sessions are shorter and more repetitive.

Figure 5: Percentage of Looping browsing sessions by session type and user type




### 3.3 A taxonomy for user intent on LLM conversations

A key advantage of the IBE dataset is that the HTML data it collects allows us to recover both search queries and full LLM conversations. This allows us to study both similarities and differences in queries between search engines and chat bots. To compare how people use LLMs and search engines, we classify queries by the type of task they reflect. A standard framework in the search literature is the taxonomy of Broder (2002), which distinguishes between *informational* queries (seeking facts or explanations), *navigational* queries (aimed at reaching a specific site, document, or tool), and *transactional* queries (intended to complete an action). This taxonomy has become a benchmark because it provides a parsimonious representation of user intent. While Broder's taxonomy captures broad similarities across platforms, it does not fully reflect the expanded functionality of LLMs. Unlike search engines, LLMs integrate reasoning, planning, content generation, and conversational interaction within a single interface, and they introduce new forms of navigation such as searching within documents or moving across tools. To capture these differences, we extend Broder's framework by adding two macro-categories—*Creative / Generative* and *Meta / Social / Chit-chat*—and by refining the original classes into ten subcategories. We implement the extended taxonomy using supervised learning. A subset of queries is hand-labeled and used to train a classifier, following a two-step procedure: (i) prompting LLaMA 3.2 to propose candidate categories based on query summaries and keywords, and (ii) refining and collapsing overlapping categories.[7]

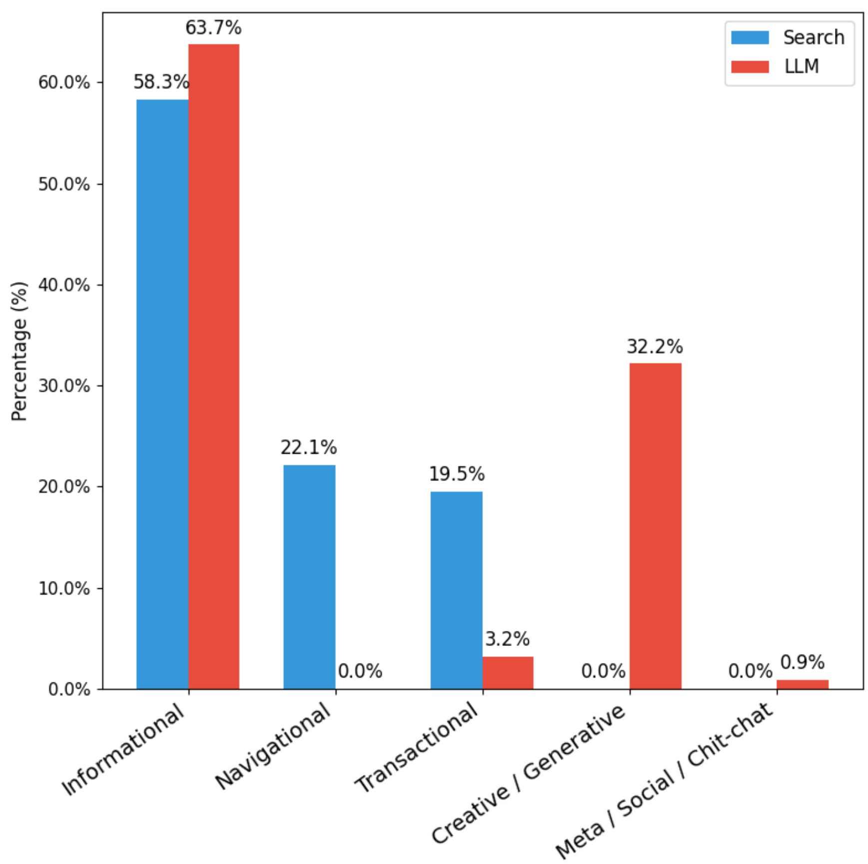



Figure 6: Distribution of macro query categories (LLM vs. Search). Informational, navigational, and transactional shares follow the framework of Broder (2002), which is extended to capture

[7]Following practices in recent applied work, see Handa et al. (2025).

Figure 6 shows the distribution of LLM conversations across the extended macro-categories. Informational queries remain the majority (66.7%), but a substantial share of activity falls into other categories, most notably generative (32.2%), with smaller fractions devoted to transactional (3.2%), and social interactions ($< 1\%$). The presence of these new categories – in particular, the *Creative / Generative* one – emphasizes once again how LLM usage extends well beyond information retrieval, motivating a finer classification at the micro-category level.

