Marketing Science Institute Working Paper Series 2026

Report No. 26-102

Impact of AI Search Summaries on Website Traffic: Evidence from Google AI Overviews and Wikipedia

Mehrzad Khosravi and Hema Yoganarasimhan

Impact of AI Search Summaries on Website Traffic: Evidence from Google AI Overviews and Wikipedia

Mehrzad Khosravi *
University of Washington

Hema Yoganarasimhan
University of Washington

January 30, 2026

Abstract

Search engines increasingly display LLM-generated answers shown above organic links, shifting search from link lists to answer-first summaries. Publishers contend these summaries substitute for source pages and cannibalize traffic, while platforms argue they are complementary by directing users through included links. We estimate the causal impact of Google's AI Overview (AIO) on Wikipedia traffic by leveraging the feature's staggered geographic rollout and Wikipedia's multilingual structure. Using a difference-in-differences design, we compare English Wikipedia articles exposed to AIO to the same underlying articles in language editions (Hindi, Indonesian, Japanese, and Portuguese) that were not exposed to AIO during the observation period. Across 161,382 matched article–language pairs, AIO exposure reduces daily traffic to English articles by approximately 15%. Effects are heterogeneous: relative declines are largest for Culture articles and substantially smaller for STEM, consistent with stronger substitution when short synthesized answers satisfy informational intent. These findings provide early causal evidence that generative-answer features in search engines can materially reallocate attention away from informational publishers, with implications for content monetization, search platform design, and policy.

Keywords: Generative AI, Search Engines, Digital Economy, Artificial Intelligence, Wikipedia, Google

*Please address all correspondence to: mehrzad@uw.edu and hemay@uw.edu.

1 Introduction

Search engines are a central layer of the modern digital economy: they intermediate how consumers discover information, evaluate alternatives, and navigate to third-party content. For many publishers, this intermediation is existential rather than incidental. A sizable share of traffic to publishers of news and informational content comes through intermediaries such as search engines, leaving publishers highly exposed to changes in platform interfaces (Nielsen and Fletcher 2022). At the same time, the long-running shift toward "zero-click" search – where users obtain what they need on the results page without visiting an external site – has steadily tightened the link between search design and publishers' search traffic (Fishkin 2024). In this environment, even small reductions in click-through can translate into meaningful losses in advertising impressions, subscription conversions, and ultimately the economic viability of web publishing. For context, U.S. internet advertising revenue totaled $258.6B in 2024, of which $74.3B came from display advertising (Interactive Advertising Bureau (IAB) 2025).

Google's AI Overviews (AIO) represent a major step in this evolution from a "list of links" toward an "answer engine." AIOs generate a synthesized response that appears prominently on the search results page, typically alongside a limited set of cited sources (Reid 2023, 2024). AIOs are embedded directly in the search interface, and alter the consumer's choice set by presenting a complete answer *before* the traditional organic links (see Web Appendix §A for an example). Google began rolling out AIOs to a subset of U.S. search traffic in March 2024 (Schwartz 2024), and then deployed it for all U.S. users in May 2024 (Reid 2024). It then expanded AI Overviews internationally in subsequent months, reaching more than 100 countries by late 2024 (Venkatachary 2024).[1] Third-party audits suggest that AIOs now appear on approximately one in five or 18% of Google searches (Tabuena 2025; Chapekis and Lieb 2025), though this number varies by query type and measurement timeframe. Alongside AIO's rollout, Google has steadily moved to integrate advertising into the AIO experience. As of late 2025, these ad integrations have expanded across devices and geographies, making ads a core part of the AIO experience (Google 2025).

These developments have important implications for users, publishers, and search engines, and the broader content economy. Publishers argue that by placing AIOs above organic results, Google may reduce

[1] Other search engines are also introducing generative summaries and AIO-like features, e.g., Microsoft's Bing also introduced Generative Search in the summer of 2024, which is similar to Google's AIO (Bing 2024). Nevertheless, Google's move is probably the most critical development since it represents 90% of the global search market (Statista 2025).

