The Impact of Google AI Overviews on Publisher Traffic and User Experience: Evidence from a Field Experiment*

Saharsh Agarwal[†] Ananya Sen[‡]

June 17, 2026

Abstract

Online search engines rely on external content to respond to user queries, while publishers depend on search engines for traffic. The integration of GenAI summaries into search interfaces, such as Google's AI Overviews (AIOs), may reduce users' incentives to visit downstream websites, yet causal evidence remains limited. We provide causal evidence by implementing a field experiment, using a custom Chrome extension, in which users are randomly assigned to either standard Google Search with AIOs or a version without them. Conditional on appearing, AIOs reduce outbound organic clicks by 39.8% and increase zero-click searches by 34.5%, without affecting sponsored clicks or overall search frequency. Engagement on downstream websites, conditional on clicks, is not different across groups. Survey evidence shows no measurable improvements in perceived search quality or ease of finding information. Exploratory evidence from an AI Mode condition, in which users are redirected to a Gemini-powered conversational AI interface, suggests even greater reductions in traffic to downstream publishers. Overall, the results suggest that AIOs divert traffic away from publishers without improving the user experience or quality of engagement for websites.

Keywords: Generative AI, Online Search, AI Overviews, Field Experiment.

---

*This study was pre-registered in the AEA RCT Registry (ID: AEARCTR-0017393). The study protocol was approved by the Institutional Review Board at the Indian School of Business (ISB). We are grateful to several seminar and conference audiences for feedback. We are indebted to the Office of Advancement and the Office of Alumni Engagement at ISB for facilitating funding for this project through the Impact Innovation Grant as well as the Emerging Scholar Research Award. We are thankful to Rohit Akash Rajendran for excellent research assistance. We gratefully acknowledge the ISB Institute of Data Science for funding the research assistance.

[†]Indian School of Business. Email: saharsh_agarwal@isb.edu

[‡]Carnegie Mellon University. Email: ananyase@andrew.cmu.edu

1 Introduction

Google Search has long served as a primary gateway to information on the internet, accounting for a dominant share of global search activity (Allcott et al., 2025). For many downstream content providers (e.g. news publishers, review platforms, informational websites, etc.), visibility on the search engine results page (SERP) is a critical driver of user acquisition and monetization. This relationship is fundamentally symbiotic: while publishers rely on search engines for traffic and discovery, search engines depend on a steady supply of high-quality content to deliver relevant results and sustain user engagement on their own platform. Over the past decade, this interdependence has intensified, giving rise to tensions over value capture, with content producers arguing that platforms appropriate and monetize their content without adequate compensation. These concerns have fueled regulatory scrutiny and legal disputes across multiple jurisdictions (De Vynck, 2021).

The recent integration of generative artificial intelligence (GenAI) into search interfaces represents a potentially significant shift in this ecosystem. In particular, Google's introduction of AI Overviews (AIOs henceforth), AI-generated summaries that appear prominently in response to certain queries, changes how users interact with search results (Google, 2024). Rather than directing users to external websites to synthesize information, AIOs consolidate content from underlying sources and present it directly on the search engine results page (SERP). While it has been suggested that such summaries may improve user convenience, downstream stakeholders have raised concerns that AIOs may further reduce referral traffic (Bellan, 2025). Moreover, AIOs use the underlying data from these content websites, which is necessary to produce the AI summary in the first place, raising concerns that search platforms may increasingly internalize content consumption while externalizing content production costs.

These concerns have quickly entered policy debates. Competition authorities in the United Kingdom have suggested that publishers should have the option to opt out of AI-generated summaries, and complaints regarding AIOs have been filed with European and Brazilian antitrust regulators (Muvija, 2026). In the United States, this issue has raised concerns that Google may be leveraging its monopoly power in the search market in ways that violate the Sherman Act (Singh and Scott Morton, 2026). Moreover, shareholders of information platforms (e.g., Reddit) have filed lawsuits against the platforms over concerns that Google AIO is stealing traffic.[1] At the core of these discussions are a few fundamental empirical questions which we aim to answer. Do AI-generated search summaries materially reduce engagement with the very sources from which they draw information? Do users find such AI summaries beneficial? Do downstream publishers benefit from more streamlined traffic in the presence of AIOs?

Despite the importance of this issue for platform design, publisher sustainability, and competition policy, systematic causal evidence on these questions remains limited.[2] Importantly, the direction and magnitude of these effects are theoretically ambiguous. On the one hand, AIOs may substitute for clicks by directly satisfying user information needs. On the other hand, by prominently citing and linking to underlying sources, AIOs may encourage users to visit downstream websites, thereby increasing publisher traffic. As such, the net effect on publisher traffic is ultimately an empirical question that depends on how these competing mechanisms interact in practice. The implications for user experience are also unclear. While removing AIOs may increase downstream engagement, it could come at the cost of reduced convenience or slower access to information, making the welfare effects of such interventions similarly ambiguous. Finally, a more streamlined search process might also increase the time users spend on downstream websites, conditional on clicking.

This paper provides novel and rigorous empirical evidence on how AI Overviews affect user behavior and downstream engagement. We do so by developing a custom-built Chrome extension and using a field experiment that directly manipulates the search experience of real users in a natural browsing environment. Our approach allows us to isolate the causal impact of AIO exposure on click behavior, time spent, and user-perceived utility. We recruited over 1,000 participants who installed the extension and consented to passive measurement of their browsing behavior over a two-week period. Participants were randomly assigned to one of three conditions. In the control group, users experienced Google Search as usual, with AIOs appearing when available. In the primary treatment group ("Hide AIO", HAIO), the extension removed AIOs in real time whenever they would otherwise appear.[3] In a third condition ("AI Mode"), users were redirected to Google's AI-focused search interface for all queries.

[1] See https://www.dandodiary.com/2025/06/articles/artificial-intelligence/reddit-downplayed-impact-of-googles-ai-related-changes-suit-alleges/ for some more details.

[2] There are some surveys and industry reports that suggest a negative relationship between AIOs and downstream website traffic. However, because these analyses rely on observational data and do not exploit exogenous variation in AIO exposure, they cannot identify the causal effect of AIOs on user behavior. See, for example, https://www.pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appears-in-the-results/ and https://ahrefs.com/blog/ai-overviews-reduce-clicks/.

[3] Importantly, the remaining search results were automatically shifted upward to fill the space, ensuring that the interface remained seamless and that no blank or disrupted layout signaled the intervention to users.

Our results yield several insights. First, descriptively, we document that AIOs were triggered in approximately 41% of observed queries, indicating that their prevalence is not marginal. Second, for queries in which an AIO was (intended to be) triggered, removing AIOs increases outbound organic clicks from 0.37 to 0.62 per search, implying that AIOs reduce clicks by 39.8%. Similarly, we find that the probability of a zero-click session is 34.5% higher in the control group compared to HAIO users for AIO-triggering queries.[4] Third, we find no meaningful change in clicks on sponsored links or other Google-owned page elements, suggesting that the primary behavioral substitution occurs between AI summaries and organic website visits rather than across the broader page ecosystem. Finally, we find no change in the number of searches that users make in the HAIO and control groups, implying that the reduction in clicks per search translates into a meaningful decline in overall outbound web traffic.

Heterogeneity analysis indicates that the treatment effect on clicks is concentrated in cases where the AIO appears at the top of the SERP, aligning with prior evidence that top-of-page real estate captures disproportionate user attention (Agarwal et al., 2011). Moreover, looking at heterogeneity by search result rank, we find that search results ranked 1, 2, and 3 experienced the largest increase in clicks in the HAIO group relative to the control.

Importantly, post-experiment survey responses indicate that users' perceived utility and satisfaction with search were unaffected by the removal of AIOs, suggesting that they divert traffic away from publishers without delivering measurable improvements in user experience. Moreover, our analysis of various measures of time spent on the downstream website, conditional on a click, shows no difference between the two groups. We find similar rates of returning to the search page, bounce rates and overall time spent, suggesting that AIOs aren't necessarily sending 'higher-quality' clicks from the downstream publisher's perspective.[5]

We complement these findings with descriptive evidence from the AI Mode condition, which represents a substantial departure from traditional search. We anticipated and found substantial differential attrition in this arm, suggesting meaningful user resistance to a fully AI-mediated search experience. Among users who remained compliant in the AI Mode condition, publishers' overall click-through rates and user satisfaction were substantially lower than in the control group. While selection limits causal interpretation in this arm, the patterns

[4] Since AIOs are triggered in approximately 41% of searches in our sample, these conditional effects translate into an average reduction in outbound organic clicks of roughly 18.5% and an average increase in zero-click searches of roughly 20.4% across all searches.

[5] This result stands in contrast to public claims made by Google about bounce rates coming from AIOs (Southern, 2026b).

provide suggestive evidence about potential equilibrium effects if AI-forward search modes become more prevalent.

This paper contributes to three related strands of literature. First, a growing body of work compares the use of large language models and AI assistants to traditional search (Agarwal and Ghosh, 2026; Aral et al., 2026; Kaiser and Schulze, 2026; Gholami et al., 2026; Padilla et al., 2025). These studies are complementary to ours, as they primarily rely on observational comparisons of search behavior across traditional platforms such as Google and AI assistants such as ChatGPT. Khosravi and Yoganarasimhan (2026) provide early evidence on this question by leveraging the rollout of AIOs in the US and comparing traffic to English Wikipedia articles with corresponding articles in other languages, which were presumably unaffected. Zhang et al. (2026) use a similar approach to understand the impact of Google AIOs on Reddit traffic using variation in Google's content moderation policy vis-a-vis Reddit. While these designs are compelling, they rely on observational variation in the rollout of AIOs and therefore cannot fully rule out contemporaneous changes in search behavior, competing AI products, platform algorithms, or broader information consumption patterns that may differentially affect treated and control content over the same period. By contrast, our study provides causal evidence from a randomized field experiment embedded directly in users' natural browsing environments using an extension.[6] Moreover, our design enables a more granular decomposition of engagement across the broader web ecosystem, including differential effects on organic versus sponsored links, bounce rates, and self-reported search experience.

Second, our work relates to studies examining AI-generated summaries in other contexts, such as product reviews (Su et al., 2024; Zhai and Ching, 2026) and GenAI search capabilities on content sharing platforms (Bai et al., 2026). We contribute to this strand by examining multiple behavioral margins—clicking, browsing, and time-based engagement—and by exploring positional mechanisms specific to search environments. Importantly, the search engine setting exhibits unique competitive dynamics among outbound links, making the welfare implications of AI summaries distinct from those in previously studied domains.[7]

Third, we contribute to the literature on conflicts between news publishers and digital platforms. A substantial body of work examines how services such as Google News, Facebook,

[6] This approach is related to a small yet growing literature that uses browser extensions to modify users' online experiences in situ (Piccardi et al., 2025; Beknazar-Yuzbashev et al., 2025).

[7] A complementary study by Ng and Wessel (2026) provides evidence on how deployment of Google AI Overviews is strategic based on search intent with potential implications for self-preferencing.

and other aggregators affect publisher traffic (Calzada and Gil, 2020; Song and Manchanda, 2025; Chiou and Tucker, 2017). Our study introduces generative AI as a new technological layer in this longstanding relationship, in which data from downstream websites is used as a key input to the GenAI product. Whereas earlier disputes focused on link aggregation and snippet display, AIOs represent a more intensive form of content synthesis that may further shift value capture toward the platform. Our paper informs ongoing regulatory discussions in the United Kingdom, the European Union, and other jurisdictions considering the competitive implications of AI-mediated search.[8]

The remainder of the paper proceeds as follows. Section 2 describes the data and experimental design. Section 3 presents the econometric framework, Section 4 reports the empirical results, and Section 5 concludes.

2 Data and Experimental Design

At the core of the data and experimental design is a custom-built Chrome desktop browser extension. The underlying architecture of our extension builds on Webmunk (Farronato et al., 2024). The extension enables the collection of high-frequency behavioral data, including URL visits, clicks, and time spent, while simultaneously allowing us to modify the content of the Google search engine results page (SERP) in a controlled and unobtrusive manner.

This data allows us to identify all Google searches, along with detailed information about the corresponding search results page. In particular, we observe whether an AI Overview (AIO) was shown, its relative position on the SERP, the full set of search results including titles, links, and their positions on the page, as well as all user clicks on both organic and sponsored results. For participants in the AI Mode condition, we also observe the full set of queries entered by users within the conversational interface (see details on data and sample construction procedures in Appendix B).

2.1 Recruitment and Study Flow

Participants were recruited from Prolific in a staggered manner between 7 January and 27 January 2026. Recruitment proceeded in multiple stages (see Figure A.1 in the Appendix for an

[8]Our paper also relates more generally to a literature that studies the potential negative impact of AI products on the quality of their own training data, especially in the context of LLMs and user-generated content online (Quinn and Gutt, 2025; Burtch et al., 2024; Peukert et al., 2024; Goldberg and Lam, 2025; Zhou et al., 2026).

overview). First, participants completed a pre-screening survey. In this stage, we restricted the sample to individuals who (i) reported Google as their primary search engine, (ii) used Chrome as their only browser, given that the extension was compatible only with Chrome, and (iii) reported using their current desktop for general browsing activities beyond simply survey platforms such as Prolific.

Participants who passed the pre-screening stage were asked to provide informed consent and then administered a baseline survey, which elicited information on their trust in online AI-generated information and their comfort with using AI tools for everyday information-seeking tasks. Following this, participants were directed to install our browser extension, which was listed on the Chrome Web Store and had undergone Google's security and compliance review. After installation, the extension conducted a browser history check based on the preceding 20 days. To qualify for the main study, participants were required to meet minimum activity thresholds based on their browser history in the last 20 days: (i) at least 10 visits to Google, (ii) visits to at least 25 unique domains, and (iii) browsing activity on at least 10 distinct days.

Participants who met these conditions were invited to continue in the main study and were eligible for an additional bonus payment, contingent on keeping the extension installed for 2 weeks. The bonus was paid on the 16th day following installation, after which participants were invited to complete an endline survey. Following the completion of the two-week study period, treatment assignments were reversed, with Control participants moved to HAIO and vice-versa, allowing us to examine within-user changes in behavior around the switch (see Section 4.2.2). Participants who uninstalled the extension before the end of the two-week study period were not eligible for the bonus. A total of 1,065 US-based participants passed the browser history checks and participated in the main two-week study (Figure B.1). Data collection began on January 7, 2026 (the first day of recruitment) and concluded on February 10, 2026, with two weeks of browsing data observed per participant.

2.2 Randomization and Treatment Arms

Eligible participants who passed all data checks were randomly assigned in real time to one of three experimental conditions. The two primary groups of interest were the control group (status quo) and the "Hide AI Overview" (HAIO) treatment group, each assigned with a probability of 36 percent. The remaining assignment probability of 28 percent corresponded to a third "AI Mode" group. This unequal allocation reflects the study's primary focus on estimating the causal effect of removing AIOs relative to the status quo, while still allowing exploratory analysis of the emerging conversational AI Mode interface. The control group experienced no changes to their browsing environment (Figure 1a). When these participants conducted searches on Google (or via their browser's search bar), they saw the standard search engine results page (SERP) exactly as Google presented it. If a given query triggered an AIO, the overview appeared normally. This group, therefore, represents the status quo user experience and provides the primary counterfactual against which treatment effects are measured.