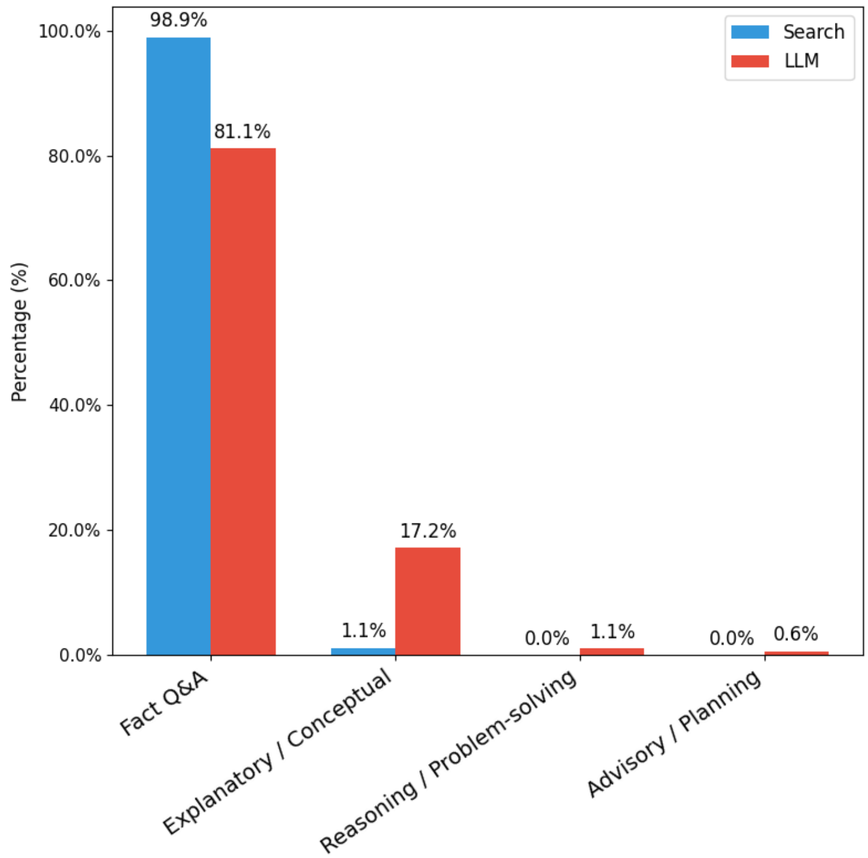



Figure 7: Distribution of subcategories under the *Informational* macrocategory, using

Looking within the informational category, activity is heavily skewed toward explanatory and conceptual requests, which account for roughly 60% of informational queries. Summarization and condensation tasks make up about 24%, reflecting users' reliance on LLMs to process and synthesize existing material. Factual question-and-answer queries represent a much smaller share (around 4.3%), while advisory queries account for approximately 1% of informational interactions. This breakdown indicates that even within informational use, LLMs are employed also for higher-level explanation and synthesis in addition to simple fact lookup. This taxonomy highlights that LLMs are not simply another channel for information retrieval, but platforms that integrate reasoning, generation, navigation, and action within a single environment.

# 4 The impact of AI adoption on web browsing patterns

In this section, we analyze the two complementary datasets, Comscore and IBE, with the goal of understanding whether the usage of AI tools might have an impact in the breadth of users' browsing activity. Firstly, we study the adoption of LLMs by Comscore panelists, analyzing how their browsing behavior changes in terms of the number of unique domains visited on a defined window of time. Secondly, we study within-participant variation for LLM adopters of the IBE sample, based on the type of browsing session and the inferred user intent – based on the classification of search engine queries and LLM conversations.

## 4.1 LLM adoption in the Comscore data

Using individual-level browsing data from Comscore, we study how users' online activity evolves following the adoption of large language models (LLMs). We focus on two outcomes measured at the user–week level: (i) the number of unique websites visited, and (ii) the number of traditional search visits. Following Padilla et al. (2025), we define LLM adoption as the first week in which a user engages with an LLM for three consecutive weeks. This definition allows us to distinguish sustained adoption from sporadic or experimental use.

Our analysis exploits staggered adoption across users between August 2023 and January 2024. We focus on this adoption window because uptake is gradual and smooth, without sharp spikes driven by major model releases or platform redesigns, allowing us to study behavioral changes associated with adoption rather than discrete technological shocks. The final sample consists of 1,770,270 desktop user–week observations, of which 4,394 users are classified as adopters during the adoption window.

**Empirical specification.** To characterize the dynamic effects of LLM adoption, we estimate an event-study specification in a staggered difference-in-differences framework. Let $i$ index users and $t$

index calendar weeks. Define event time as

$$\tau_{it} = t - T_i,$$

where $T_i$ denotes the adoption week for user $i$. The outcome variable $y_{it}$ is either (i) the number of traditional search visits in week $t$ or (ii) the number of distinct URL hosts visited in week $t$ excluding search engines and LLM platforms. We estimate the following two-way fixed effects regression:

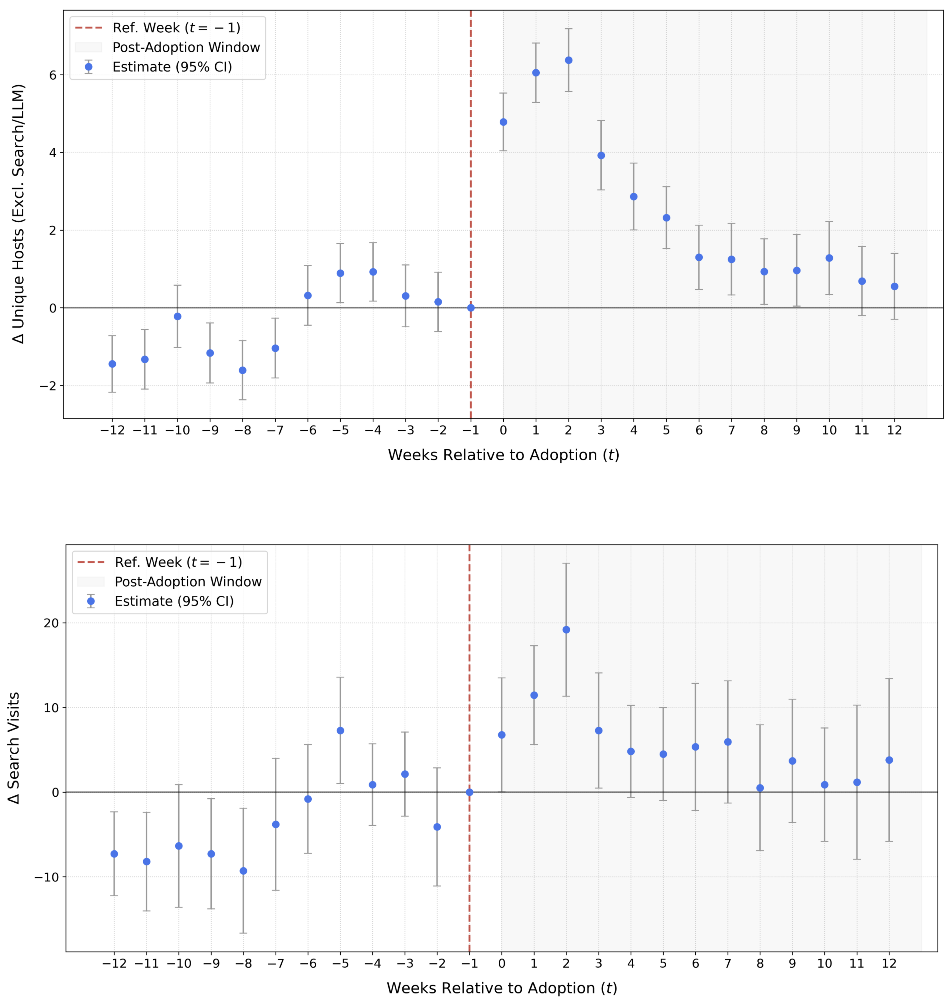
$$y_{it} = \sum_{\substack{k=-12 \\ k \neq -1}}^{12} \beta_k \cdot \mathbb{1}\{\tau_{it} = k\} \; + \; \gamma \, m_{it} \; + \; \alpha_i \; + \; \lambda_t \; + \; \varepsilon_{it},$$