scrolling and click-throughs to third-party sites, even when sources are cited, shifting eyeballs and advertising revenue toward the search platform. In their view, AIO serves as a *substitute* for publisher content and cannibalizes traffic (Davies 2025; Miller 2025); consistent with industry reports of declining organic visits as generative features diffuse across search platforms (Chapekis and Lieb 2025; SimilarWeb 2025). Google (and other platforms) counter that AIO and generative summaries are *complementary*: lowering search costs, improving user satisfaction, and sending users to helpful pages through included links, potentially expanding overall search activity and traffic (Reid 2024, 2025). The economic and policy stakes of this disagreement are substantial. If AIO systematically diverts attention away from publishers, it could shift advertising revenue and market power further into the hands of large search platforms and away from publishers. In the long run, it could also accelerate a negative feedback loop in which the content supply that search relies on becomes harder to fund. If AIO instead reallocates clicks toward higher-quality destinations, the feature may represent a net efficiency gain in information search.

This debate echoes prior work on news aggregators, which asks whether aggregation cannibalizes publisher demand (substitution) or expands it by improving discovery and lowering search costs (complementarity). Empirical evidence largely supports complementarity: removing Associated Press content from Google News reduced traffic to other publishers (Chiou and Tucker 2017), and the 2014 unavailability of Google News in Spain decreased traffic to Spanish publishers (Calzada and Gil 2020; Athey et al. 2021). However, AIO changes the mechanism in a way that makes direct extrapolation difficult. Unlike an aggregator that primarily routes users to linked pages, AIO presents a synthesized answer before organic links and can satisfy informational queries without a click. As a result, whether AIO ultimately substitutes for publisher visits or expands overall demand is an empirical question.

In this paper, we seek to empirically document the causal effect of Google's AI Overviews on publishers' inbound traffic, using Wikipedia as an exemplar. We focus on Wikipedia for three reasons. First, Wikipedia is among the top ten most visited websites globally, with over 130 billion page views per year (Wikipedia 2025a,b). Further, it is one of the most reliable and open sources of information on the Internet, plays a central role in the digital information economy, and as such, has been extensively studied in the literature (Zhang and Zhu 2011; Greenstein and Zhu 2012). Second, Wikipedia offers a uniquely transparent measurement environment: its daily article-level page view data are publicly available at a global scale, enabling traffic tracking without relying on proprietary publisher data (Wikimedia Foundation 2024a). Third, Wikipedia's multilingual structure allows us to observe the same underlying topics across language editions. This, combined with AIO's staggered rollout, supports a within-article difference-in-differences design that differences out persistent article popularity and nets out common time shocks.

A priori, it is not obvious whether AIO helps or hurts Wikipedia. Wikipedia receives a large share of its traffic from search engines; so any changes to SERP (Search Engine Results Page) that affect click-through should meaningfully affect its page views (McMahon et al. 2017; Vincent and Hecht 2021).[2] The direction, however, is ambiguous. On the one hand, AIO can substitute for Wikipedia by providing similar factual summaries directly at the top of the results page, potentially cannibalizing visits. On the other hand, Wikipedia is a highly reputable and frequently cited source in AIOs (Harsel et al. 2025); if being cited increases exposure or if AIO expands overall search traffic and engagement, Wikipedia could benefit.

We construct a large article–language–day panel from Wikimedia's public pageview API covering 46.5 million observations on 161,382 article-language pairs (52,262 English articles with matched versions in Hindi, Indonesian, Japanese, and Portuguese) from October 28, 2023, to August 14, 2024. Because Wikimedia pageviews are not disaggregated by readers' country, we proxy AIO exposure using language editions: English Wikipedia readership is most closely aligned with the first wave of the AIO rollout in the United States, beginning March 22, 2024, while the comparison languages were primarily exposed only with later international expansion. We then estimate a difference-in-differences model that compares each English article's traffic with that of its corresponding non-English versions, controlling for article–language and calendar-date fixed effects to absorb persistent cross-language differences in baseline popularity and common daily shocks.