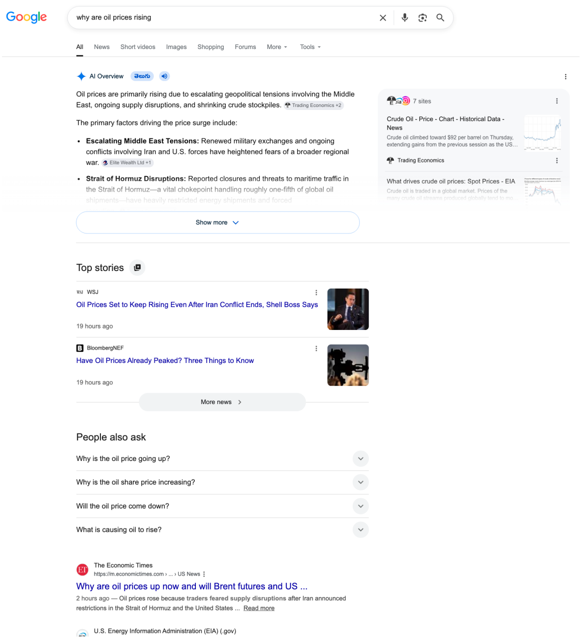
Participants assigned to the HAIO group experienced the study's main experimental manipulation. For these users, the Chrome extension automatically removed any AIOs that would otherwise have appeared on the SERP during the two-week experimental window (Figure 1b). Importantly, the remaining search results were automatically shifted upward to fill the space previously occupied by the AIO, ensuring that the interface remained seamless and visually indistinguishable from a standard results page without AIOs. All other aspects of the search page remained unchanged. This design isolates the causal impact of AIO exposure by holding the broader search environment constant.

The third group—the AI Mode group—was exposed to Google's Gemini-powered conversational search interface whenever they conducted searches (Figure 2). This condition represents a more substantial departure from the traditional SERP experience. Because this interface is relatively new, we anticipated higher attrition in this arm. Accordingly, analyses involving the AI Mode group are conceptualized as primarily descriptive rather than strictly causal. Nevertheless, including this group provides useful contextual evidence about how users respond to a fully AI-mediated search experience.

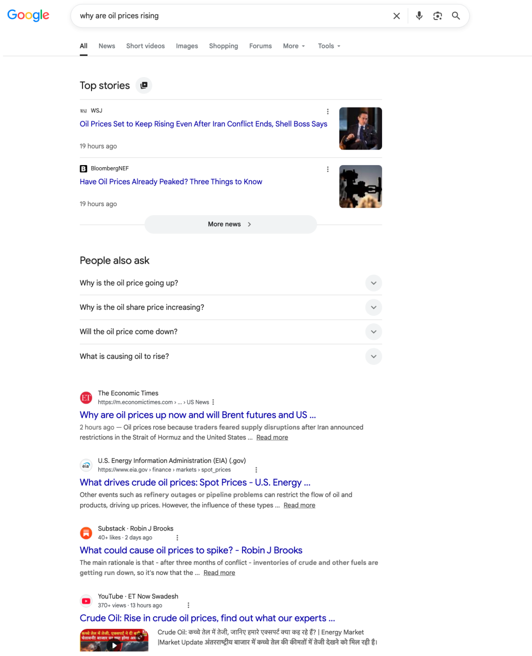
2.3 Internal Validity

The extension was designed to operate seamlessly in the background, without introducing any noticeable changes to page load times or overall browsing performance.[9] All modifications to the search interface were executed client-side and in real time, ensuring that users did not experience any visible lag, flicker, or disruption when interacting with the SERP. This was critical not only for preserving a natural browsing environment and minimizing the likelihood that participants would detect the intervention, but also for maintaining an "all else equal" comparison

[9] Mean page-load times were statistically indistinguishable across groups (HAIO: 2.02 s; Control: 1.97 s; $p = 0.84$).



(a) Control



(b) Hide AIO (HAIO)

Figure 1: Search interfaces in the Control and HAIO conditions.

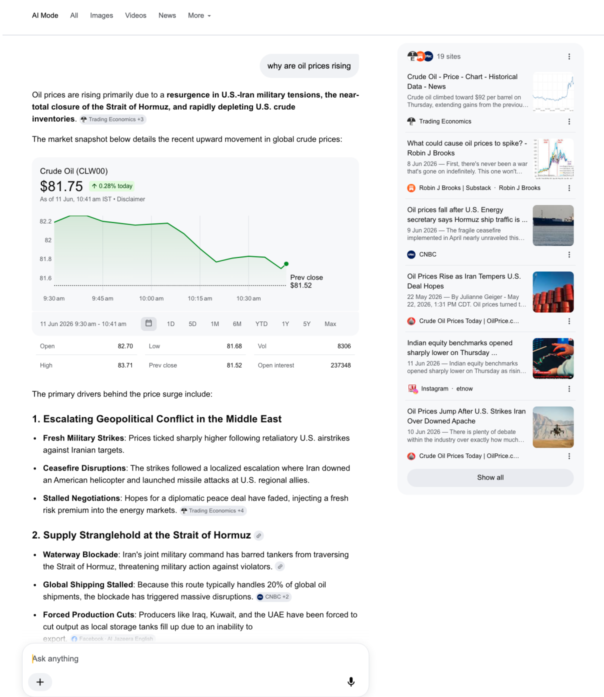


Figure 2: AI Mode interface.

across conditions, thereby strengthening internal validity.[10]

To ensure comparability across experimental conditions, we conducted balance tests on a range of pre-treatment characteristics, including demographic attributes (e.g., age, gender from Prolific), baseline attitudes toward AI (e.g., AI trust from our baseline survey), and prior browsing behavior (from the extension). Table 1 reports the results of the balance tests. Across all observable characteristics, we find no statistically significant differences between the experimental groups, consistent with successful randomization.

Table 1: Balance in Means Across Experimental Groups

Variable	N	Control (C) (N = 396)	AI Mode (A) (N = 295)	HAIO (H) (N = 374)	p(C - A)	p(C - H)	p(A - H)
Age	1,063	38.22	38.08	39.03	0.87	0.37	0.32
Male	1,063	0.43	0.40	0.41	0.50	0.55	0.91
White	1,060	0.70	0.75	0.71	0.21	0.87	0.27
Latitude	1,065	37.36	37.31	37.71	0.90	0.33	0.32
Longitude	1,065	-91.04	-91.80	-89.85	0.53	0.28	0.12
Trust in AI info	1,065	3.37	3.39	3.34	0.71	0.75	0.51
Comfort with AI tools	1,065	3.59	3.52	3.64	0.38	0.46	0.12
Prior Google visits	1,045	242.51	241.54	219.02	0.97	0.29	0.40
Prior active browsing days	1,045	18.78	18.55	18.88	0.35	0.67	0.19
Prior unique domains visited	1,045	203.10	201.72	214.01	0.89	0.24	0.22

Notes: Trust in AI information and comfort with AI tools are measured on a Likert scale from 1 to 5. All prior browsing variables are computed over the 20 days preceding enrollment. Prior Google visits refer to the total number of unique URL visits to Google (including refreshes). Prior active browsing days measure the number of distinct days with any recorded browsing activity. Prior unique domains visited refers to the number of distinct domains visited.

While randomization ensures balance at baseline, differential attrition across experimental groups can undermine this comparability. We therefore examine participant retention across treatment arms. We measure attrition using two complementary indicators. First, we track whether participants uninstall the extension, an action directly observed by the extension itself. Second, we capture attempts to circumvent the intervention.[11]

Figure A.2 plots survival rates—defined as the proportion of participants who remain active (i.e., have neither uninstalled the extension nor engaged in bypass behavior)—over time since installation. Attrition patterns for the Control and HAIO groups are broadly similar over the

[10] To further illustrate the user experience under each condition, we provide short video demonstrations of the extension in operation. Video demonstrations of the browsing experience under each experimental condition are available at https://youtu.be/drqSQN6ZaIc (Control), https://youtu.be/2z5-Pf64RcI (HAIO), and https://youtu.be/ADjnRaA0_HI (AI Mode).

[11] This channel is primarily a concern in the AI Mode condition, where users are exposed to a substantially different interface; in contrast, participants in the Control and HAIO groups experience minimal or no perceptible changes (confirmed by our endline survey). In particular, some users in the AI Mode group install alternative browser extensions that bypass our modifications. These extensions work by appending "udm=14" to Google search queries, which forces the search interface into a plain web-only mode. We treat the first occurrence of such a query as an attrition event, as it reflects a deliberate attempt to bypass the assigned condition.

study period. In contrast, the AI Mode group exhibits substantially higher attrition, with a sharper decline in the share of active users over time. Given this differential attrition, we treat the results from the AI Mode arm as primarily exploratory rather than causal.

We complement these behavioral checks with evidence from the endline survey, which provides additional insights into compliance and potential behavioral responses to the intervention. Over 95% of users in both the Control and HAIO groups report either not noticing any changes or being unsure (Figure A.5).[12] Finally, we find no difference across groups in self-reported compliance with the study (Figure A.6) and shifts to a different browser or device as seen in Figures A.3 and A.4 in the Appendix.

3 Econometric Framework

We begin by focusing on the HAIO treatment arm relative to the control group, which provides the cleanest setting for causal inference. We return to the AI Mode condition later and present the results as exploratory.

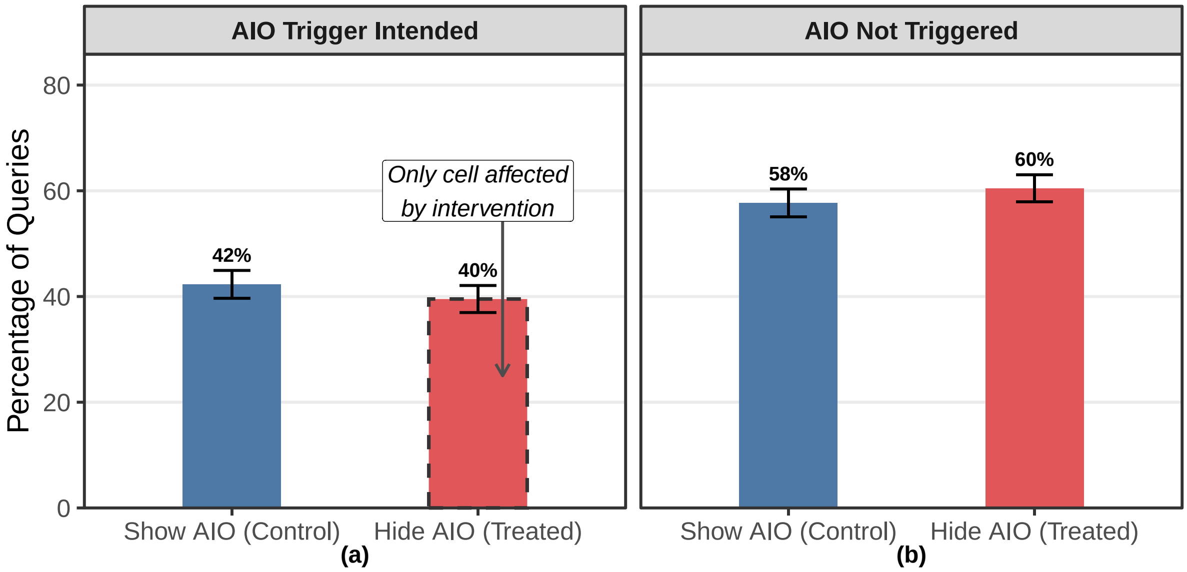
We first document the prevalence of AIOs in our sample. Figure 3 shows that AIOs are triggered in approximately 41% of observed queries. This estimate is somewhat lower than that reported in Aral et al. (2026) (67%), likely reflecting differences in query composition: while Aral et al. (2026) rely on a curated set of queries, our measure is based on organically generated user queries in a natural browsing environment. Importantly, this prevalence is similar across the Control and HAIO groups, suggesting that removing AIOs does not meaningfully alter the types of queries users submit. This provides initial assurance that our estimates will primarily capture differences in responses to search results rather than in underlying search intent.

At a conceptual level, total clicks from Google can be decomposed into the product of the number of searches and the number of clicks per search. AIOs may affect both margins: they may reduce (or increase) the number of clicks on a given query (intensive margin), and they may also alter users' overall search frequency (extensive margin). Hence, we structure our empirical analysis to examine these two channels separately.

Our empirical strategy exploits the fact that the HAIO intervention only affects queries for which an AIO would have been shown (panel (a) of Figure 3). We therefore structure all subsequent analyses by partitioning queries into two groups: those for which an AIO was intended

[12] For those who reported being unsure, we manually reviewed open-ended responses and found no evidence that participants correctly inferred the nature of the intervention.

Figure 3: AIO Prevalence



Notes: The figure reports the share of queries in each condition for which an AIO was (intended to be) triggered (panel a) and not triggered (panel b). The estimates are computed using search-level data across the Control and HAIO groups over the two-week period following enrollment. For queries in the HAIO group, AIO presence is inferred based on whether an overview would have been shown in the absence of the intervention. Error bars denote 95% confidence intervals clustered at the individual level.

to be triggered, and those for which it was not.[13] Queries in the latter category serve as a natural placebo group (panel (b) of Figure 3), while we expect treatment effects to be concentrated among queries in the former group.

Across the 396 users in the control group and 374 in HAIO, we observe a total of 68,089 unique searches over the two-week period following each participant's enrollment. We focus on the search-level data to estimate the following regression at the individual-i search-s level:

$$Y_{is} = \gamma + \beta T_i + \epsilon_{is} \tag{1}$$

where T_i is an indicator for assignment to the HAIO treatment and β captures the causal effect of the HAIO on the outcome of interest. Because treatment is randomly assigned at the individual level, we estimate this specification without additional control variables in our baseline. Standard errors are clustered at the individual level to account for within-user correlation in outcomes across searches.

[13] We use the wording 'intended to be triggered' because in the HAIO group, the AIO was not actually shown.

4 Empirical Results

4.1 Search-level Analysis: Engagement with SERP Content

We examine three primary outcomes at the search level. First, we measure the number of outbound organic clicks, defined as clicks to external websites from anywhere on the SERP. Second, we consider the likelihood of a zero-click search, defined as an indicator for searches with no outbound organic clicks. Finally, we examine the number of sponsored clicks.

Figure 4 presents the main results. We begin by focusing on queries for which an AIO was (intended to be) triggered in panel (a). In these queries, removing AIOs leads to a large and statistically significant increase in outbound organic clicks. The average number of external clicks per search increases by 39.8% from 0.37 in the control group to 0.62 in the HAIO group. Consistent with this, the likelihood of a zero-click search declines by 34.5% from 0.73 to 0.54 in the HAIO group (panel (b)). In contrast, we find no meaningful differences in outcomes (number of organic clicks and probability of zero-click search) between the Control and HAIO groups for queries in which an AIO was not intended to be triggered, as seen in Appendix Figure A.7. The absence of effects in this placebo group strengthens the causal interpretation of our estimates.