where $\mathbb{1}\{\tau_{it} = k\}$ are event-time indicators relative to adoption, with week $-1$ omitted as the reference period. The variable $m_{it}$ measures the number of visits to email providers' websites (e.g. mail.gmail.com) by a user in a given week and proxies for time-varying individual online activity. We include this control for two reasons. First, it captures short-run changes in a user's availability (such as vacations or fluctuations in workload) that are not absorbed by user or week fixed effects. Second, email use reflects a routine, habitual activity that is distinct from information-seeking behavior. This makes it a useful control for overall online activity while avoiding the mechanical endogeneity. The specification includes user fixed effects $\alpha_i$, which absorb time-invariant heterogeneity in browsing behavior, and calendar-week fixed effects $\lambda_t$, which account for aggregate shocks common to all users. Standard errors are clustered at the user level. This specification compares treated users to not-yet-treated users within the same calendar week, leveraging the staggered timing of adoption. Identification relies on the assumptions of conditional parallel trends across adoption cohorts and no anticipatory behavioral responses prior to adoption.

**Effects on website breadth.** The top panel of Figure 8 shows the dynamic effects of LLM adoption on the number of unique websites visited per week. In the pre-adoption period, coefficients exhibit some movement but remain small and largely centered around zero.

Following adoption, users visit a broader set of websites relative to the pre-adoption baseline. The increase appears immediately after adoption and remains positive throughout the post-adoption window. Although the magnitude of the effect declines over time, estimates remain elevated for

Figure 8: Effect of LLM Adoption on Traditional Search Activity and Web Diversity




*Notes:* The top panel reports corresponding estimates for the number of unique websites visited. The bottom panel reports event-study estimates of the effect of LLM adoption on the number of traditional search visits per week. Coefficients are plotted relative to the week immediately prior to adoption (week $-1$), which is omitted as the reference category. Vertical dashed lines indicate the timing of adoption. Error bars represent 95% confidence intervals based on standard errors clustered at the user level.

several weeks, indicating that the change in browsing breadth is not purely transitory.

At the same time, these patterns admit multiple interpretations. One possibility is that LLM adoption directly lowers the cost of discovering relevant sources (by summarizing information, recommending links, or directing users toward specific domains) thereby encouraging broader exploration. A complementary interpretation is that users adopt LLMs precisely when they face more complex informational tasks, such as research, comparison, planning, or troubleshooting, which naturally require synthesizing information from multiple sources. Even conditional on overall online activity, these tasks may lead users to reallocate their browsing toward a larger set of distinct websites, reflecting a shift in how information is gathered rather than a change in the volume of activity.

Under either interpretation, the observed post-adoption increase in the number of unique websites visited does not necessarily reflect a mechanical effect of LLMs on navigation alone. Instead, it may capture changes in the types of tasks users undertake once LLMs become part of their workflow. In this sense, the jump in browsing breadth could arise from a compositional shift (fewer repeated visits to the same sites and more one off visits across a broader range of domains) rather than from heightened browsing intensity per se.

For this reason, the event-study estimates are best interpreted as documenting changes in browsing behavior around the timing of adoption, rather than isolating a single causal mechanism. In particular, as discussed in Section 3.2.1, when LLM usage is combined with traditional search, sessions become longer and span a larger number of distinct websites, indicating that increases in browsing breadth are closely tied to more intensive and exploratory information-gathering activity rather than to a mechanical effect of LLMs on navigation alone.

**Effects on traditional search activity.** The bottom panel of Figure 8 plots event-study estimates of the effect of LLM adoption on the number of traditional search visits per week. In the pre-adoption period, coefficients fluctuate around zero and are generally small, though somewhat noisy, indicating no clear evidence of strong differential pre-trends across adoption cohorts.

Following adoption, the estimated coefficients remain noisy, particularly in weeks 8–12. Search activity increases around the time of adoption and then gradually declines in the weeks that follow. However, estimates become increasingly imprecise toward the end of the post-adoption window, with wide confidence intervals that limit strong inference. As a result, it is difficult to make a sharp causal claim about the effect of LLM adoption on search activity based on these estimates alone.

Overall, the event-study evidence indicates that search behavior around adoption is volatile rather than exhibiting a clear monotonic response. The combination of a short-run jump and a subsequent decline is consistent with multiple interpretations, including short-lived bursts of search during adoption or learning phases, followed by reduced reliance on traditional search. Given the noise in the estimates, we interpret these results as documenting changes in search activity around the timing of adoption, rather than as definitive evidence of a persistent causal reduction in search usage.

Overall, the event-study evidence indicates that LLM adoption coincides with a reorganization of users' browsing behavior around the time of adoption. Both traditional search activity and the number of unique websites visited increase sharply at adoption, suggesting a period of intensified and broadened information seeking. In the weeks that follow, search activity gradually declines, while browsing breadth remains elevated relative to the pre-adoption period.