Our estimates indicate that AIO meaningfully reduced downstream visits to Wikipedia's English pages by about 15%. This finding is robust to alternative timing and aggregation choices, e.g., redefining the post period to begin at the full U.S. launch yields a very similar decline (about 16–17%), and aggregating to weekly data produces an estimated decline of about 15%. Taken together, these patterns suggest that Google's AIO feature substituted for click-through to source pages (even for publishers like Wikipedia, which tend to be heavily cited in AIOs). Scaling the average estimate to our entire sample of 52,262 English language

[2]According to industry reports, 70-80% of Wikipedia's inbound traffic originates from organic search (Similarweb 2025; SemRush 2025).

articles implies roughly 11.5 million fewer daily visits for this sample after the introduction of AIO (or about 4.21 billion fewer visits per year). While Wikipedia does not show ads, this shift in traffic corresponds to economically meaningful revenue-equivalent losses for ad-supported publishers at a comparable scale. Further, these average effect masks substantial heterogeneity across content types, consistent with substitution being strongest when a short, synthesized answer can satisfy informational intent.

In sum, our study provides early causal evidence that generative-answer features like AIO, embedded in search platforms and intermediaries, can materially reallocate attention and traffic away from publisher pages, even for reputable websites like Wikipedia that tend to be highly cited in search summaries. Our findings have implications for publisher monetization, search-platform design, and the use of AI in summaries. From a policy perspective, our results suggest that digital platforms and policymakers may need to devise fairer and more incentive-compatible mechanisms to share ad revenues between digital intermediaries and content creators; e.g., through attribution mechanisms such as Shapley valuation (Ye and Yoganarasimhan 2025).

2 Setting and Data

Table 1 summarizes the main milestones in the evolution of Google's AI Overviews (AIO).[3] As we can see, AIO's rollout was staggered over time and across geographies. We use this staggered rollout to motivate the treatment window in our difference-in-differences design. In particular, we leverage the time period from March 22, 2024, to August 2024, where AIO was rolled out (partially or fully) in the United States, but not in other countries. This staggered rollout serves as a natural experiment for studying the impact of AIO on website traffic.

Table 1: Key milestones in Google's AIO rollout.

Date	Milestone
Mar. 22, 2024	Google begins testing AI Overviews in the main search interface on a subset of queries and a subset of search traffic in the U.S (Schwartz 2024).
May 15, 2024	Google rolls out AI Overviews for all users in the U.S. (Reid 2024).
Aug. 15, 2024	Google expands AI Overviews beyond the U.S. to India, Mexico, Japan, Indonesia, Brazil, and the UK (Budaraju 2024).
Oct. 28, 2024	Google rolls out AI Overviews to more than 100 countries and territories and additional languages (Venkatachary 2024).

We construct an article-language-day panel using publicly available Wikipedia page view data from the

[3]As a precursor to AIO, Google launched Search Generative Experience (SGE) in May 2023 to test and showcase some of its generative AI capabilities in search (Reid 2023). However, this was a small-scale, U.S.-only, opt-in experiment.

Wikimedia Analytics API. This REST API provides open access to project-level and page-level consumption metrics such as page views (Wikimedia Foundation 2024b). Our outcome variable, $\text{views}_{i,l,t}$, is the number of daily page views for article i in language edition l on calendar day t. We obtain these counts from the API's pageviews endpoint for individual pages (the `metrics/pageviews/per-article` route), which returns a daily time series of page views for a specified language project (e.g., `en.wikipedia.org`) and article title over a chosen window (Wikimedia Foundation 2025). We focus on all access types (desktop and mobile), consistent with the goal of capturing overall inbound attention to Wikipedia articles.

Table 2: Treatment structure by language group and period.

Group (Languages)	Pre-treatment period Oct 28, 2023 – Mar 21, 2024	Treatment period Mar 22 – Aug 14, 2024
Treated: English	✗	✓
Control: Hindi, Japanese, Portuguese, Indonesian	✗	✗

Pre-treatment and treatment periods: Our data collection period runs from October 28 2023 to August 14 2024. We choose Oct 28 2023 as the starting point since it gives us symmetric pre- and post-treatment periods (146 days each). Our pre-treatment period runs from October 28 2023 to March 21 2024 (when AIOs were introduced in the U.S), and our post-treatment period runs from March 22 2024 to August 14 2024 (one day before AIO was rolled out in India, Mexico, Japan, Brazil, and Indonesia).

Treatment and control languages: The Wikimedia API reports page views by project (language edition) and article, but does not provide filters based on the reader's country. Therefore, we proxy for exposure to AIO using language edition rather than geography.