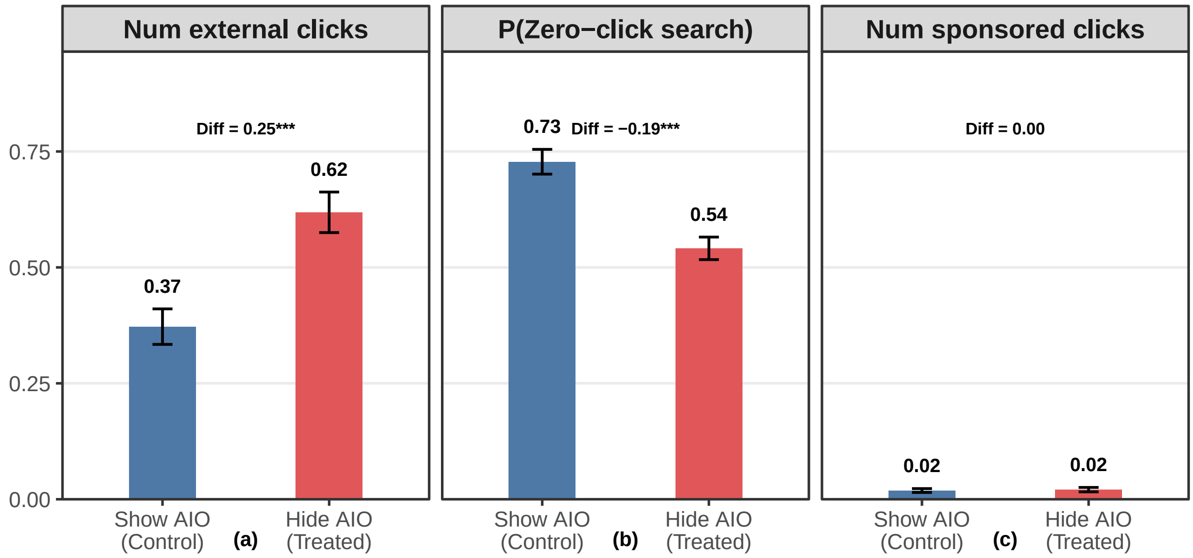
We find no meaningful changes in other forms of engagement. Sponsored clicks are unaffected (panel (c) of Figure 4), indicating that the primary behavioral substitution occurs between AI-generated summaries and organic website visits, rather than across the broader page ecosystem. Because sponsored clicks are rare (approximately 0.02 clicks per search), detecting effects is difficult without substantially larger samples.[14]

Our estimates also highlight that the aggregate impact of AIOs depends critically on their prevalence. Currently, AIOs are triggered in roughly 41% of queries in our sample. As the share of queries displaying AIOs increases over time (Aral et al., 2026), the overall reduction in outbound traffic could become substantially larger, even if the per-query effect remains constant. By reducing the need to click through to underlying content, AIOs shift the role of search engines from intermediaries that facilitate discovery to destinations that aggregate and deliver information. This shift has important implications for content producers. Much of the online

[14]We also analyze the number of internal clicks, which capture interactions within Google's ecosystem, such as navigating to images, news, pagination, or other result tabs, as well as engaging with query refinements or suggested searches. Similarly to sponsored clicks, the number of internal clicks remains largely unchanged, as seen in Appendix Figure A.8.

content ecosystem is sustained by referral traffic from search, which underpins both advertising revenue and subscription models. To the extent that AIOs reduce outbound clicks, they may weaken these incentives by lowering the returns to producing high-quality, discoverable content, potentially affecting both the quantity and quality of content over time.

Figure 4: Search-level Treatment Effects



Notes: The figure includes queries for which an AIO was (intended to be) triggered. Each bar represents the corresponding average of search-level outcomes for the Control (Show AIO) and HAIO (Hide AIO) groups. Error bars denote 95% confidence intervals clustered at the individual level. The difference (HAIO − Control) shown between the bars corresponds to the estimate from Equation 1. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

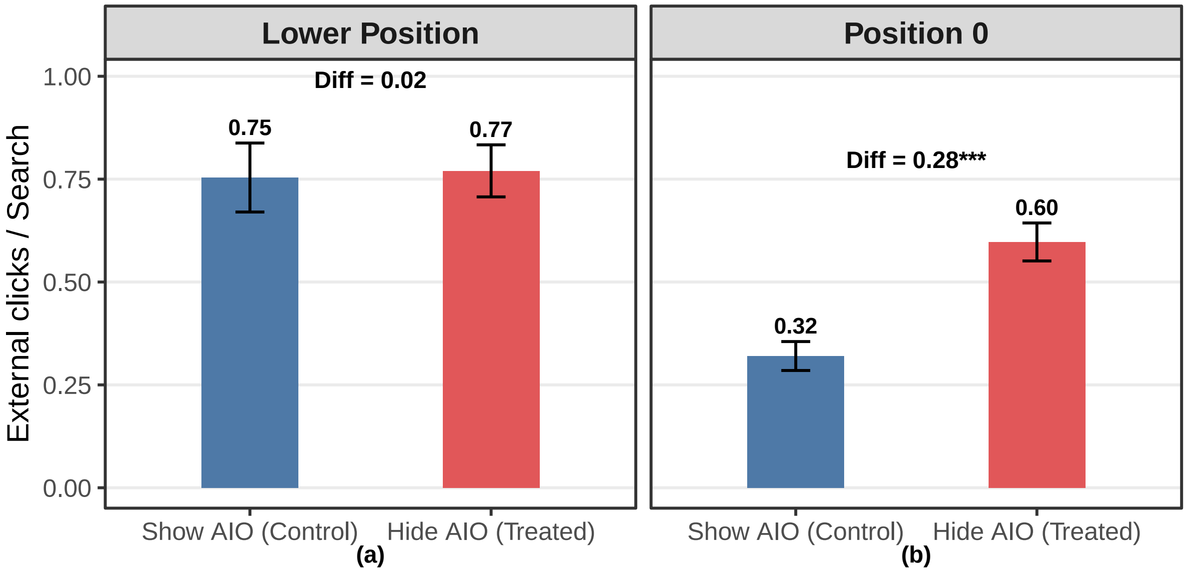
4.2 Heterogeneity and Robustness

4.2.1 Heterogeneous Treatment Effects

We first examine (pre-registered) heterogeneity in treatment effects by the position of the AI Overview (AIO) on the SERP, using the number of organic external clicks as the focal outcome of interest. This dimension of heterogeneity is particularly relevant given the importance of salience, often proxied by position on the webpage, in shaping user attention and click behavior on SERPs and other online pages (Agarwal et al., 2011; Ursu, 2018). In our setting, this consideration is especially important because AIO placement is highly skewed. As shown in Figure A.9 in the Appendix, more than 87% of queries for which an AIO was (intended to be) triggered display the overview in the top position (position 0), above all organic results. This motivates a natural partition of queries into two groups: those where the AIO appears at the top of the page and those where it appears in a lower position.

We find that the effects documented in the previous section are almost entirely driven by queries in which the AIO appears above the search results (panel (b) of Figure 5). For these queries, removing the AIO leads to large and statistically significant increases in outbound organic clicks (88% increase relative to the control). In contrast, for queries where the AIO appears in a lower position, we observe no effect on user behavior (panel (a)).[15]

Figure 5: Position of AIOs on the SERP



Notes: The figure reports treatment effects separately by the position of the AIO on the search results page. Queries are split into those where the AIO appears above the search results (position 0) (panel (b)) and those where it appears elsewhere (panel (a)). Each bar represents the corresponding search-level average number of organic external clicks for the Control (Show AIO) and HAIO (Hide AIO) groups. Error bars denote 95% confidence intervals clustered at the individual level. The difference (HAIO − Control) shown between the bars corresponds to the estimate from Equation 1. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

We next examine how treatment effects vary across the position of organic search results on the page. In our data, most queries display between 7 and 10 organic results. We estimate treatment effects separately by the position of the clicked result on the SERP. Focusing first on cases where the AIO is shown at the top (Figure A.12 in the Appendix), we find that the increase in outbound clicks is concentrated among the highest-ranked results. The largest absolute increase occurs at position 1, followed by positions 2 and 3, with progressively smaller effects at lower ranks even if statistically significant. These results provide further evidence that the effects of AIOs might be driven by their prominence.[16]

[15] It is important to note that this analysis captures heterogeneity with respect to observed AIO position, rather than the causal effect of position itself. Since AIO placement is not randomly assigned, differences across positions may reflect underlying differences in query types or user intent.

[16] We also examine heterogeneity by demographic characteristics (age, gender, and race) and baseline attitudes toward AI in Appendix Table A.1. Across all specifications, we continue to find a positive and statistically significant baseline treatment effect. While some heterogeneity emerges along age and prior comfort with AI, the core finding that AIOs substitute for external website visits holds consistently across groups.

Next, we analyze heterogeneity by query category. We utilize an LLM with detailed prompts and a sample of query links to categorize queries into three buckets: informational, transactional, and navigational (Rose and Levinson, 2004). Navigational searches are directed toward a specific website or online destination, informational searches seek knowledge or answers to a question, and transactional searches reflect an intent to complete an action such as making a purchase, booking a service, or downloading content[17]. In Figure A.13 in the Appendix, we find a large and positive effect on informational queries but statistically insignificant effects for the other two categories (though noisier due to a smaller sample size).[18]

4.2.2 Robustness: Extensive Margin and Intervention Switch

If AIOs make search more convenient by reducing the need to click on a link, they may increase the total number of searches users conduct. To examine this extensive-margin channel, we aggregate behavior to the user-day level and focus on participants who remained active for the full two-week study period. We construct an unbalanced user-day panel such that a user is included on a given day only if the extension recorded any browsing activity on that day. This approach ensures that extensive-margin outcomes are measured relative to days on which users were actively observed.[19]

Column (1) in Table 2 reports the total number of searches conducted at the user-day level in the Control and HAIO groups finding no meaningful difference. The point estimate is small in magnitude and statistically insignificant. We then analyze corresponding user-day-level totals for outbound organic clicks to find a strong pattern that mirrors the search-level results. For queries in which an AIO was (intended to be) triggered, users in the HAIO group generate significantly more external clicks over the two-week period than users in the Control group (column 2). In contrast, for queries where AIOs were not triggered, total external clicks are statistically similar between the two groups (column 3). These findings suggest that the effects of AIOs operate primarily through the intensive rather than the extensive margin.

As an additional robustness check, immediately following the main two-week experimental period, we implemented a treatment switch in which participants assigned to HAIO were

[17] Informational queries account for 71% of all searches in the sample, followed by navigational (18%) and transactional (12%) queries. The corresponding AIO trigger rates are 53%, 15%, and 6%, respectively.

[18] We also categorize queries into over 20 topics (e.g., health, finance, online communities, news). In Figure A.10, we see that the treatment had a statistically and economically significant effect on external clicks across all topic groups relative to the control, with limited differences across groups. This comes with the caveat that we lose power and hence have wider confidence bands when analyzing sub-samples of the data.

[19] We show that our results are robust to using user-level analysis in Appendix Table A.2.

assigned to the Control condition and vice versa. We then examine within-user changes in behavior before and after the switch allowing us to utilize user fixed effects. To ensure meaningful exposure to both conditions, we restrict the sample to participants who remained active through the end of the main study and for at least one week after the switch. We further restrict the analysis to searches conducted during the one-week period immediately before and after the switch for which an AIO was intended to be triggered. One caveat is that this exercise relies on a selected sample of participants who remained active throughout the extended study period and therefore does not include all originally enrolled participants.

Table 2: Extensive Margin Effects

	(1) Total Searches	(2) Total Clicks (AIO Triggered)	(3) Total Clicks (AIO Not Triggered)
Hide AIO (Treated)	-0.386	0.612***	0.161
	(0.670)	(0.157)	(0.324)
Constant	8.068***	1.275***	2.960***
	(0.487)	(0.0842)	(0.197)
Observations	8,127	8,127	8,127
R-squared	0.001	0.010	0.001

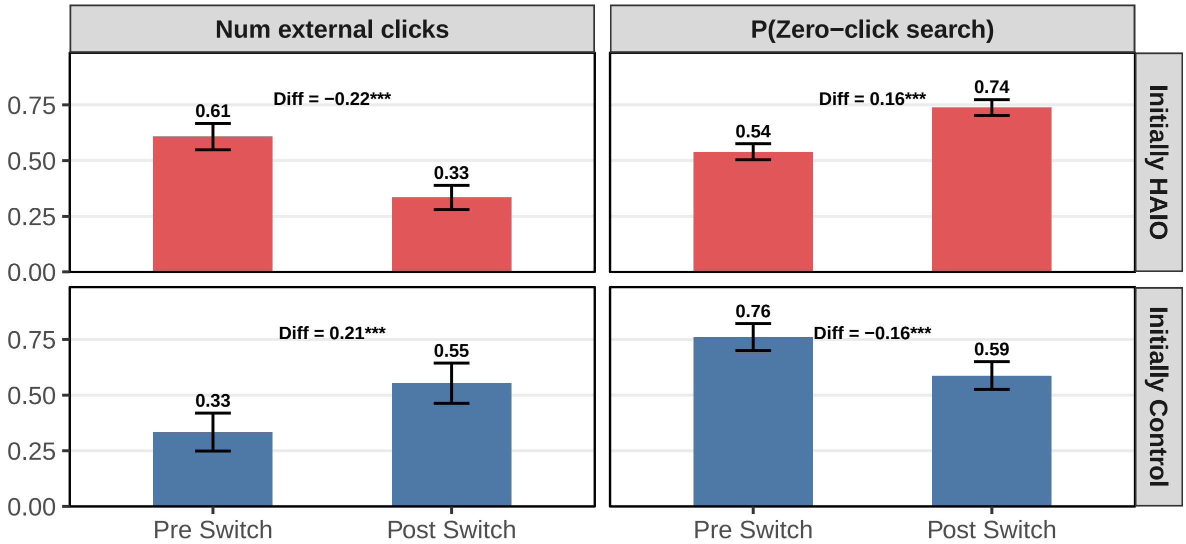
Notes: Each column reports the coefficient from a user-day regression of the indicated outcome on an indicator for assignment to the HAIO treatment. Column (1) reports effects on the total number of searches conducted while columns (2) and (3) report effects on the total number of outbound organic clicks generated from queries for which an AIO was (intended to be) triggered and not triggered, respectively. Standard errors clustered at the individual level are reported in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

As shown in Figure 6, individuals who switched from HAIO to AIO reduced their number of external clicks (from 0.61 to 0.33), while those who switched from AIO to HAIO increased theirs (from 0.33 to 0.55), consistent with the direction and magnitude of the effects found in the primary analysis. The probability of a zero-click search responded symmetrically to the reassignment, rising among those moved to AIO and falling among those moved to HAIO. These results closely mirror our main findings, strengthening the causal interpretation.