These patterns are consistent with users adopting LLMs in the context of complex informational tasks that involve active exploration and engagement with multiple sources, followed by a partial reallocation away from repeated search queries as LLMs become integrated into users' workflows. Taken together, the results suggest that LLMs do not simply replace existing online activity, but reshape how users access and navigate information on the web over time.

**Robustness: Continuous LLM Usage and Weekly Browsing Outcomes** As a complementary robustness exercise, we examine how weekly search activity and web exploration vary with the *intensity* of LLM usage, rather than with discrete adoption timing. Specifically, we estimate the following two-way fixed effects specification at the user–week level:

$$Y_{it} = \beta \log(1 + \text{LLMVisits}_{it}) \; + \; \gamma \, \text{Email Visits}_{it} \; + \; \alpha_i \; + \; \lambda_t \; + \; \varepsilon_{it},$$

where the outcome variable $Y_{it}$ is alternatively (i) the number of traditional search visits or (ii) the number of unique websites visited in week $t$. The variable Email $\text{Visits}_{it}$ proxy for users' overall online activity and time spent online, allowing us to net out changes in baseline engagement that are unrelated to information acquisition. All specifications include user fixed effects $\alpha_i$ and calendar-week fixed effects $\lambda_t$, with standard errors clustered at the user level.

These regressions serve two purposes. First, they provide a transparent check on whether the patterns documented in the event-study analysis also appear when LLM usage is treated as a continuous measure rather than a discrete adoption event. Second, by conditioning on overall online activity, they help assess whether the observed relationships between LLM usage and browsing outcomes merely reflect periods of unusually high online engagement.

Table 2 shows that weeks with higher LLM usage are associated with more traditional search and broader web exploration. A one-unit increase in $\log(1 + \text{LLM Visits})$ is associated with 9.57 additional search visits (SE = 0.78, p = 0.000) and 3.47 additional unique non-search/non-LLM hosts (SE = 0.11, p = 0.000). Interpreting magnitudes, moving from 0 to 1 LLM visit in a week corresponds to roughly +6.6 search visits and +2.4 unique hosts, though these estimates are descriptive rather than causal because usage intensity is endogenous to information needs. Overall, this robustness exercise reinforces the view that LLM usage is associated with richer patterns of information seeking rather than a simple displacement of traditional web navigation.

Table 2: LLM Usage, Search Activity, and Web Diversity

| | (1) Search Visits | (2) Unique Hosts (Excl. Search & LLM) |
|---|---|---|
| $\log(1 + \text{LLM Visits})$ | 9.57 | 3.47 |
| | (0.78) | (0.11) |
| Email Visits | 0.03 | 0.01 |
| | (0.008) | (0.004) |
| User fixed effects | Yes | Yes |
| Calendar-week fixed effects | Yes | Yes |
| SE clustered at user level | Yes | Yes |
| Observations | 158,216 | 158,216 |
| Users | 4,394 | 4,394 |
| $R^2$ | 0.0046 | 0.03 |

*Notes:* The table reports two-way fixed effects regressions at the user–week level. Column (1) uses the number of traditional search visits as the outcome. Column (2) uses the number of unique websites visited excluding search engines and LLM platforms. The key regressor is log LLM usage in a given user–week. All specifications include user and calendar-week fixed effects and control for email visits. Standard errors are clustered at the user level.

### 4.2 AI-generated answers and browsing activity

In this section, we will analyze how AI-generated answers (either LLM answers and AI Overviews) influence the extent and the depth of how people browse the internet.[8] The rationale for this is hat AI-generated answers may reduce users' need to visit external websites by providing direct responses to their queries. If LLMs and AI Overviews successfully answer users' questions, we would expect to observe fewer click-throughs to traditional web content, potentially reshaping traffic patterns across the web ecosystem. We have already presented some descriptive evidence that browsing session where participants use LLM chat bots (i.e., 'llm only' and 'mixed' sessions) are characterized, respectively, by a lower and a higher number of unique hosts visited compared to 'search only' sessions (see Table 3b). We infer whether search results pages contain *AI Overviews* by inspecting the presence of dedicated selectors.[9] Figure 9 shows the average number of unique hosts – excluding search engines and LLM platforms – that participants visited, based on whether search engines provided or not AI Overviews.

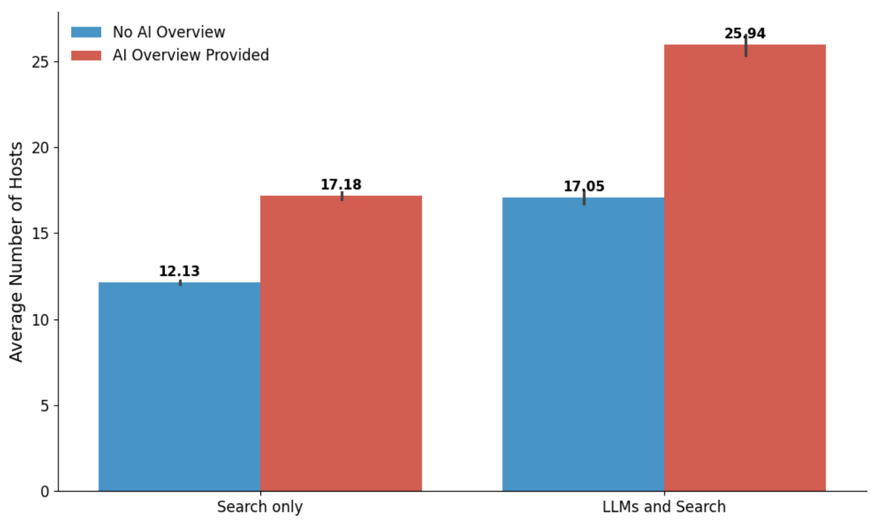



Figure 9: Average number of unique hosts (excluding search and LLM hosts) visited, based on whether AI Overviews are provided by search engines.