We define the treated group as the English-language edition of Wikipedia (`en.wikipedia.org`), and the control group as the corresponding editions of the same underlying articles in Hindi, Japanese, Portuguese, and Indonesian (`hi`, `ja`, `pt`, `id`), when available. We chose these languages as the control group for two reasons. First, except Hindi, these are among the top 15 Wikipedia languages based on page views (Wikipedia 2024). Thus, they get sufficient volume of Wikipedia page views to provide data for the analysis and form a reliable control group for the English language edition. Second, since Google launched the AIO feature in the corresponding countries in August 2024, they can be viewed as the closest to English queries from the search engine's perspective. We exclude Spanish from the control group because Spanish Wikipedia has substantial viewership across multiple countries (e.g., the U.S., Spain, and Mexico); as such, it cannot be cleanly categorized as treatment or control.

Table 3: Summary statistics of daily pageviews by language: pre- and post-treatment periods (pre: $t <$ March 22, 2024; post: $t \geq$ March 22, 2024).

Pre-treatment (Oct. 28, 2023 – Mar. 21, 2024)						
stat	All	English	Portuguese	Japanese	Indonesian	Hindi
Mean	579.94	1,458.43	101.71	261.91	71.89	116.68
Std.	12,905.89	22,351.99	631.97	2,909.78	2,377.70	800.71
Min.	0.00	0.00	0.00	0.00	0.00	0.00
25%	10.00	33.00	8.00	18.00	2.00	3.00
50%	55.00	289.00	31.00	73.00	10.00	24.00
75%	269.00	1,276.00	96.00	226.00	44.00	104.00
Max.	12,579,026.00	12,579,026.00	182,144.00	1,161,021.00	3,325,857.00	457,026.00
Article-Language	161,382	52,262	32,350	38,805	25,442	12,523
Post-treatment (Mar. 22, 2024 – Aug. 14, 2024)						
Mean	487.45	1,216.15	91.95	238.79	52.46	71.95
Std.	12,771.75	22,171.94	650.03	2,828.64	650.97	321.01
Min.	0.00	0.00	0.00	0.00	0.00	0.00
25%	9.00	31.00	8.00	18.00	2.00	2.00
50%	48.00	257.00	28.00	70.00	8.00	15.00
75%	238.00	1,073.00	85.00	215.00	35.00	64.00
Max.	6,946,666.00	6,946,666.00	292,792.00	947,963.00	888,463.00	101,196.00
Article-Language	161,382	52,262	32,350	38,805	25,442	12,523

Article selection and panel construction: To define a stable set of comparable articles without conditioning on post-treatment outcomes, we construct our article universe using only pre-treatment popularity information. For each of the five language editions in our sample (English, Hindi, Japanese, Portuguese, Indonesian), we collect daily lists of the top 1,000 most-viewed pages using the API's "most-viewed pages" endpoint (`metrics/pageviews/top`) (Wikimedia Foundation 2024c). We do so for each day between July 1, 2023 and March 21, 2024, and take the union of all articles that appear in the top 1,000 at least once during this pre-treatment window.[4]

We then restrict attention to comparable cross-language units by retaining only articles that (i) have an English version and (ii) have at least one corresponding version in the control languages (Hindi, Japanese, Portuguese, or Indonesian). In the resulting analysis sample, we observe 52,262 English article pages and

[4]We choose July 1, 2023, as our article selection start date since it helps to have a larger sample of articles compared to setting it the same as our estimations start date, October 28th, 2023.

161,382 total article-language pairs across all five languages (Table 3). For each retained article-language pair, we query the Wikimedia Analytics API for daily page views over our analysis window surrounding the March 22, 2024 AIO test start, and assemble these into an article-language-day panel.

Table 3 reports summary statistics of daily page views by language, separately for the pre-treatment period (October 28, 2023 to March 21, 2024) and the post-treatment period (March 22 to August 14, 2024). We observe three patterns in the data. First, the magnitude of views of the English language articles is much larger than that of articles from other languages. Second, across all languages, daily page views are right-skewed: the mean is larger than the median, reflecting a small number of high-traffic pages. However, the mass on zeroes is not very high; i.e., most article-languages combinations get non-zero views on most days during the observation period. Third, average daily views are slightly lower in the post period compared to the pre-period for all language editions, indicating that our study window spans broader shifts in aggregate Wikipedia traffic.