4.3 Downstream Website Engagement Analysis

The above results show an increase in traffic to downstream publishers from the HAIO group. However, a larger number of clicks does not necessarily imply greater value for publishers. It is possible that clicks originating from search pages with AIOs reflect more targeted information seeking and therefore generate higher-quality engagement with downstream websites. We therefore examine back-to-search, bounce rates, and time spent on clicked websites across the

Figure 6: Post-Experiment Treatment Switch



Notes: The top row includes queries for those who were initially in the HAIO group, while the bottom panel includes those who were initially in the Control (Show AIO) group. Only queries in the one week window before and after the switch are considered for each user. Error bars denote 95% confidence intervals clustered at the individual level. The difference shown between the bars corresponds to the estimated pre- versus post-switch effect from a regression with user fixed effects and therefore may not equal the arithmetic difference between the displayed group means. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

two groups. This analysis is restricted to clicks originating from queries for which an AIO was intended to be triggered. The results will be suggestive, since the regressions will be conditional on a click, which may introduce selection bias.

A first measure, often used in the industry as an indicator of click quality (Seobility Wiki Team, 2026), simply captures the probability that the user navigates back to the search page by clicking the back button on the same tab. A return to the search page may indicate that the search remained unresolved. As discussed in the Data Appendix, this measure can only be constructed for same-tab clicks, since back-button behavior is not observable for clicks opened in a new tab. We next turn to bounce, a commonly used measure of potentially low-quality website visits. While there is no standard industry-wide definition of a bounce, we use and adapt a metric based on Google Analytics definitions (Google, 2024). In particular, we define a bounce as a click that leads to a session lasting less than 10 seconds on the website, with no further interaction. We also run a simple regression using the overall time spent with the caveat that total time spent on any webpage may be affected by measurement error, as is well known in the industry (Aridor et al., 2024; Brands, 2022).

The results in Table 3 below provide a clear picture of click quality across groups. In particular, in column (1), we observe no statistically significant difference in the probability of returning to the search page via the back button between HAIO and control. A null result between the treatment and control groups is also seen in column (2) when we analyze bounce. Finally, when analyzing overall time spent in column (3), we see no economically or statistically significant difference between the treatment and control groups.

Overall, the absence of meaningful differences in bounce rates and time spent on downstream websites suggests that the additional clicks generated by HAIO are comparable in quality to those observed in the control group. This finding is at odds with the view that AIOs primarily eliminate low-engagement website visits (Southern, 2026a).

Table 3: Quality of Click Analysis

	(1)	(2)	(3)
VARIABLES	Back to Search	Bounce	Log(duration)
HAIO	-0.011	-0.010	-0.023
	(0.021)	(0.013)	(0.059)
Constant	0.398***	0.184***	3.61***
	(0.014)	(0.009)	(0.040)
Observations	10,937	12,781	12,781
R-squared	0.0001	0.0002	0.0001

*Notes.*The dependent variable is an indicator equal to one if the individual returns to the SERP using the back button in column (1), an indicator equal to one if the individual navigates away from the webpage in less than ten seconds without any additional engagement in column (2), and the logarithm of one plus duration (in seconds) in column (3). HAIO is equal to 1 if the click came from an individual i in the HAIO (treated) group and zero otherwise. Only clicks coming from AIO trigger intended queries were considered. Standard errors clustered at the individual level are reported in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

4.4 AI Mode Analysis

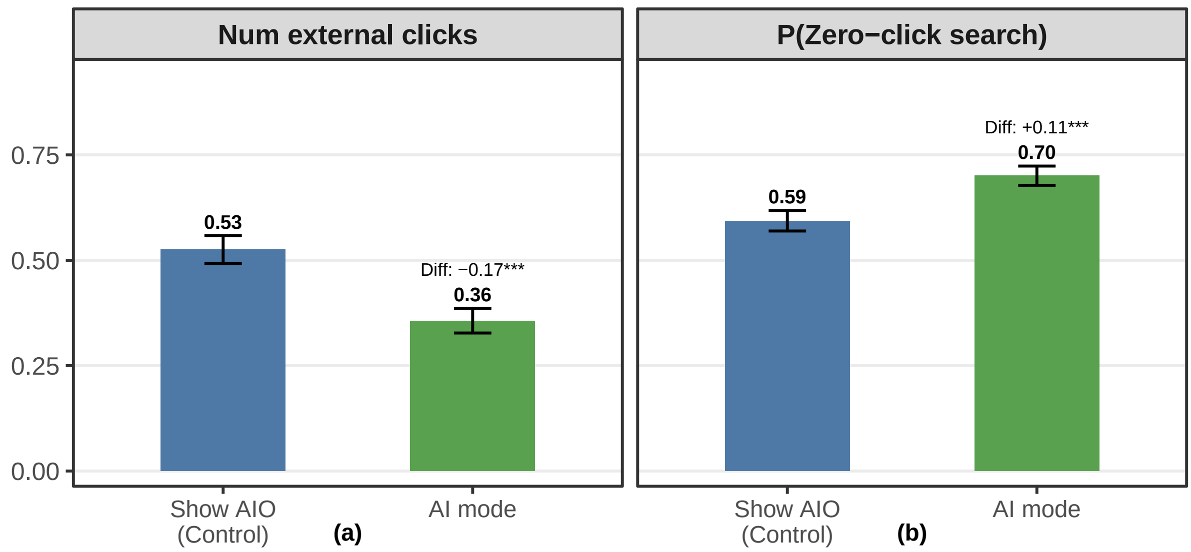
We now turn to the exploratory analysis of our third arm, in which individuals were assigned to use AI Mode and thus experienced significant attrition. Analyzing behavior in AI Mode presents measurement challenges relative to the standard search interface. In the traditional SERP setting, each search instance can be cleanly defined by a query. In contrast, AI Mode features a conversational interface that allows users to issue multiple follow-up inputs within the same interaction. These follow-ups can include genuine new queries, refinements of the original query, or non-informational inputs (e.g., acknowledgments or clarifications). As a result, defining a “search” is less straightforward in this setting.

Hence, we adopt a simplified and conservative approach. We define a search instance in AI Mode as the initial query that initiates the conversational window, and we aggregate all subsequent interactions and clicks within that session into a single observation. Thus, even if a user issues multiple follow-up prompts, all resulting clicks are attributed to the same search instance. Overall, we observed 12,507 unique search instances among AI Mode users during the experimental period. This approach may overstate clicks per search if users perform multiple distinct searches within a single session, because all clicks are attributed to the original query.

We find that overall engagement with external content is substantially lower in AI Mode relative to the Control. As shown in Figure 7, the average number of external clicks per search is markedly lower in AI Mode relative to the default with AIO (panel (a)), while the likelihood of a zero-click search is correspondingly higher (panel (b)).[20]

While these results should be seen as descriptive rather than causal, they provide suggestive evidence that as large language model interfaces become more prevalent, the overall impact on outbound traffic could be substantially larger than we estimate in the HAIO setting.

Figure 7: Treatment Effects for AI Mode Group



Notes: The figure compares engagement outcomes across the Control (Show AIO) and AI Mode groups. Outcomes are aggregated at the level of a search instance. Each bar represents the corresponding average outcome at the search level. Error bars denote 95% confidence intervals clustered at the individual level. The differences across groups are based on regressions analogous to Equation 1. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

4.5 Endline Survey: User Experience on Google Search

While our results show that AIOs reduce outbound clicks to external websites, an important part of the overall welfare question is whether they improve user experience by making information easier to access or improving satisfaction.

[20]In Appendix Figure A.14, we provide additional comparison with the HAIO group using the same definition of a search instance. In line with expectations, the difference between HAIO and AI Mode is even more accentuated relative to the Control baseline. We do not partition Control and HAIO queries by AIO trigger status in this analysis, as AI Mode does not have a direct analog to this distinction.

To examine this, we analyze responses from the endline survey, which elicited participants' perceptions of their search experience. We had a response rate of 90% in the endline survey. We restrict this analysis to participants who remained active for the full two-week study period (i.e., did not uninstall the extension or engage in bypass behavior), so that responses are more likely to reflect sustained exposure to the assigned condition. Figure 8 reports average responses across several dimensions, including overall satisfaction with Google Search (panel (a)), perceived quality of information on Google (panel (b)), and ease of finding information on Google (panel (c)). Across all measures, we find no meaningful differences between the Control and HAIO groups, with precisely estimated nulls implying limited impact on users' perceived search experience. In contrast, we observe a markedly different pattern for participants assigned to the AI Mode condition. Across all dimensions—overall satisfaction, perceived quality of information, and ease of finding information—responses are substantially lower than those of both the Control and HAIO groups. This suggests that users find the fully conversational, AI-mediated search experience less satisfactory than the traditional SERP. It is important to note that these estimates are likely upward-biased due to selection, as the AI Mode group experienced significantly higher dropout rates over the study period.[21]

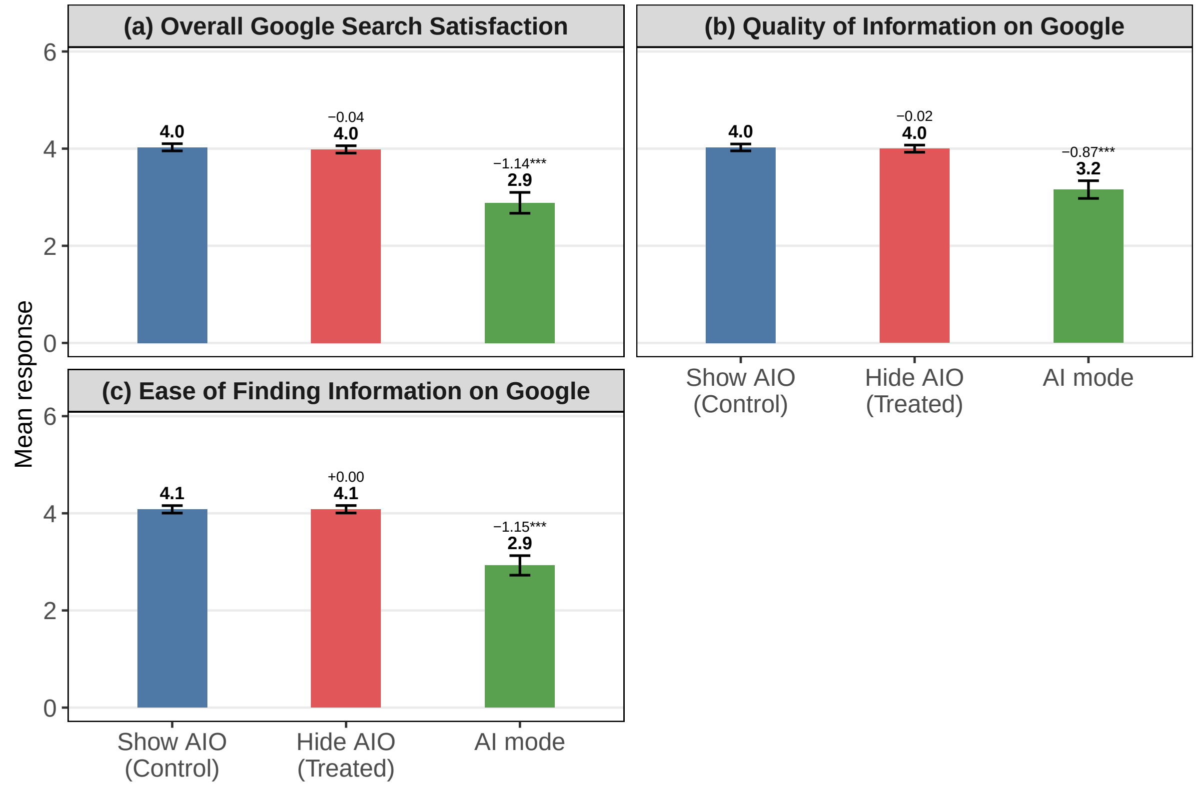
Overall, these results suggest that generative AI in search reduces referral traffic to publishers without providing commensurate benefits to users, at least on the dimensions measured above. This raises concerns about the sustainability of the content ecosystem that underpins search, as well as about platform incentives and market power.

5 Conclusion

This paper aims to provide evidence on how integrating generative AI into search interfaces reshapes user behavior and the broader digital information ecosystem. Leveraging a field experiment with users in their natural search environment, we isolate the causal impact of AI Overviews on engagement patterns, offering insights that move beyond correlational claims. Our findings demonstrate that the prevalence of AIOs substantially reduces clicks to organic search results, increasing the prevalence of zero-click searches while leaving interactions with sponsored content unchanged. These results suggest that AI-generated summaries primarily replace visits to external websites rather than redistribute attention across the search page.

[21] In Figure A.11 in the Appendix, we examine responses related to trust in AI-generated information and comfort with AI-based tools to find no meaningful differences in across the three groups.

Figure 8: Endline Survey Responses



Notes: The figure reports average responses from the endline survey across groups. Outcomes include overall satisfaction with Google Search, perceived quality of information, and ease of finding information. All measures are collected on Likert scales ranging from 1 to 5. Each bar represents the mean response for the corresponding group. The difference indicated is for each group relative to the Control AIO mean response. Error bars denote 95% confidence intervals. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Given that AIOs appear in a significant share of queries, a share that is likely to rise in the coming months and years, their aggregate impact on publisher traffic is likely to be substantial. At the same time, the absence of measurable improvements in user satisfaction or increased engagement on the downstream webpage after a click challenges the notion that these features deliver clear gains to users or downstream publishers. Instead, our evidence points to a redistribution of value within the ecosystem, in which search platforms may increasingly capture user attention – a pattern also suggested by our exploratory AI Mode results, which may foreshadow effects as search environments become more AI-mediated.

Taken together, our findings contribute to ongoing debates in platform regulation and digital market design. They highlight a fundamental tension: while AI integration can streamline access to information, it may simultaneously undermine the economic incentives that sustain high-quality content production. Ultimately, as generative AI becomes more deeply embedded in core digital infrastructures, understanding its behavioral and economic consequences will be critical for shaping a more innovative and equitable online information environment.

References

Agarwal, Ashish, Kartik Hosanagar, and Michael D Smith, "Location, location, location: An analysis of profitability of position in online advertising markets," *Journal of marketing research*, 2011, *48* (6), 1057–1073.

Agarwal, Saharsh and Saptashya Ghosh, "AI at the Fingertips: How do LLM Apps Change Smartphone Usage," 2026.

Allcott, Hunt, Juan Camilo Castillo, Matthew Gentzkow, Leon Musolff, and Tobias Salz, "Sources of Market Power in Web Search: Evidence from a Field Experiment," Working Paper 33410, National Bureau of Economic Research January 2025.

Aral, Sinan, Haiwen Li, and Rui Zuo, "The Rise of AI Search: Implications for Information Markets and Human Judgement at Scale," *arXiv preprint arXiv:2602.13415*, 2026.

Aridor, Guy, Rafael Jiménez Durán, Ro'ee Levy, and Lena Song, "Experiments on social media," 2024.

Bai, Bing, Zhiyu Zeng, Dennis Zhang, and Zhiwei Xu, "The Impact of Generative AI Search on Content-Sharing Platforms," *Available at SSRN 6171026*, 2026.