Sessions where AI Overviews were provided are associated with a higher average number of unique hosts visited. In search-only sessions, participants visited 12.1 hosts on average when no

[8]AI Overviews are defined as AI-generated summaries that are provided to users of search engines in response to particular search queries. Examples are the *Google AI Overview* for Google and *Copilot Search* for Bing.

[9]For instance, we detect the presence of AI Overviews in Google search results pages by looking for particular selectors which also contain the 'Google AI Overview' text.

AI Overview was shown, compared to 17.2 hosts when an AI Overview was present. A similar pattern emerges in mixed sessions (LLMs and Search), where the presence of AI Overviews is associated with approximately 9 additional hosts visited (25.9 vs 17.1). While these results may seem counterintuitive, it is important to note that our unit of analysis is the entire browsing session rather than solely the user's actions on the search results page. This design allows us to capture the broader content consumption effects of integrating AI tools. The figures suggest that while AI summaries may reduce the likelihood of clicking a specific search result, they do not curb overall consumption. Instead, users likely leverage the efficiency of the summary to save time on their initial query, subsequently investing that time in visiting additional, distinct domains.

Additionally, these results may reflect selection effects. AI Overviews are not randomly assigned: Google's algorithm tends to generate them for queries that are more informational, complex, or exploratory in nature. These same query characteristics are independently associated with more extensive browsing behavior. In other words, the type of query that triggers an AI Overview is also the type of query that leads users to visit multiple websites, regardless of whether an AI Overview is displayed. To address this concern, we estimate regression models that control for session duration and user fixed effects. In particular, we focus on sessions where participants either use search engines or search engines and LLM chat bots, thus excluding those sessions where users We estimate the following regression

$$\text{Unique hosts}_{i,s} = \alpha + \boldsymbol{\beta} SessionType_{i,s} + \gamma \cdot log(\text{duration})_{i,s} + \text{AI Overviews}_{i,s} + \delta_i + \epsilon_{i,s} \quad (1)$$

where: 'Unique hosts' is the number of unique domains visited by participant $i$ during session $s$, excluding search engines and LLM platforms; $SessionType$ can be 'search only', 'llm only', 'mixed' or 'none', depending on whether participant $i$ in session $s$ visited only search engines, only LLM platforms, both search engines and LLM platforms or none of them respectively; 'AI Overviews' is a dummy variable for whether AI Overviews have been displayed at least once during session $s$ to participant $i$.

Table 3 shows the results of the estimation of (1), over the subsample of LLM adopters. We focus on this subsample as we recognize that comparing LLM users to non-LLM users would conflate the effect of LLM usage with unobserved heterogeneity in user characteristics—LLM adopters may differ systematically in their browsing habits, information needs, and technological sophistication. By restricting the analysis to LLM users, we can exploit within-user variation in session types, comparing how the same users behave when they use search only versus when they combine search with LLM assistance. The baseline category is *Search only* sessions. Column (I) presents the base specification, while Column (II) adds controls for search query categories. Sessions where users combine both search and LLM (*Mixed sessions*) are associated with approximately 2.1–2.3 more unique hosts visited compared to search-only sessions ($p < 0.01$). In contrast, *LLM only* sessions show no significant difference from search-only sessions. The presence of AI Overview is positively associated with unique hosts visited (1.7–2.6 additional hosts, $p < 0.01$). This suggests that AI-generated summaries in search results may stimulate further exploration rather than substituting for it, though this additional exploration may be unrelated to the original query. Session duration exhibits a strong positive effect: longer sessions are associated with more hosts visited, as expected.

These results are descriptive rather than causal. The choice to use an LLM within a session is endogenous to the underlying task, and differences across session categories may reflect variation in task complexity, information needs, or user intent that is not fully captured by observable controls. Accordingly, the estimated coefficients should be interpreted as characterizing systematic differences in browsing behavior across session types, rather than as causal effects of LLM usage on web traffic. Nonetheless, we find no evidence that the adoption of LLM tools is associated with a reduction in the number of websites visited per session.

Table 3: Session Type, AI Overview, and Browsing Outcome (LLM Users Only)

| Unique hosts visited | (I) | (II) |
|---|---|---|
| AI Overview Displayed | 2.614 | 1.663 |
| | (0.404) | (0.405) |
| log(Session Duration) | 7.080 | 6.505 |
| | (0.448) | (0.438) |
| LLM only sessions | −0.408 | −0.604 |
| | (1.010) | (0.987) |
| Mixed sessions (Search + LLM) | 2.330 | 2.141 |
| | (0.550) | (0.544) |
| Has Informational query | | 1.485 |
| | | (0.249) |
| Has Transactional query | | 2.744 |
| | | (0.225) |
| Has Navigational query | | 2.227 |
| | | (0.199) |
| Participant Fixed Effects | Yes | Yes |
| Observations | 20,805 | 20,805 |
| $R^2$ | 0.522 | 0.533 |
| Adj. $R^2$ | 0.517 | 0.529 |

Notes: Sample restricted to LLM users. The omitted session type category is *Search only*. Standard errors clustered by participant in parentheses. $p$-values are reported below standard errors and rounded to four decimals. For "Has Navigational query", the $p$-value is not available (insufficient within-participant variation under fixed effects).

### 4.3 Concentration Measures

The robustness analysis in section 4.1 suggests that weeks with higher LLM usage are associated with both increased search activity and broader web engagement, even after conditioning on overall online activity. However, user–week regressions cannot reveal *how* these patterns arise. In particular, they do not distinguish whether broader web activity reflects longer browsing spells, changes in navigation within a session, or differences in how users combine LLMs with other tools when pursuing specific tasks.

To shed light on the micro-level mechanisms behind these patterns, we turn to session-level data from the IBE and examine how browsing behavior varies across individual browsing sessions. We focus on downstream web activity within a session, measured by the number of unique URLs visited, and how it differs depending on whether users rely on search engines, LLMs, or a combination of both. Sessions are classified into four mutually exclusive categories (*search-only*, *LLM-only*, *mixed* (search and LLM), and sessions involving neither search nor LLM) and downstream behavior is summarized using three complementary measures:

- Externalization rate: This is the share of queries that lead to at least one external website visit in the five minutes following the query. It provides a simple check on whether users remain within the issuing platform or continue their activity elsewhere. A higher externalization rate means users are more likely to leave the platform and visit outside domains.