3 Empirical Analysis

3.1 Model-Free Evidence

We first present descriptive evidence on how Wikipedia traffic evolved around the launch of Google AIO. Figure 1 plots total daily page views for articles in our sample, separately for English and for the pooled set of control languages. For each group, we subtract its own pre-treatment sample mean so that both series are demeaned and fluctuate around zero in the pre-treatment period; this makes the trajectories comparable despite the much higher overall magnitude of English page views.

Prior to the AIO test (first vertical dashed line), English and non-English page views largely move in parallel, with similar high-frequency variation. After the launch, especially during the test period, the English series declines significantly relative to the control languages. This divergence suggests that there was a significant and visible decline in English Wikipedia traffic following the introduction of Google AIO (in comparison to other languages).

3.2 Main Effect on Wikipedia Page Views

We now specify a differences-in-differences model with article-language-level daily page views as the outcome variable. In our analysis, we directly use levels (rather than log-transformed variables) as the

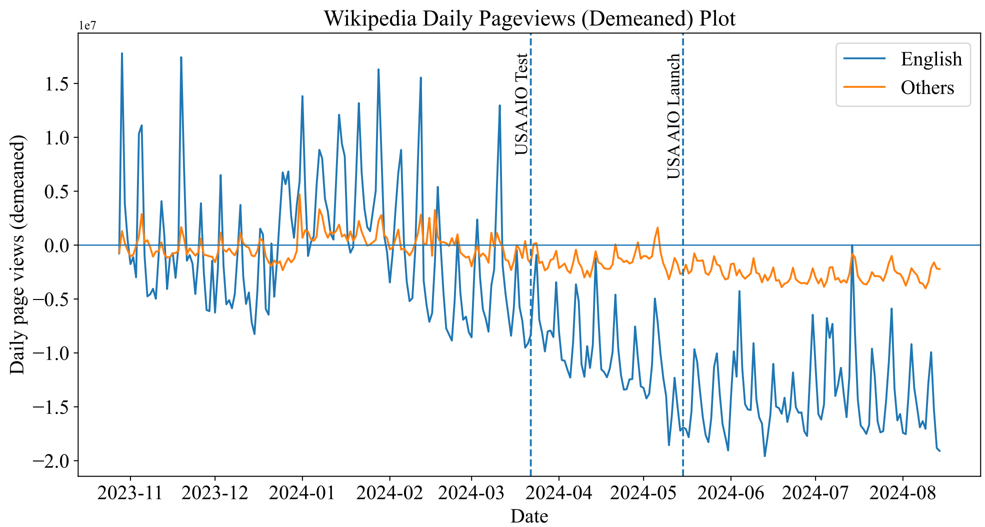

Figure 1: Plot of the page views evolution for English vs. the control group. The vertical lines identify the start of the Google AIO's test and launch periods.

outcome variable for three reasons. First, this choice aligns with the publisher-relevant estimand: advertising and subscription revenue scale with total page impressions, so an additive effect is directly interpretable as the number of visits gained or lost and aggregates transparently across pages. Second, AIO exposure varies across queries and time and is not observed at the page level. In this setting, log-based specifications effectively target average proportional changes and can place disproportionate influence on low-traffic pages, which are more likely to reflect measurement noise or lack of exposure than meaningful treatment intensity. Third, and most importantly, recent research (McConnell 2024) shows that using log-transformed outcome variables can flip the sign of interest when the levels of the treated and control groups differ significantly in magnitude (like ours). As such, McConnell (2024) recommends using levels analysis or a multiplicative model. We consider alternative specifications, including a multiplicative PPML model and a weighted-log model, as robustness checks in Web Appendix §C.1.

We consider the following difference-in-differences model, specified at the article-language-day level:

$$\text{views}_{i,l,t} = \alpha_{i,l} + \delta_t + \beta \left(\text{English}_l \times \text{Post}_t\right) + \varepsilon_{i,l,t}, \tag{1}$$

where i indexes the underlying article, l indexes language editions, and t indexes calendar days. The unit fixed effects $\alpha_{i,l}$ absorb all time-invariant differences across article–language pairs (e.g., baseline popularity and language-specific audience size of article i), and the date fixed effects δ_t absorb shocks common to all units on day t (e.g., seasonality, platform-wide demand shifts, Google's algorithm updates, or Wikipedia-wide outages). English_l equals one for the English edition and zero for the control languages, and serves as the treatment indicator. Post_t equals one on and after March 22, 2024, and indicates the treatment period. The coefficient of interest, β, captures the average post-launch change in English page views relative to contemporaneous changes in the corresponding non-English editions.