Beknazar-Yuzbashev, George, Rafael Jiménez-Durán, Jesse McCrosky, and Mateusz Stalinski, "Toxic content and user engagement on social media: Evidence from a field experiment," Technical Report, CESifo Working Paper 2025.

Bellan, Rebecca, "Google's AI Overviews are killing traffic for publishers," https://techcrunch.com/2025/06/10/googles-ai-overviews-are-killing-traffic-for-publishers/ June 2025. Accessed: 2026-02-24.

Brands, Iron, "Google Analytics' Time on Page Metric is Flawed," https://www.simpleanalytics.com/blog/you-are-getting-fooled-by-google-analytics-time-on-page-metric February 2022. Simple Analytics Blog. Last edited January 23, 2025. Accessed: May 12, 2026.

Burtch, Gordon, Dokyun Lee, and Zhichen Chen, "The consequences of generative AI for online knowledge communities," *Scientific Reports*, 2024, *14* (1), 10413.

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

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

Farronato, Chiara, Andrey Fradkin, and Chris Karr, "Webmunk: A new tool for studying online behavior and digital platforms," Technical Report, National Bureau of Economic Research 2024.

Gholami, Samira, Cristiana Firullo, Cristobal Cheyre, and Alessandro Acquisti, "Beyond search: LLM adoption and the concentration of web traffic," Technical Report, Working paper (Draft dated January 21, 2026).[Unpublished manuscript] 2026.

Goldberg, Samuel and H Tai Lam, "Generative ai in equilibrium: Evidence from a creative goods marketplace," *Equilibrium: Evidence from a Creative Goods Marketplace (February 24, 2025)*, 2025.

Google, "AI Overviews in Search," https://blog.google/products-and-platforms/products/search/generative-ai-google-search-may-2024/ May 2024. Accessed: 2026-02-24.

Kaiser, Maximilian and Christian Schulze, "Frontiers: ChatGPT Referrals to E-Commerce Websites: How Do LLMs Compare Against Traditional Channels?," *Marketing Science*, 2026.

Khosravi, Mehrzad and Hema Yoganarasimhan, "Impact of AI Search Summaries on Website Traffic: Evidence from Google AI Overviews and Wikipedia," *Available at SSRN 6164926*, 2026.

Muvija, M., "UK regulator proposes changes to Google search for publishers," https://www.reuters.com/legal/litigation/uk-regulator-proposes-changes-google-search-publishers-2026-01-28/ January 2026. Accessed: 2026-02-24.

Ng, Robin and Michael Wessel, "AI Overview or Overreach? Google's Strategic Deployment of Generative AI in Search," Technical Report, University of Bonn and University of Mannheim, Germany 2026.

Padilla, Nicolas, H Tai Lam, Anja Lambrecht, and Brett Hollenbeck, "The impact of llm adoption on online user behavior," *Available at SSRN*, 2025.

Peukert, Christian, Florian Abeillon, Jérémie Haese, Franziska Kaiser, and Alexander Staub, "Strategic behavior and ai training data," Technical Report, CESifo Working Paper 2024.

Piccardi, Tiziano, Martin Saveski, Chenyan Jia, Jeffrey Hancock, Jeanne L Tsai, and Michael S Bernstein, "Reranking partisan animosity in algorithmic social media feeds alters affective polarization," *Science*, 2025, *390* (6776), eadu5584.

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

Rose, Daniel E. and Danny Levinson, "Understanding User Goals in Web Search," in "Proceedings of the 13th International Conference on World Wide Web (WWW '04)" ACM New York, NY, USA 2004, pp. 13–19.

Seobility Wiki Team, "Pogo Sticking," 2026.

Singh, Madhavi and Fiona M. Scott Morton, "A Roadmap for a Monopolization Case Against Google: Monopsony Power and AI Overviews," Technical Report, SSRN January 2026. Forthcoming in University of Pennsylvania Law Review, Vol. 175.

Song, Yu and Puneet Manchanda, "Frontiers: Does Carrying News Increase Engagement with Non-News Content on Social Media Platforms?," *Marketing Science*, 2025, *44* (4), 748–759.

Southern, Matt G., "AI Overviews Clicks Get Tested, Earnings Tell Two Stories – SEO Pulse," https://www.searchenginejournal.com/seo-pulse-ai-overviews-clicks-get-tested-earnings-tell-two-stories/573505/ May 2026. Search Engine Journal. Accessed: May 12, 2026.

_ , "Google Pushes "Bounce Clicks" Explanation for AI Overview Traffic Loss," Search Engine Journal April 2026. Accessed: 2026-05-20.

Su, Yi, Qili Wang, Liangfei Qiu, and Runyu Chen, "Navigating the sea of reviews: Unveiling the effects of introducing AI-generated summaries in E-commerce," *Available at SSRN 4872205*, 2024.

Ursu, Raluca M, "The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions," *Marketing Science*, 2018, *37* (4), 530–552.

Vynck, Gerrit De, "In the fight for tech giants to pay for news, Google relented. Facebook didn't.," *The Washington Post*, 2021, pp. NA–NA.

Zhai, Sihan and Andrew T Ching, "AI Summary Overrepresents Fake Reviews: Evidence from Amazon," 2026.

Zhang, Peibo, Ruomeng Cui, and Dennis J Zhang, "The Impact of AI Search on the Online Content Ecosystem: Evidence from Google and Reddit," *arXiv preprint arXiv:2605.16428*, 2026.

Zhou, Pengxiang, Davide Proserpio, and Ali Goli, "LLMs as Gatekeepers: Source Concentration, Factual Quality, and Political Slant in Information Search," *Available at SSRN*, 2026.

A Appendix I

Table A.1: Heterogeneous Treatment Effects

	(1) Age	(2) Male	(3) White	(4) Trust	(5) Comfort
Treatment	0.196***	0.254***	0.278***	0.204***	0.157***
	(0.037)	(0.042)	(0.065)	(0.044)	(0.040)
Moderator	0.003	0.077*	0.049	-0.089*	-0.140***
	(0.038)	(0.037)	(0.044)	(0.036)	(0.032)
Treatment × Moderator	0.112	-0.011	-0.047	0.066	0.125*
	(0.058)	(0.056)	(0.071)	(0.056)	(0.055)
Constant	0.371***	0.338***	0.339***	0.425***	0.471***
	(0.028)	(0.027)	(0.039)	(0.024)	(0.023)
Observations	27,898	27,898	27,859	27,898	27,898

Notes: Each column reports results from a separate regression interacting the treatment indicator with a moderator. The dependent variable is the number of organic external clicks at the search level. Moderators correspond to indicators for being above the median age (37), male, white, above-median trust in AI (4+), and above-median comfort with AI tools (4+). Standard errors are clustered at the individual level. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table A.2: Extensive Margin Effects: User Level

	(1) Total Searches	(2) Total Clicks (AIO Triggered)	(3) Total Clicks (AIO Not Triggered)
Hide AIO (Treated)	-6.09	6.93***	0.95
	(8.43)	(1.93)	(3.92)
Constant	95.74***	15.16***	35.28***
	(5.89)	(1.35)	(2.74)
Observations	713	713	713

Notes: Each column reports the coefficient from a user regression of the indicated outcome on an indicator for assignment to the HAIO treatment. Only users who were part of the endline survey sample (those who did not uninstall) are included in this analysis. Column (1) reports effects on the total number of searches conducted while columns (2) and (3) report effects on the total number of outbound organic clicks generated from queries for which an AIO was (intended to be) triggered and not triggered, respectively. Standard errors clustered at the individual level reported in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Figure A.1: Participant Recruitment and Study Flow

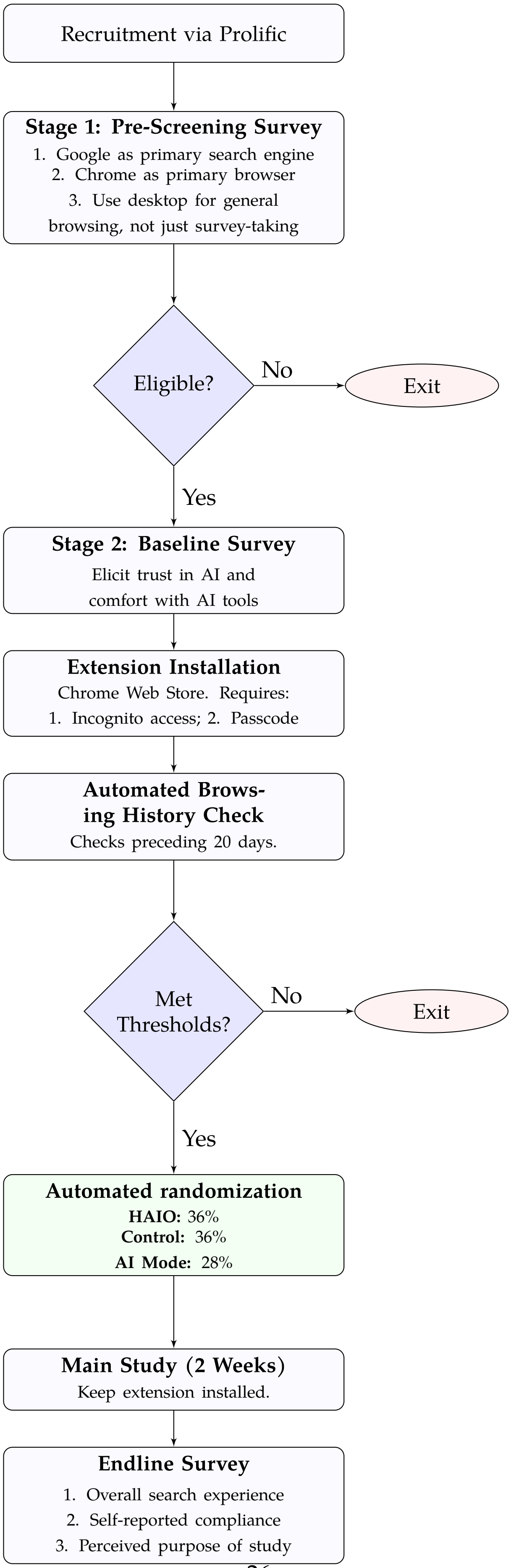
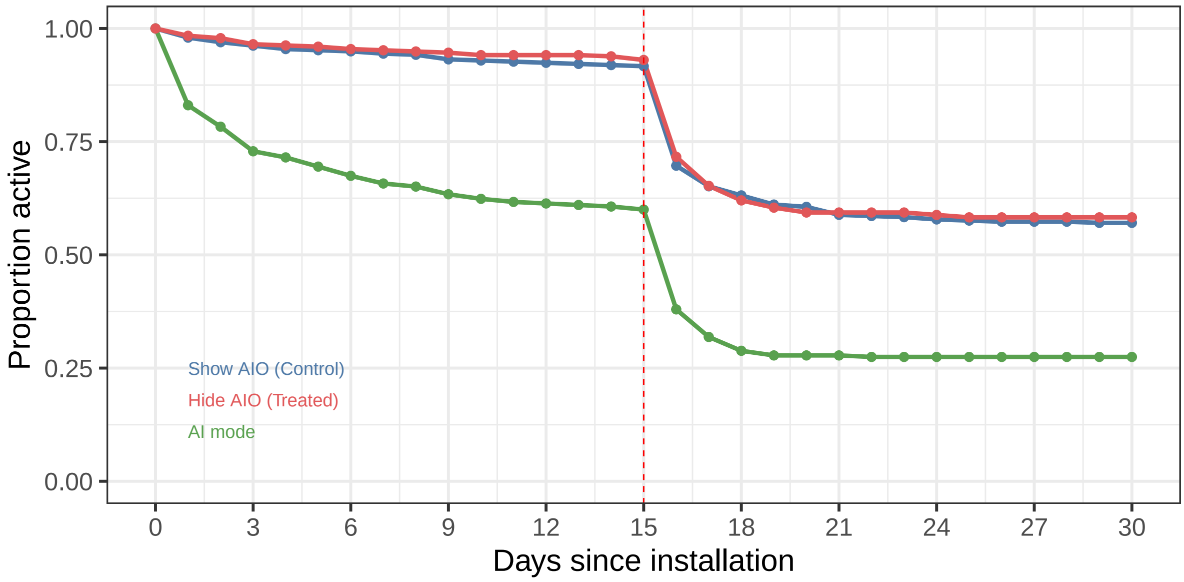
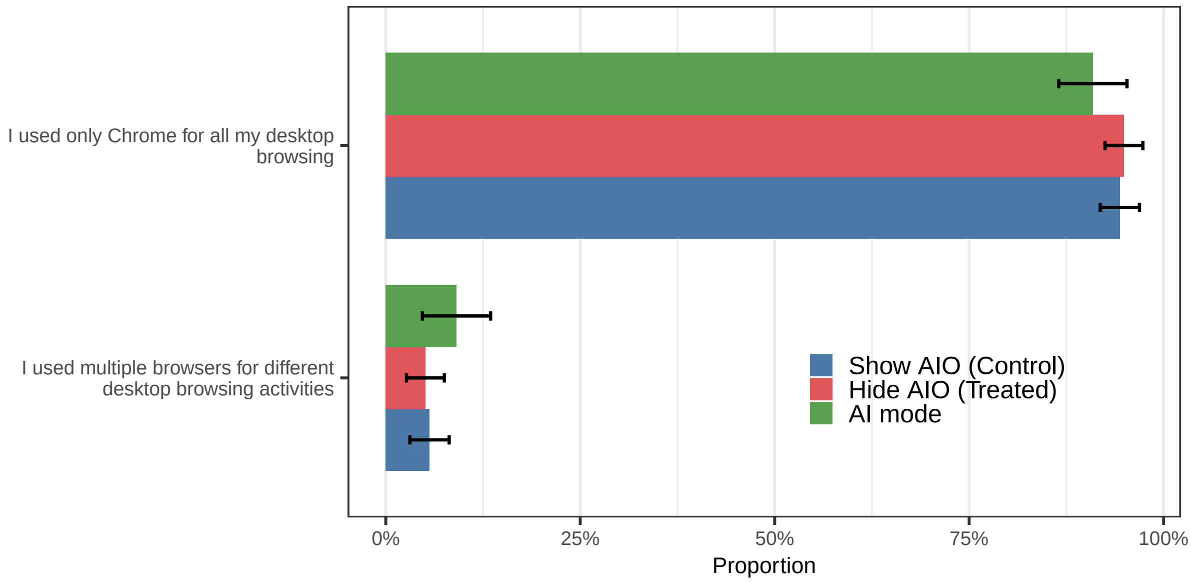


Figure A.2: Participant Survival Over Time by Treatment Group



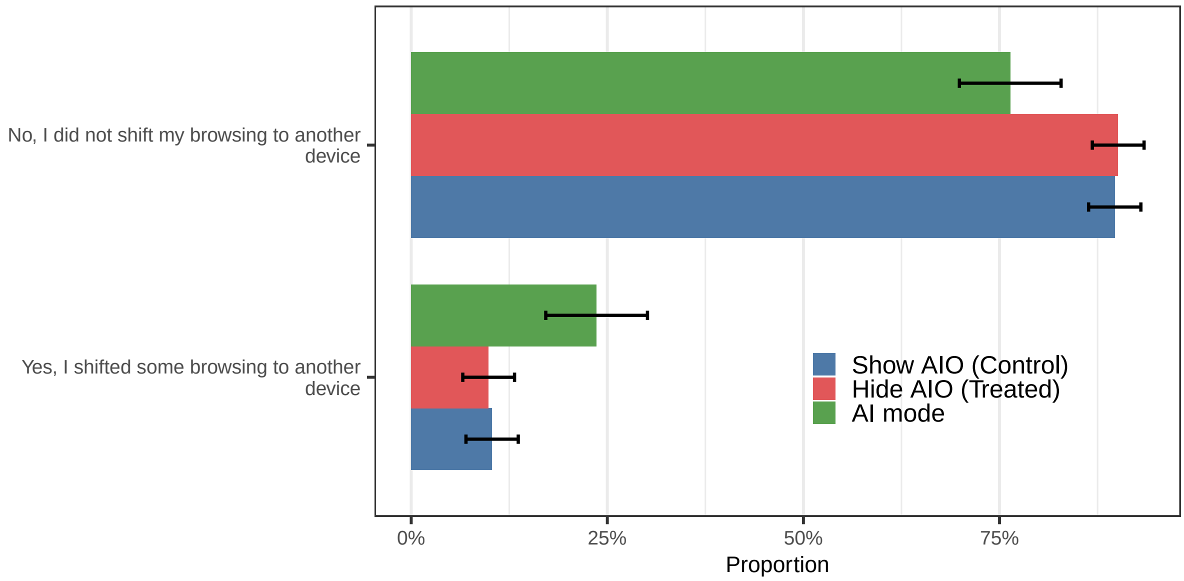
Notes: The figure shows study participant survival across the three groups (AIO, HAIO, AI Mode) during the two weeks of the main experimental period and then in the two weeks after where we switch their treatment status.