- Herfindahl–Hirschman Index (HHI): This measures how concentrated follow-up traffic is across domains. Formally,

$$HHI = \sum_d s_d^2,$$

  where $s_d$ is the share of external visits to domain $d$ within the five-minute window. If traffic is spread evenly across many sites, each $s_d$ is small and $HHI$ will be low. If most traffic goes to only one or two domains, $HHI$ rises toward one.

- Shannon entropy: This measures how dispersed traffic is across domains. It is defined as

$$H = -\sum_d s_d \log(s_d).$$

Higher entropy means visits are spread more evenly across many destinations, while lower entropy means activity is concentrated on a narrow set of sites. For example, if all visits go to a single domain, entropy equals zero.

Table 4: Externalization and Diversity Metrics by Platform and User Type (Session vs. 5-min Window)

| | **Entire session** | | | **5-min window** | | |
|---|---|---|---|---|---|---|
| **Platform** | **Ext. Rate** | **HHI** | **Shannon** | **Ext. Rate** | **HHI** | **Shannon** |
| Search (non LLM users) | 0.603 | 0.654 | 0.682 | 0.597 | 0.674 | 0.623 |
| Search (LLM users) | 0.641 | 0.630 | 0.709 | 0.636 | 0.644 | 0.673 |
| LLM (LLM users) | 0.705 | 0.634 | 0.726 | 0.699 | 0.639 | 0.711 |

Notes: Ext. Rate = mean externalization rate. HHI and Shannon are computed over category shares within the relevant time window.

Table 4 compares how often users leave the originating platform and how concentrated the resulting destinations are, measured both over the entire session and in the 5 minutes following each query. Externalization is common for both tools: in the 5-minute window, externalization ranges from 0.597–0.636 for search and reaches 0.699 for LLM queries. However, destinations reached after LLM interactions appear more diverse. In the 5-minute window, LLM queries exhibit a lower HHI (0.639) and a higher Shannon entropy (0.711) than search (HHI 0.644/0.674; entropy 0.673/0.623), with similar patterns in the session-wide measures. This suggests that, when LLMs do route users outward, they may disperse attention across a broader set of sites than traditional search, though these differences remain descriptive. Both platforms externalize users at comparable rates, but through opposite mechanisms: search concentrates traffic toward specific destinations; LLMs disperse users across a wider range of websites. As LLM adoption grows, this diversification effect may reshape traffic patterns across the web.

## 5 Conclusion

This paper studies how the adoption of large language models (LLMs) is associated with changes in online information-seeking behavior and web navigation. Combining nationally representative Comscore panel data with fine-grained session-level evidence from the IBE experiment, we document how users integrate LLMs into their browsing workflows and how this integration relates to search activity, downstream navigation, and the breadth of web exploration. Our analysis provides novel empirical evidence by being the first to jointly observe LLM use, traditional search, and full downstream browsing behavior at the individual level, combining population-representative data with fine-grained session-level evidence. Whereas prior work typically examines LLM interactions, search behavior, or platform outcomes in isolation, our approach allows us to trace how LLM adoption reorganizes information-seeking workflows and downstream content consumption across the web.

At the individual level, our results point to a reorganization—rather than a contraction—of online activity following LLM adoption. Exploiting staggered adoption timing, we show that users visit a broader set of distinct websites after adopting LLMs, alongside a short-lived increase in traditional search activity. These findings suggest that early narratives emphasizing the rise of zero-click searches and predicting a collapse of traffic to the open web are not supported by the observed patterns in our data. This underscores the importance of examining how users actually leverage LLMs relative to traditional search, with attention to browsing behavior across entire sessions rather than focusing solely on immediate post-query outcomes.

Our session-level evidence provides some telling insights regarding these dynamics: LLM use most often occurs in combination with search, and workflows that combine the two tools are longer and more exploratory than either search-only or LLM-only sessions. By contrast, LLM-only sessions tend to be shorter and narrower, suggesting that users rarely rely on LLMs as a stand-alone replacement for web navigation. Taken together, these patterns suggest that LLMs are best understood as complements to existing navigation tools rather than substitutes for them. By lowering the cognitive and time costs of formulating queries, synthesizing information, and coordinating multi-step tasks, LLMs appear to support richer information-seeking workflows. Users may complete specific subtasks more efficiently within the LLM interface, while continuing to navigate across external websites to evaluate sources, access content, or act on information. In this sense, LLMs seem to facilitate more structured and extended browsing sessions, even as they streamline parts of the search and discovery process.

We also document differences in how LLMs and traditional search engines route users to the external web. While both platforms frequently lead users to external websites, conditional on externalization (i.e., the probability of visiting external websites), LLM-assisted workflows are associated with more diverse downstream exploration. Importantly, these findings do not imply a permanent reallocation of aggregate web traffic, but rather reflect how LLMs currently support exploratory behavior within users' existing browsing routines. One plausible interpretation is that by reducing the time and cognitive effort required to complete specific sub-tasks, LLMs allow users to move on more quickly and allocate attention to additional activities, including further web navigation. More broadly, this distinction highlights that the implications of LLM adoption for content discovery depend not only on whether users leave the platform, but also on how referral traffic is distributed across destinations—suggesting a redistribution of attention rather than a contraction of the open web.

Overall, our work provides novel empirical evidence on how consumers integrate LLMs into real-world information-seeking behavior at a critical juncture where platform design is evolving faster than established measurement frameworks. Rather than replacing search or centralizing online activity, LLMs appear to augment the consumer's ability to engage in complex tasks, supporting richer and more exploratory browsing patterns that redistribute attention across the open web. By jointly observing LLM use, traditional search, and full browsing behavior at the individual level, our analysis establishes the empirical facts necessary to inform subsequent theoretical modeling, causal identification, and the normative evaluation of generative AI in digital markets. As these interfaces become more deeply integrated into the search experience, understanding how they shape the efficiency, structure, and scope of the customer journey remains a vital direction for marketing research.