Figure A.3: Endline: Exclusive Use of Chrome Browser



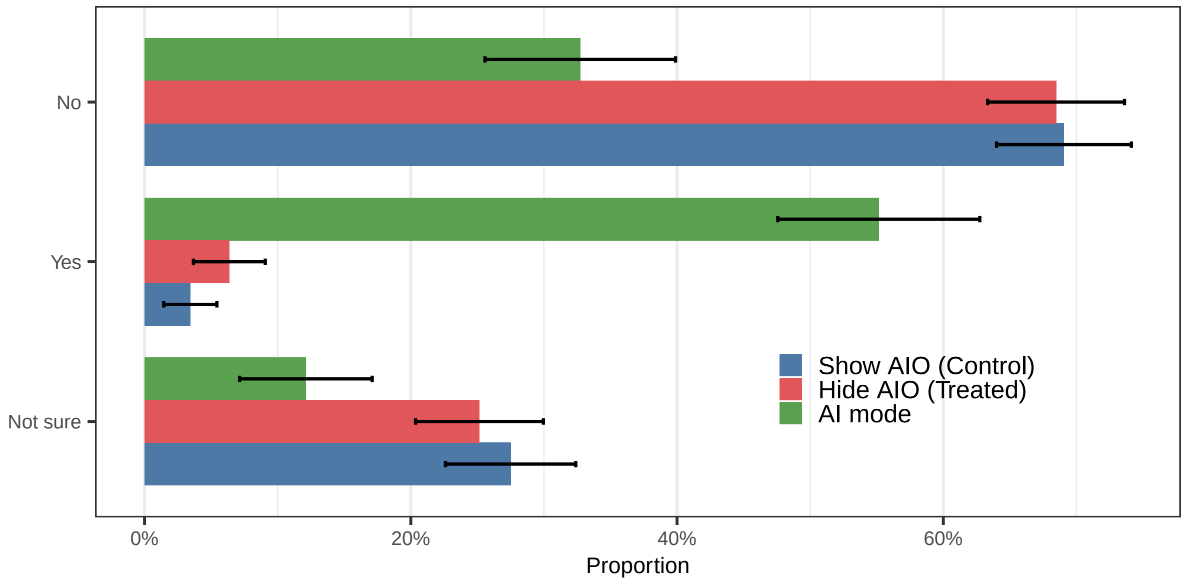
Notes: The figure shows endline survey response of participants across the three groups (AIO, HAIO, AI Mode) for the question related to whether they used the Chrome browser exclusively during the experimental period.

Figure A.4: Endline: Shift to Another Device



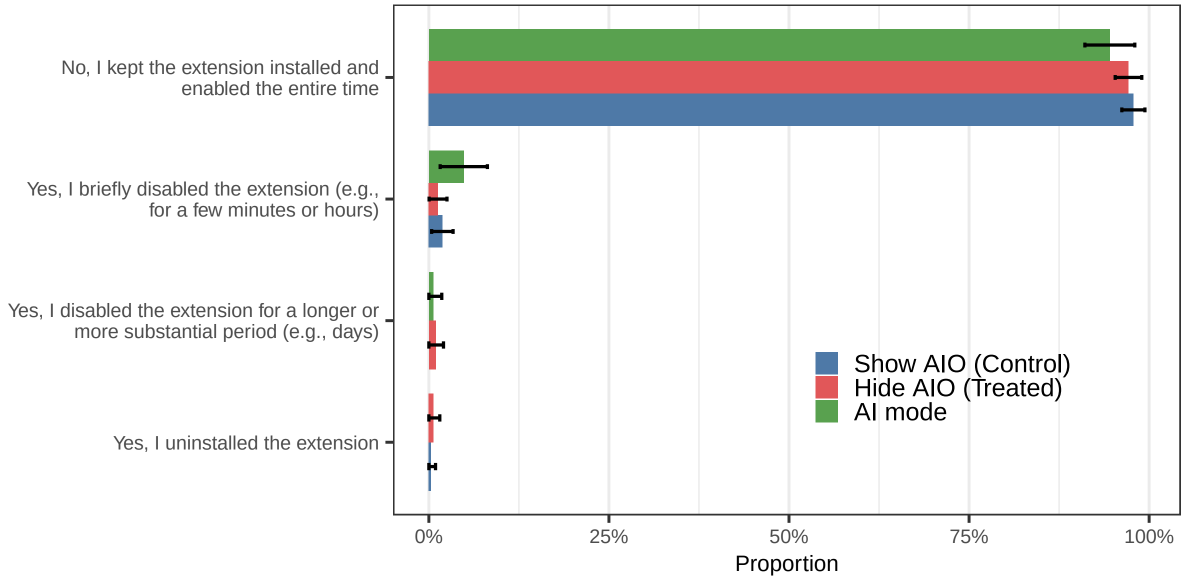
Notes: The figure shows endline survey response of participants across the three groups (AIO, HAIO, AI Mode) for the question related to whether they shifted their online activities to another device.

Figure A.5: Endline: Noticed Changes



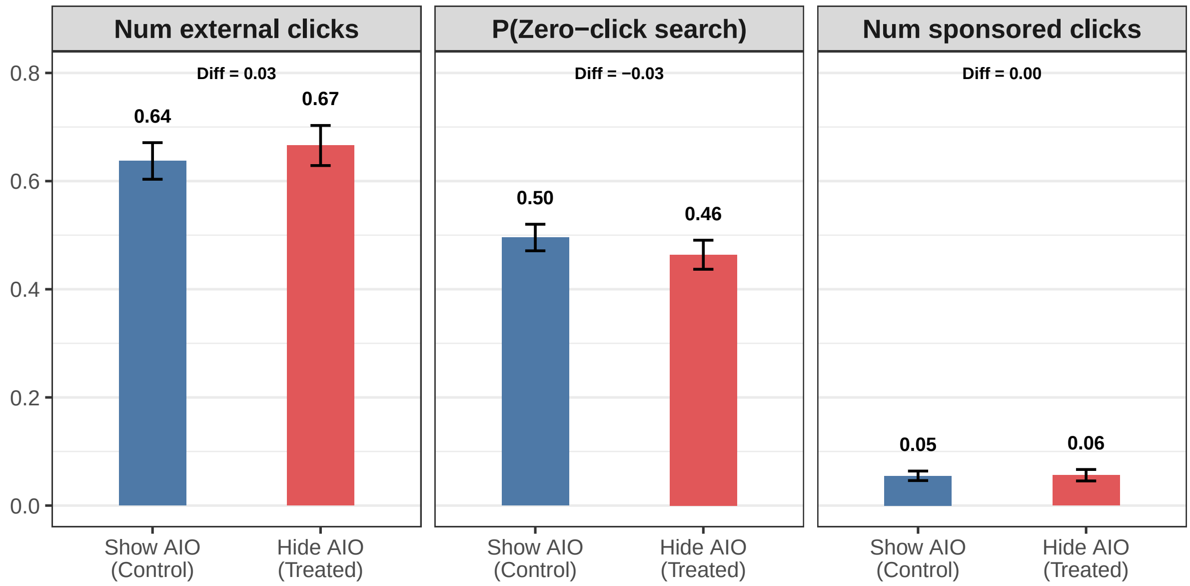
Notes: The figure shows endline survey response of participants across the three groups (AIO, HAIO, AI Mode) for the question related to whether they noticed any changes in their search experience.

Figure A.6: Endline: Self-reported Compliance



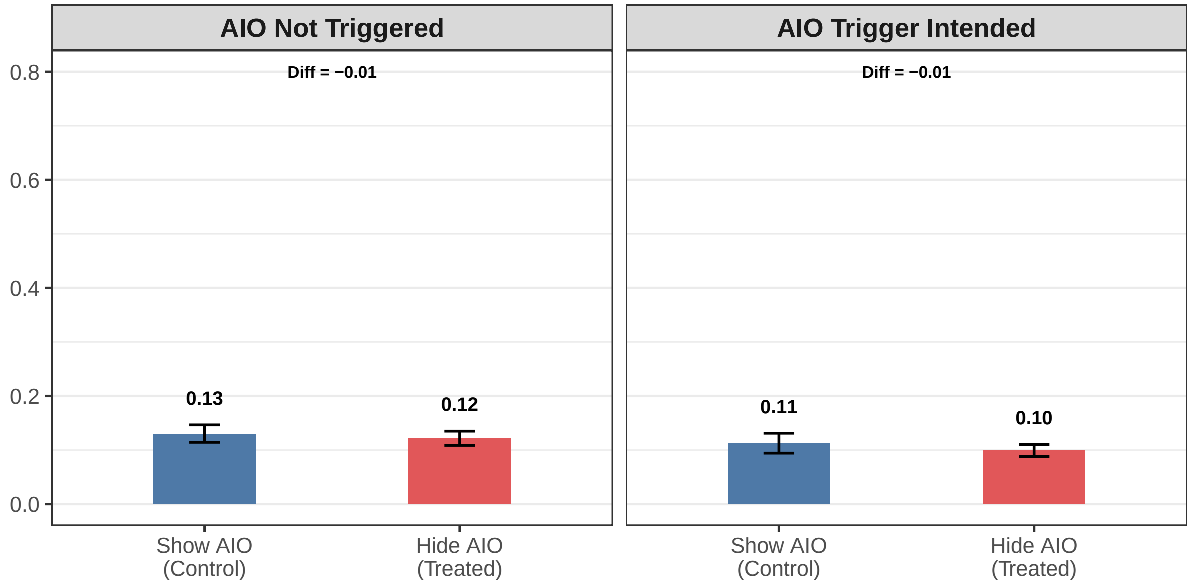
Notes: The figure shows endline survey responses of participants across the three groups (AIO, HAIO, AI Mode) for the question related to whether they complied with the experimental guidelines about having the extension installed and enabled through the two-week period.

Figure A.7: Treatment Effect on Organic Clicks for Non-Triggered Queries



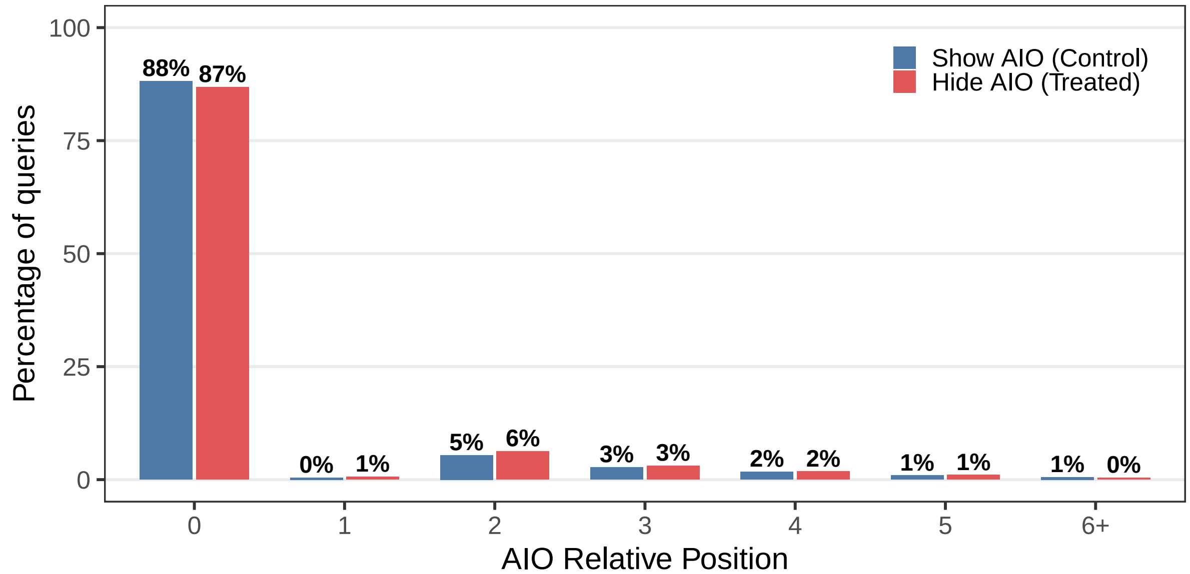
Notes: The figure reports treatment effects on external organic clicks for queries that didn't trigger AIOs. Each bar represents the number of external clicks per search for the HAIO (Hide AIO) relative to the control (AIO) group. Error bars denote 95% confidence intervals clustered at the individual level. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Figure A.8: Treatment Effect on Internal Clicks



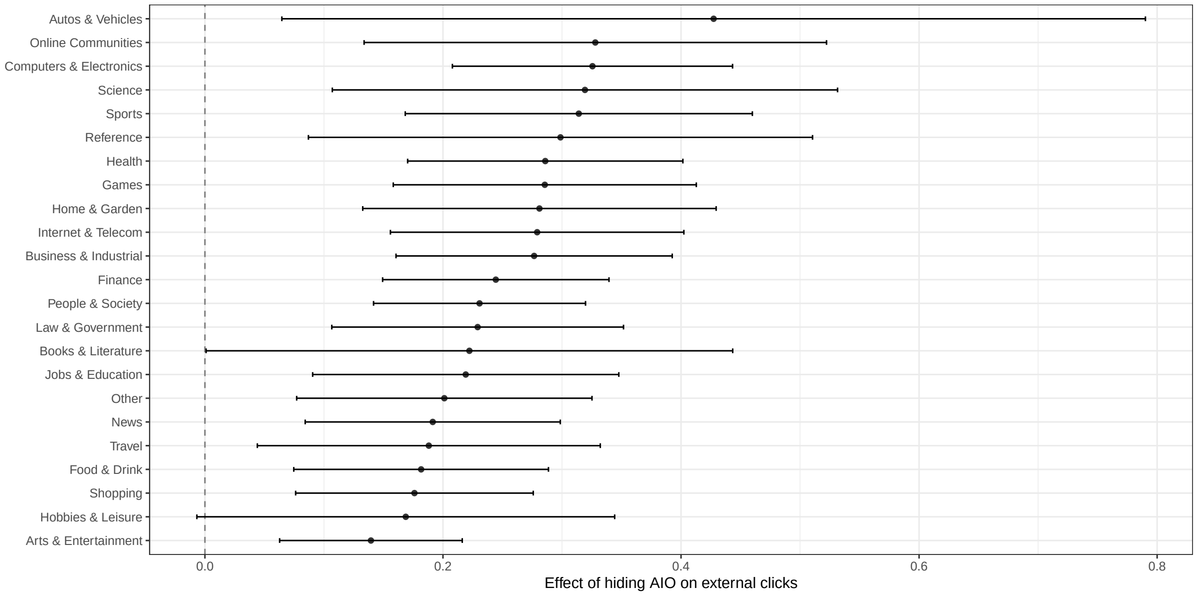
Notes: The figure reports treatment effects on internal navigation clicks. Each bar represents the number of internal clicks per search for the HAIO (Hide AIO) relative to the control (AIO) group. Error bars denote 95% confidence intervals clustered at the individual level. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Figure A.9: AIO Position Distribution



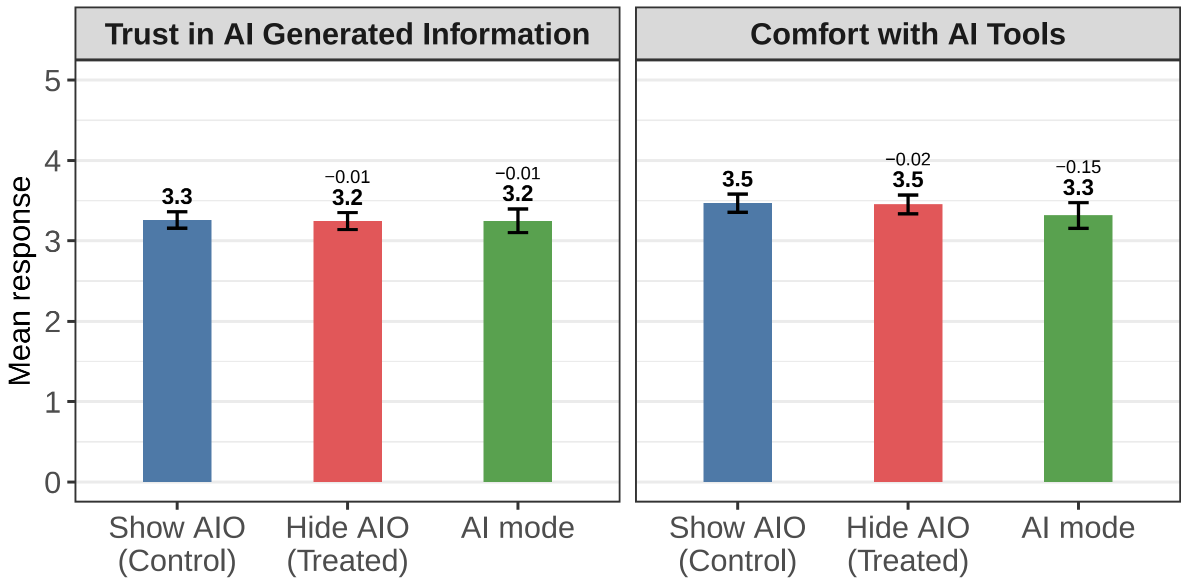
Notes: The figure reports the distribution of AI Overview (AIO) positions on the search engine results page (SERP) across queries for which an AIO was (intended to be) triggered. Relative position 0 denotes placement in the space above the first organic search result. Relative position 1 denotes placement at the first organic-result slot, while relative position 2 denotes placement after the first organic result and before the second. More generally, for relative positions greater than 0, relative position x denotes placement after the (x-1)-th organic result and before the x-th organic result.

Figure A.10: Treatment Effect on Organic Clicks by Query Topic



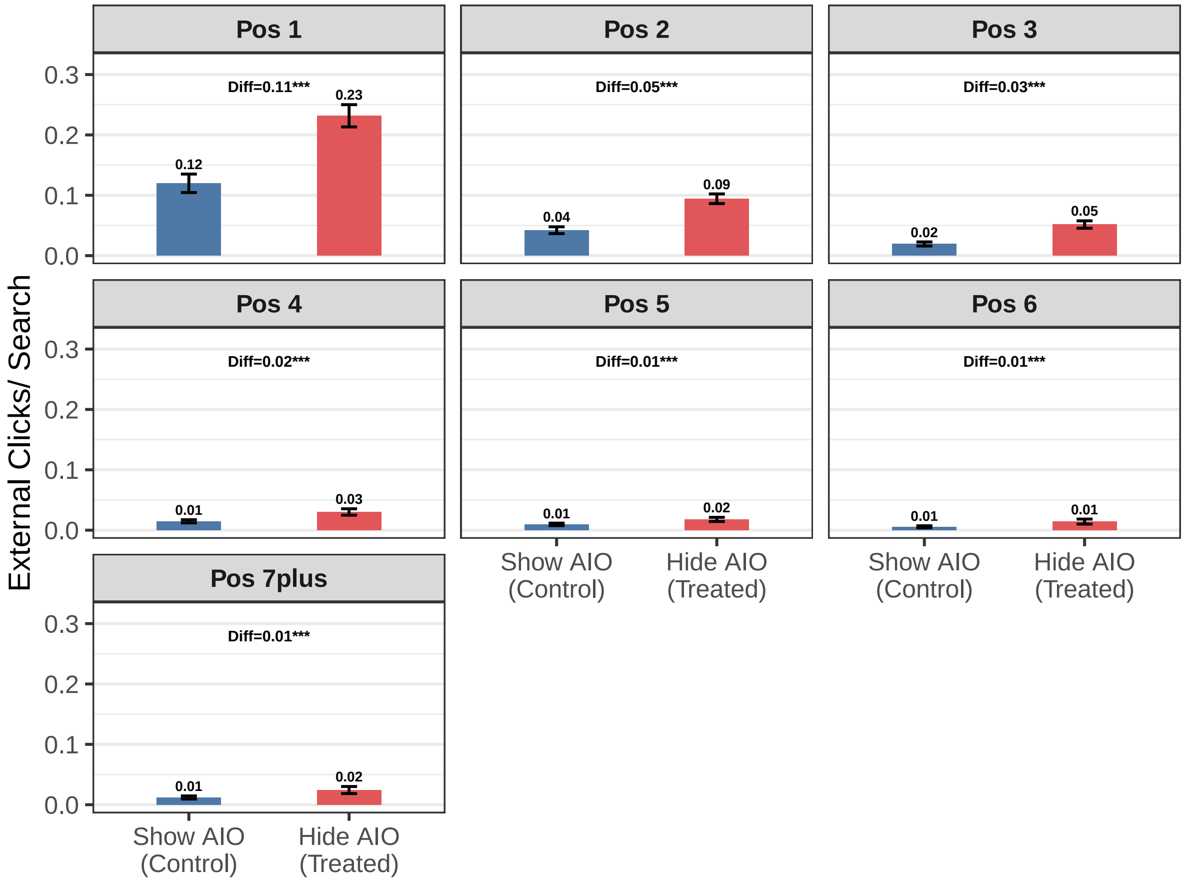
Notes: The figure reports treatment effects on external organic clicks by the topic of the query. Each bar represents the number of external clicks per search for the HAIO (Hide AIO) relative to the control (AIO) group. Error bars denote 95% confidence intervals clustered at the individual level. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Figure A.11: Additional Endline Survey Measures



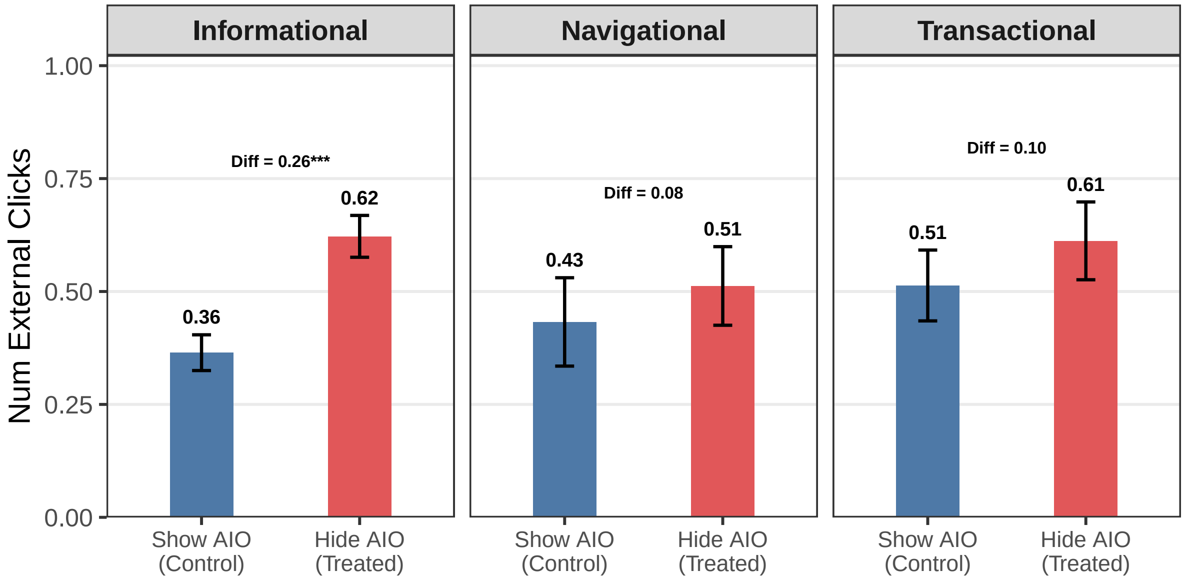
Notes: The figure reports average responses from the endline survey across treatment groups. All measures are collected on Likert scales ranging from 1 to 5. Each bar represents the mean response for the corresponding group. The difference indicated is for each group relative to the Control AIO mean response. Error bars denote 95% confidence intervals. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Figure A.12: Organic Clicks by Search Position: AIO Shown at Top Position



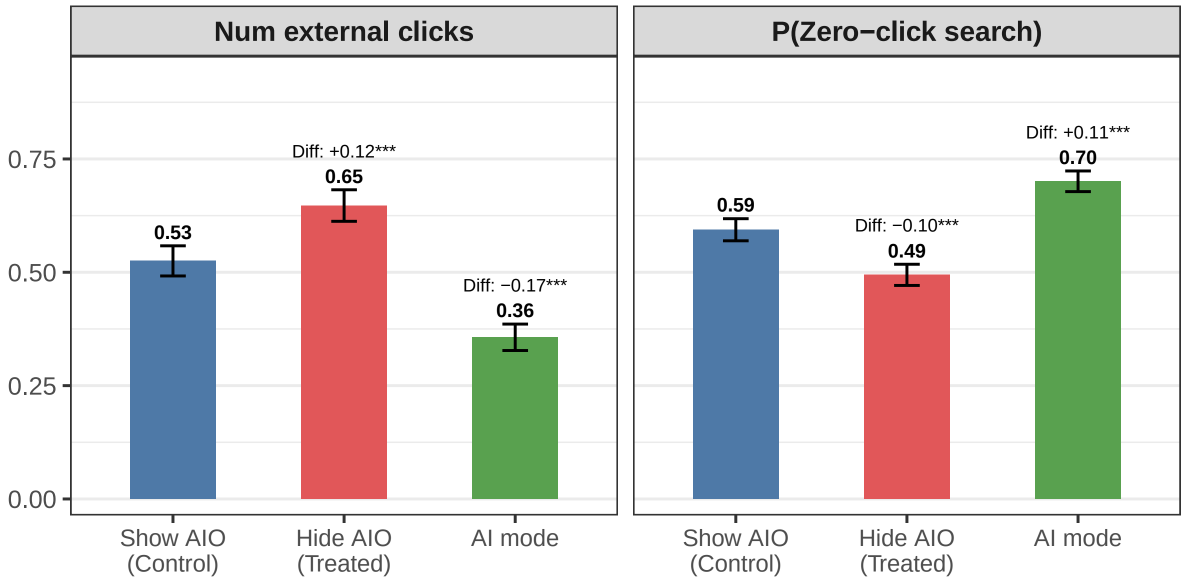
Notes: The figure reports treatment effects on external organic clicks by the position of the clicked result on the search engine results page (SERP). Panels correspond to result positions (1 through 6 and 7+). The sample is restricted to queries for which an AIO was (intended to be) triggered at the top position. Each bar represents the average number of external clicks per search for the Control (Show AIO) and HAIO (Hide AIO) groups at the corresponding position. Error bars denote 95% confidence intervals clustered at the individual level. The reported differences (HAIO − Control) correspond to estimates from Equation 1. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Figure A.13: Treatment Effect on Organic Clicks by Query Category



Notes: The figure reports treatment effects on external organic clicks by the category of the query. Each bar represents the number of external clicks per search for the HAIO (Hide AIO) relative to the control (AIO) group. Error bars denote 95% confidence intervals clustered at the individual level. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Figure A.14: Treatment Effects for AI Mode Group



Notes: The figure compares engagement outcomes across the Control (Show AIO), HAIO, and AI Mode groups. Outcomes are aggregated at the level of a search instance. In the AI Mode condition, a search instance is defined as the initial query that initiates a conversational session, with all subsequent interactions and clicks within that session aggregated into a single observation. For comparability, queries in the Control and HAIO groups are not partitioned by AIO trigger status. Each bar represents the corresponding average outcome at the search level. Error bars denote 95% confidence intervals clustered at the individual level. The differences across groups are based on regressions analogous to Equation 1. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

B Appendix II: Data Collection and Cleaning

B.1 The Extension

Behavioral data were collected using a custom Chrome browser extension built on Webmunk (Farronato et al., 2024). The extension operates through a series of JavaScript-based event 'listeners' embedded within the browser, which asynchronously monitor user interactions and webpage state changes in real time. These listeners capture page navigations, URL visits, tab and window activity, clicks on webpage elements, scrolling behavior and scroll depth, page load events, and interactions with search engine result pages (SERPs). AI Mode access was disabled for the control and HAIO groups to preserve a common search environment and ensure that treatment effects can be attributed to differences in AI Overview exposure rather than the availability of an alternative AI-powered search interface.

For Google Search pages specifically, the extension includes a dedicated Google Search results listener that is triggered whenever a search results page loads or is updated. This listener records the search query, search result URLs, the presence and position of AI Overviews (AIOs), sponsored results, related-question modules, and other SERP elements. The listener also captures the relative positions of these elements on the page, enabling the reconstruction of the search environment presented to the user.