## References

Alexander Bick, Adam Blandin, and David J. Deming. The rapid adoption of generative ai. NBER Working Paper 32966, National Bureau of Economic Research, Cambridge, MA, 2024. URL http://www.nber.org/papers/w32966.

Andrei Broder. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, 2002. URL https://sigir.org/files/forum/F2002/broder.pdf.

Gordon Burtch, Ashish Agarwal, Yili Hong, and Dongjie Liu. The consequences of generative ai for online knowledge and community engagement. *Scientific Reports*, 14(1):61221, 2024. doi: 10.1038/s41598-024-61221-0. URL https://doi.org/10.1038/s41598-024-61221-0.

Aaron Chatterji, Tom Cunningham, David J. Deming, Zoe Hitzig, Christina Ong, Carl Yan Shan, and Kevin Wadman. How people use chatgpt. (w34255), 2025. doi: 10.3386/w34255. URL https://www.nber.org/papers/w34255.

Mahe Chen, Xiaoxuan Wang, Kaiwen Chen, and Nick Koudas. Generative engine optimization: How to dominate ai search. *ACM Conference Proceedings*, 2025. URL https://arxiv.org/abs/2509.08919. Forthcoming / Preprint.

Cristobal Cheyre, Li Jiang, Florian Schaub, Zijun Ding, Cristiana Firullo, Yucheng Li, and Alessandro Acquisti. An experimental infrastructure to investigate the impact of online tracking, targeting, and advertising on consumer behavior and consumer welfare. *Working Paper*, 2024.

Francesco Decarolis and Gabriele Rovigatti. From mad men to maths men: Concentration and buyer power in online advertising. *American Economic Review*, 111(10):3299–3327, October 2021. doi: 10.1257/aer.20190811.

Kunal Handa, Alex Tamkin, Miles McCain, Saffron Huang, Esin Durmus, Sarah Heck, Jared Mueller, Jerry Hong, Stuart Ritchie, Tim Belonax, Kevin K. Troy, Dario Amodei, Jared Kaplan, Jack Clark, and Deep Ganguli. Which economic tasks are performed with ai? evidence from millions of claude conversations. *arXiv preprint*, 2025. URL https://arxiv.org/pdf/2503.04761.

Shiri Melumad and Jin Ho Yun. Experimental evidence of the effects of large language models versus web search on depth of learning. *PNAS Nexus*, 4(10):pgaf316, 2025. doi: 10.1093/pnasnexus/pgaf316.

Nicolas Padilla, H. Tai Lam, Anja Lambrecht, and Brett Hollenbeck. The impact of llm adoption on online user behavior. *Working Paper*, 2025. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5393256. London Business School and UCLA.

Sofia Eleni Spatharioti, David Rothschild, Daniel G Goldstein, and Jake M Hofman. Effects of llm-based search on decision making: Speed, accuracy, and overreliance. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan, 2025. ACM. doi: 10.1145/3706598.3714082.

Eduardo Vasconcelos, Evandro Cunha, Esteban Moro, Jussara Almeida, and Fabrício Benevenuto. Are large language models a threat to digital public goods? the impact of chatgpt on stack overflow. *arXiv preprint arXiv:2307.07367*, 2023. URL https://arxiv.org/abs/2307.07367.

Ruiyun Xu, Yue Feng, and Hailiang Chen. Chatgpt vs. google: A comparative study of search performance and user experience, 2023. Available at SSRN: https://ssrn.com/abstract=4498671.

Ziang Zhang, Minda Jia, Beining Yao, Sauvik Das, Adam Lerner, Dakuo Wang, and Tian Li. It's a fair game, or is it? examining how users navigate disclosure risks and benefits when using llm-based conversational agents. *arXiv preprint arXiv:2309.11653*, 2023. URL https://arxiv.org/abs/2309.11653.

# A Appendix

## A.1 Recruitment and Descriptive Statistics

Our analysis draws on two data sources: the comScore panel covering the period 2019–2024 and the Internet Behavior Experiment (IBE). As of January 2026, the IBE is ongoing and includes 308 participants out of a target sample of 1,200. To date, the primary recruitment channel has been social media, using a dedicated advertising campaign on Meta. Figure 10 shows an example of an advertisement used in this campaign. Beginning in January 2026, we are expanding recruitment through university subject pools in the United States, including Stanford University, Cornell University, the University of Michigan, and Carnegie Mellon University.


Figure 10: Example of the recruitment advertisement shown to users.

Table 5 shows the main demographic characteristics of participants in the Comscore and the IBE samples. Overall, the IBE sample is older and more female than the Comscore panel. Nevertheless, the distribution of key demographic characteristics is broadly similar across the two samples, suggesting that the IBE participants are reasonably representative of the underlying Comscore population.