B.2 Participant Recruitment and User Exclusions

Figure B.1 summarizes participant recruitment and sample construction. A total of 3,710 individuals initiated the screening survey on Prolific. Of these, 1,574 were screened out because they did not meet the pre-specified eligibility criteria, including using Google as their primary search engine, Chrome as their primary browser, and reporting regular desktop browsing activity. An additional 447 participants either did not complete the survey or did not successfully install the browser extension. A further 581 participants failed the browser-history validation checks, which required minimum levels of recent Google usage, browsing activity, and domain diversity. After a small number of data-integrity-related exclusions (43 participants, as described below), the final experimental sample consisted of 1,065 participants.

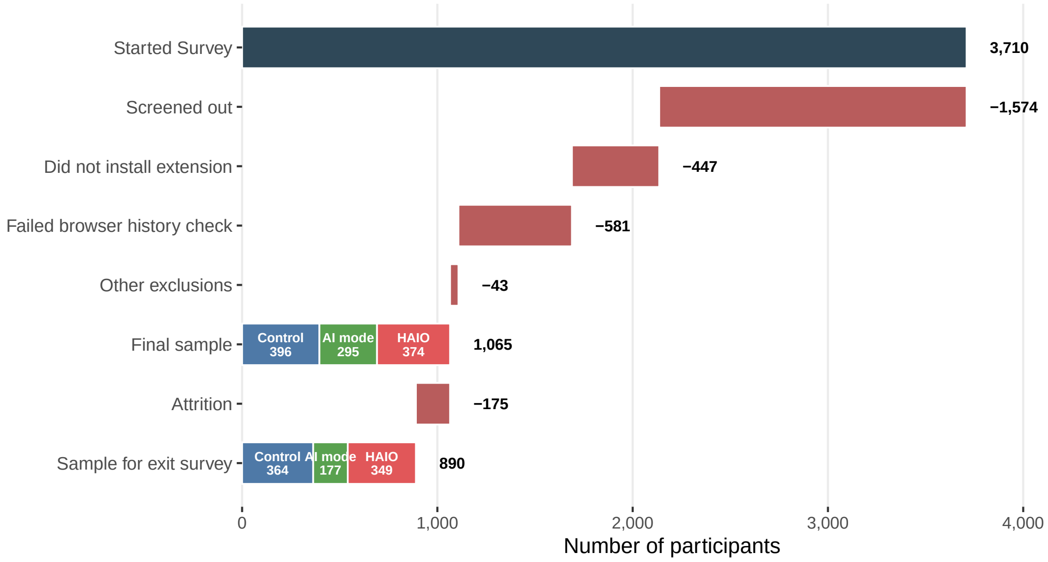
- **17 users (pilot participants or duplicates):** These users either participated in a pilot version of the study earlier, or completed the survey and installed the extension twice. They were inadvertently included in the main sample when rejected Prolific submissions reopened study slots that were subsequently filled by individuals who had already participated in an earlier wave of the study or a pilot.
- **7 users (conflicting extension users):** These users exhibited unusually high frequencies of searches containing the `udm=14` parameter during the pre-treatment period, indicating prior use of browser extensions or tools that modified Google Search results.
- **19 users (Google RaterHub workers):** These users were identified as participants in Google's RaterHub program[22] based on their URL-visits data. Because their search activity is influenced by professional search-evaluation tasks, their behavior may not be representative of typical consumer search behavior.

Participants were then randomly assigned to one of three experimental conditions: Control (n = 396), AI Mode (n = 295), and Hide AI Overview (HAIO) (n = 374). During the subsequent two-week observation period, 175 participants attrited by uninstalling the extension or engaging in bypass behavior. The resulting endline survey sample comprised 890 participants: 364 in the Control group, 177 in the AI Mode group, and 349 in the HAIO group. Two

[22] https://www.linkedin.com/pulse/what-google-raterhub-david-bassey-8e0oc/

participants uninstalled the extension at approximately 11:30 PM on the final day of the study period; because they completed virtually the entire observation window, they are not counted in attrition and are included in the exit survey sample.

Figure B.1: Sample Funnel



B.3 Construction of the Search-Level Dataset

The Google Search listener was triggered whenever a Google Search results page (SERP) was loaded. Because the same search could generate multiple listener events—for example, through page refreshes, back-button navigation, reopening a tab, or incremental page updates—the raw telemetry data do not correspond one-to-one with underlying searches. Constructing the search-level dataset therefore required four steps.

Restricting to standard web searches. We first restricted the sample to standard Google web search pages and excluded specialized verticals such as Images, Videos, Shopping, Books, and News. Search queries were extracted from URL parameters and standardized through URL decoding and text normalization.

Identifying unique searches. The same search could generate multiple listener events within a short period of time. To identify unique searches, observations were grouped by user, browser window, browser tab, query text, and AI Overview status. Observations occurring within a short interval were treated as belonging to the same search instance, while observations separated by longer intervals were treated as distinct searches.[23]

This distinction is important because Google search results are dynamic. Repeated executions of the same query may differ in the ordering of search results, the presence and position of AI Overviews, sponsored content, and other SERP elements. Consequently, observations separated by longer intervals are more likely to reflect genuinely distinct searches than repeated exposure to the same results page.

[23]The cutoff corresponded to the 95th percentile of within-query inter-arrival times (433 seconds).

Collapsing repeated observations. When multiple listener events were determined to belong to the same search instance, they were collapsed into a single search-level record. Search-result and sponsored-result lists were merged across observations to account for cases in which page elements appeared incrementally during loading. The resulting dataset contains one observation per unique search.

Removing ambiguous observations. A small number of searches could not be reliably classified and were excluded from the analysis.

First, approximately 500 searches returned no observable SERP content. Many of these cases were associated with malformed or incomplete queries, while others likely reflect routine telemetry failures in which the listener was triggered but failed to capture the page contents. Because these observations contained neither search results nor click opportunities, and because AI Overview status could not be determined, they were excluded.

Second, 1,518 searches exhibited ambiguous AI Overview status. The extension was designed to classify an AI Overview as present only when it appeared as a standalone module within the main search results page. In these cases, however, content identified as an AI Overview appeared within the "People Also Ask" module, likely reflecting changes in Google's underlying HTML structure during the study period. Because it was not possible to determine whether a conventional AI Overview was simultaneously present elsewhere on the page, these searches were excluded.

After de-duplication and data-quality exclusions, the final dataset contains 68,089 unique searches generated by 1,065 users during the two-week study period.

B.4 Construction and Cleaning of Click Data

Click activity was recorded using two separate event listeners: one for standard left-clicks and one for right-clicks. Both listeners operated similarly, recording the user identifier, browser window and tab identifiers, timestamp, timezone, the URL of the page on which the click occurred, and the HTML content of the clicked element. Constructing the click-level dataset required the following steps.

Restricting to search-related clicks. We first restricted the sample to click events originating from standard Google web-search pages and excluded clicks recorded on specialized search verticals such as Images, Videos, Shopping, News, and Books. We then extracted the destination URL from the clicked element and recovered the associated search query from the SERP URL recorded at the time of the click. Query strings were standardized through URL decoding and text normalization.

Classifying click destinations. Each click was classified into one of three categories used throughout the analysis.

1. **Sponsored clicks.** Clicks with destination URLs containing the path `/aclk` were classified as sponsored clicks. This path corresponds to Google's advertising redirect infrastructure and is used to route traffic from sponsored search results.

2. **External clicks.** Clicks leading to non-Google domains were classified as external clicks. These clicks represent traffic directed to downstream websites and constitute the primary outcome in the search-level analyses.

3. **Internal clicks.** All remaining clicks were classified as internal clicks. This category includes pagination, redirects to a new Google search, transitions to Google verticals (e.g., Images, News, Shopping, Maps, and AI Mode), interactions with knowledge panels and other Google modules, and any other click that remains within Google's ecosystem.

This classification distinguishes between the key behavioral margins examined in the paper: traffic sent to external publishers, engagement with sponsored content, and continued interaction within Google's own search environment.

B.5 Merging Search and Click Data

A common challenge in browser telemetry studies is that search and click events are generated by independent listeners and therefore lack a common identifier that would allow them to be linked directly. As a result, clicks cannot be deterministically assigned to a specific search instance. Instead, we reconstructed search–click relationships using information observed by both listeners, including the user identifier, browser window identifier, browser tab identifier, query text, and timestamps, following the steps below.

Matching clicks to searches. For each click event, we retrieved the Google query associated with the search results page where the click occurred. Clicks were then matched to searches using the user identifier, browser window identifier, browser tab identifier, and query text.

When only one search matched a given combination of identifiers, all associated clicks were assigned to that search. When multiple searches shared the same identifiers and query, clicks were assigned to the most recent search occurring before the click timestamp. This procedure reflects the natural assumption that a click originates from the search results page most recently displayed to the user within that browser tab.

Removing implausible matches. After matching, we calculated the elapsed time between each search and its associated clicks. A small number of observations exhibited negative time gaps, likely reflecting minor timing discrepancies caused by asynchronous listener execution. Clicks occurring up to two seconds before the recorded search event were retained and assigned a zero-second gap.

Clicks occurring more than one hour after the matched search were excluded. Although some of these clicks may reflect users returning to a previously viewed search results page, long delays substantially increase uncertainty regarding the correct search–click linkage because users may have navigated across multiple pages, tabs, or search sessions in the interim. The one-hour threshold is conservative relative to observed browsing behavior: the 99th percentile of the search–click time-gap distribution is approximately 17 minutes.

B.4 Determining Overview Position

To accurately map the vertical location of the AIOs, the extension dynamically calculates a relative position metric for the AIO by evaluating its hierarchical placement within the Document Object Model (DOM). If the AIO appears above the primary search results container, it is assigned a relative position of 0. If the AIO appears in the main search results section, its position is determined relative to other SERP elements, including organic search results and modules such as image carousels, video results, and People Also Ask boxes. For example, a relative position of 1 indicates that the AIO appears as the first result in the ranked search results, whereas a relative position of 0 indicates that it appears above the search results entirely. A relative position of 2 indicates that the AIO appears after the first result and before the second. This approach allows us to consistently determine whether the AIO appears above the search results or is integrated into the ranked results.

B.5 Classifying Query into Category and Topic

User queries were systematically classified using an automated pipeline with the help of the "GPT-5-mini" model via OpenAI's API. Recognizing that raw search queries are frequently short or ambiguous, the model was provided with an enriched data schema that included the user's explicit query text alongside contextual signals extracted from the immediate search environment. Specifically, the model evaluated the textual titles of the organic search results and binary indicators based on the presence of SERP elements such as "Products" (shopping) section and "Top Stories" section to better infer the context of the search. For example, a query such as "tariffs" is inherently ambiguous. The presence of a "Top Stories" module containing recent headlines can help distinguish a news-related search from a more general informational query.

Using this enriched context, the model classified each search event along two independent dimensions: search intent and topic. To categorize queries by search intent, we adopted the taxonomy and category descriptions proposed by Rose and Levinson (2004) to broadly classify based on primary user goals: Navigational, Informational and Transactional. For topic classification, the model mapped the query against the 25 specific topic categories utilized by Google Trends. The model was instructed to assign up to five relevant topics per search event from this predefined list.

B.6 Construction of Click Quality Measures

The click-quality analysis is restricted to external clicks from Google Search to non-Google websites. To construct these measures, we use a browser navigation listener that records all major browsing events, including URL visits, page refreshes, back and forward navigation, new tab creation, tab switches, and tab closures. These events are timestamped and linked to browser tabs and windows, allowing us to reconstruct the sequence of actions that occur after a user clicks an external search result. The objective is to identify the start and end of each downstream browsing session and then measure active engagement within that session.

Identifying click-session starts. We begin with the set of external clicks observed on Google Search. For clicks that open in the same browser tab, we identify the first subsequent `navigation_committed` event occurring in the same user, window, and tab whose URL matches the clicked destination. To accommodate redirects and URL normalization, we additionally allow matches based on the destination domain when an exact URL match is unavailable. Matches must occur within 10 seconds of the recorded click. For same-tab clicks (about 90% of all clicks), engagement duration begins at this matched `navigation_committed` event, which marks the user's arrival at the destination page.

New-tab clicks require a separate procedure. When a user opens a search result in a new tab, the navigation listener records a `navigation_created_new` event that contains both the source tab and the newly created destination tab. We first use this event to map the originating Google Search tab to the destination tab and then identify the first `navigation_committed` event in that destination tab that matches the clicked URL or domain. As before, when an exact URL match is not found, we match on the domain name. Importantly, users frequently open links in background tabs without immediately viewing them. We therefore define the start of engagement as the first `tab_switch` event into the destination tab rather than the `navigation_committed` or tab-creation event itself. New tabs that load successfully but are never activated are retained as valid clicks, but are assigned a zero engagement duration. Using this procedure, we successfully identify session starts for 37,338 of the 39,724 external clicks observed in the Control and HAIO groups (94.0%). For the remaining clicks, we are unable to link the recorded click to a corresponding navigation event. This can occur because of missing navigation telemetry, redirects or URL transformations that prevent reliable matching between the click and subsequent navigation records, or situations in which the destination page fails to load successfully. These unmatched clicks are excluded from all click-quality analyses.

Identifying click-session endings. After identifying the start of each click session, we determine when engagement with the clicked website ends. A session terminates at the first subsequent event that either closes the tab/window or navigates away from the clicked website. To distinguish genuine exits from continued browsing within the same website, we classify a navigation as internal whenever consecutive `navigation_committed` events occur on the same domain within a tab. Internal navigations, therefore, remain part of the same session and do not generate a session end. Importantly, switching to another tab or interacting with content in other tabs does not end the session. The session remains active until the focal tab itself is closed or navigates to a different domain.

Measuring active engagement. For each reconstructed session, we collect all navigation events that occur between the session start and end. Active engagement time is calculated only for periods when the clicked tab is the active tab. Durations are computed from the time elapsed between consecutive events and summed across the session. This approach excludes time during which the clicked webpage remains open in a background tab while the user interacts elsewhere. To reduce the influence of long periods of inactivity and telemetry gaps, we cap the elapsed time between any two consecutive events at four minutes before summing durations across the session.[24]

Final measures. Using the reconstructed sessions, we construct measures of click quality analogous to those commonly used in web analytics. A click is classified as a bounce if the user spends less than ten seconds actively engaged with the destination website without any additional navigation within the destination domain.

For same-tab clicks, we construct a measure of *Back to Google Search*. This occurs when the session ends through a browser back-navigation event and the destination of that navigation is the original Google Search results page. We restrict this measure to same-tab clicks (90% of all external clicks) because the browser back button provides a direct signal that the user intentionally returned to the search results after visiting the clicked webpage. In contrast, for clicks that open in a new tab, users often return to Google simply by closing the destination tab, which automatically reveals the original search-results tab underneath. Because such behavior does not necessarily indicate an active decision to return to search, new tab clicks are excluded from the back-to-search analysis.

[24] 4 minutes is the 95th percentile. Results are robust to removing this cap.