Table 5: Participant Demographics and Characteristics (Full Sample vs. LLM Users)

| Characteristic | CS Sample | | IBE Full Sample | | LLM Users | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| **Age Groups** | | | | | | |
| Under 30 years | 251,335 | 31% | 21 | 6.8% | 17 | 9.1% |
| 30–59 years | 419,705 | 52% | 198 | 64.3% | 118 | 63.1% |
| 60+ years | 137,939 | 17% | 88 | 28.6% | 51 | 27.3% |
| **Gender** | | | | | | |
| Male | 460,577 | 57% | 123 | 39.9% | 72 | 38.5% |
| Female | 348,402 | 43% | 180 | 58.4% | 114 | 61.0% |
| **Education Level** | | | | | | |
| High school diploma | | | 70 | 22.7% | 37 | 19.8% |
| Bachelor's degree | | | 115 | 37.3% | 70 | 37.4% |
| Master's degree or above | | | 79 | 25.6% | 51 | 27.3% |
| Other | | | 44 | 14.3% | 29 | 15.5% |
| **Employment Status** | | | | | | |
| Employed full-time | | | 120 | 39.0% | 78 | 41.7% |
| Employed part-time | | | 59 | 19.2% | 39 | 20.9% |
| Retired | | | 56 | 18.2% | 30 | 16.0% |
| Student | | | 13 | 4.2% | 8 | 4.3% |
| Disabled | | | 24 | 7.8% | 10 | 5.3% |
| Unemployed | | | 34 | 11.0% | 21 | 11.2% |
| Prefer not to answer | | | 2 | 0.6% | 1 | 0.5% |
| **Annual Income** | | | | | | |
| Under $30,000 | 150,532 | 18.6% | 82 | 26.6% | 46 | 24.6% |
| $30,000 to $49,999 | 140,488 | 17.3% | 46 | 14.9% | 24 | 12.8% |
| $50,000 to $74,999 | 102,250 | 12.6% | 61 | 19.8% | 38 | 20.3% |
| $75,000 to $99,999 | 67,480 | 8.3% | 47 | 15.3% | 29 | 15.5% |
| $100,000 to $149,999 | 101,710 | 12.6% | 42 | 13.6% | 30 | 16.0% |
| $150,000 or more | 204,673 | 25.3% | 22 | 7.1% | 14 | 7.5% |
| Prefer not to answer | 41,846 | 5.2% | 8 | 2.6% | 6 | 3.2% |
| **Sample Size** | **808,979** | | **308** | | **187** | |

### A.2 Potential extensions

One possible extension of our current analysis focuses on the externalization rate and the various types of browsing sessions. To understand whether the session type predicts the likelihood of visiting external websites, we regress the externalization according to the following model

$$Ext_{i,t} = \alpha + \boldsymbol{\beta} SessionType + \gamma \cdot Prompts_{i,t} + \delta \cdot (Prompts \times SessionType) + \zeta_i + \epsilon_{i,t} \quad (2)$$

The dependent variable is an indicator equal to one if a session includes at least one visit to a website other than search engines or LLM platforms. We estimate a linear probability model with participant fixed effects, using search-only sessions as the baseline category. We control for the total number of prompts in the session, defined as the sum of search queries and LLM messages. In principle, more prompts could increase externalization by generating more opportunities to click outward, or decrease it if users remain within the querying platform without finding satisfactory results.

### A.3 Query classification prompts

We classify user queries using a local LLM (Llama 3.2 3B) combined with deterministic keyword-based rules. We employ distinct classification strategies for LLM conversations and search queries, reflecting their different structures and intent taxonomies.

#### A.3.1 LLM conversation classification

LLM conversations are classified using a two-stage pipeline.

**Stage 1: Summarization.** Multi-turn conversations are first condensed into a single intent statement. We prompt the model with:

```
What does the user want? Answer in 5-10 words starting with a verb.
```

This produces action-oriented summaries (e.g., "Write an email to my boss", "Explain quantum entanglement") that serve as input to the classification stage.

**Stage 2: Classification.** The summary is classified into one of six intent categories:

```
Classify into A, B, C, D, E, or F:

A = LEARN: questions, explain, understand, verify

B = CREATE: write, rewrite, edit, generate, code, translate, summarize

C = BUY/FIND: purchase, find store/restaurant

D = WEBSITE: go to specific website

E = CHAT: greetings, hello, small talk

F = EMOTIONAL: personal depression, anxiety, sadness

Request:  "{summary}"

Answer (one letter):
```

**Hybrid classification logic.** To improve reliability, we implement a keyword-based detection layer that takes precedence over LLM outputs for unambiguous patterns. Queries beginning with interrogative structures (e.g., "what is", "how does") are directly classified as Informational, while explicit creative verbs (e.g., "write", "code", "generate") trigger Creative classification. This hybrid approach reduces misclassification errors common with small language models, particularly the tendency to over-assign Creative labels to explanatory requests.

**Subcategory inference.** Within each macrocategory, subcategories are assigned using deterministic keyword matching on the original query text. For instance, Informational queries containing advisory language ("should I", "recommend") are labeled as *Advisory/Planning*, while those with computational terms ("solve", "calculate") are labeled as *Reasoning/Problem-solving*. Creative queries are subdivided into *Writing/Content Creation*, *Coding/Programming*, *Summarization*, *Translation*, and *Image Generation* based on task-specific keywords.

**Preprocessing.** Prior to LLM processing, we sanitize queries by replacing profanity and sensitive content with neutral placeholders to prevent model refusals. Temperature is set to 0.1 to maximize classification consistency.

#### A.3.2 Search query classification

For search queries, we employ a simplified three-category taxonomy based on Broder (2002): Informational, Transactional, and Navigational.

```
Classify this search query into ONE category.

QUERY: "{query}"

CATEGORIES:

A = INFORMATIONAL: User wants to learn or understand something

- "what is X", "how to Y", "why does Z", definitions, explanations

B = TRANSACTIONAL: User wants to BUY something or FIND a place

- "best laptop", "restaurants near me", "buy iPhone"

C = NAVIGATIONAL: User wants to go to a specific website

- "facebook login", "gmail", "amazon"

Choose A if unsure.

ANSWER (one letter A/B/C):
```

Given the brevity of search queries, no summarization stage is required. Subcategories are inferred deterministically: Informational queries are subdivided into *Fact Q&A*, *Explanatory/Conceptual*, *How-to/Tutorial*, and *Research/Learning*; Transactional queries into *Product Search*, *Service Search*, *Local Business/Place*, and *Entertainment/Media*.

To efficiently process the large volume of search queries ($N > 200{,}000$), we implement a deduplication strategy: queries are normalized to lowercase, classified once per unique string, and mapped back to all original instances. This reduces the number of LLM calls while maintaining full coverage.