Identification follows the standard parallel-trends logic: absent AIO, the treated and control editions of the same articles would have evolved similarly over time (Angrist and Pischke 2009). While Figure 1 provides visual evidence in support of the parallel trends, we further assess this assumption using pre-trend tests and HonestDiD tests in Web Appendix §B. Throughout, we report cluster-robust standard errors that allow for arbitrary within-article correlation over time[5] and common shocks across articles within a day. This can matter in long panels where outcomes and interventions can be serially correlated (Bertrand et al. 2004). Concretely, we two-way cluster by article and date, consistent with the multi-way clustering framework in Cameron et al. (2011).

We report our estimation results in Table 4. First, Model 1 shows the estimates from a simple model without article-language and date fixed effects. Here, the estimated β coefficient implies a 16.6% decline relative to the pre-period English baseline level ($-242.28/(143.70+1314.734)$). Next, Model 2 includes article$\times$language and date fixed effects. The estimated effect remains negative and statistically significant, and similar in magnitude. Together, these findings suggest that the introduction of Google AIO led to a significant and large reduction in traffic to Wikipedia.

Further, we note that this estimated effect is likely to be conservative for a couple of reasons. First, AIO exposure within the U.S. was gradual across users and query classes, so not all English traffic was treated uniformly throughout the treatment period. (In §3.4 and Web Appendix §C.2, we also consider a model where we only consider the data from May 15th 2024 to see if there are differences in the effect sizes if we drop the intervening test period.) Second, English Wikipedia has a global readership, so a nontrivial share of English page views originates from countries with different AIO exposure timing; this mechanically attenuates the measured treatment intensity.

[5]This allows for shocks to be plausibly correlated within an underlying article across languages on a given day (topic-level news, seasonality, global interest in the subject, editing/featured status spillovers, etc.)

Table 4: Main effect of Google AIO launch on English articles' page views.

	Model 1 – Without FE		Model 2 – With FE	
English	1314.735***	(87.0274)		
post × English (DiD)	−242.283***	(11.4507)	−220.545***	(15.1456)
Constant	143.698***	(6.0364)	569.762***	(2.4135)
Observations	46,534,093		46,534,093	
Article – language FE	No		Yes	
Date FE	No		Yes	
Clusters (article and date)	52,262 and 292		52,262 and 292	
R^2	0.0019		0.9166	

Robust standard errors (in parentheses) are two-way clustered by article and date.
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

3.3 Heterogeneous Treatment Effects by Topic

The estimates in §3.2 average effects across a wide range of informational content, but substitution toward AIOs likely varies by topic. Some queries (e.g., celebrity and cultural reference) may be satisfied by a short summary, whereas others (e.g., STEM or socially contextual topics) may still prompt users to click for nuance, context, or verification. Moreover, if AIO makes search more effective and increases searching, any market-expansion effects may also differ across categories. We therefore re-estimate the analysis by topic using Wikipedia's categorization, which assigns each page to one of four topics: Culture, Geography, History & Society, and STEM.

We start with descriptive evidence. Figure 2 plots demeaned daily page views for English versus pooled control-language editions separately for each of the four categories (Culture, Geography, History & Society, and STEM). Consistent with the pooled pattern, English and non-English editions move similarly prior to March 22, 2024, and then English page views fall relative to controls within each topic, with magnitudes that differ visibly across categories. Further, the effects are more visually obvious for Culture and History/Society.

Next, we re-estimate our difference-in-differences specification separately for each topic as follows: for each topic k, we restrict the sample to articles with $\text{Topic}_i = k$ and estimate Equation (1) with article×language and date fixed effects. This yields a topic-specific average treatment effect that compares the English edition to the same articles' non-English editions within that topic, netting out common time shocks. We report the topic-specific estimates in Table 5. We find a negative and statistically significant effect in all four categories, but the magnitude varies substantially. The largest percentage decline compared to baseline numbers occurs

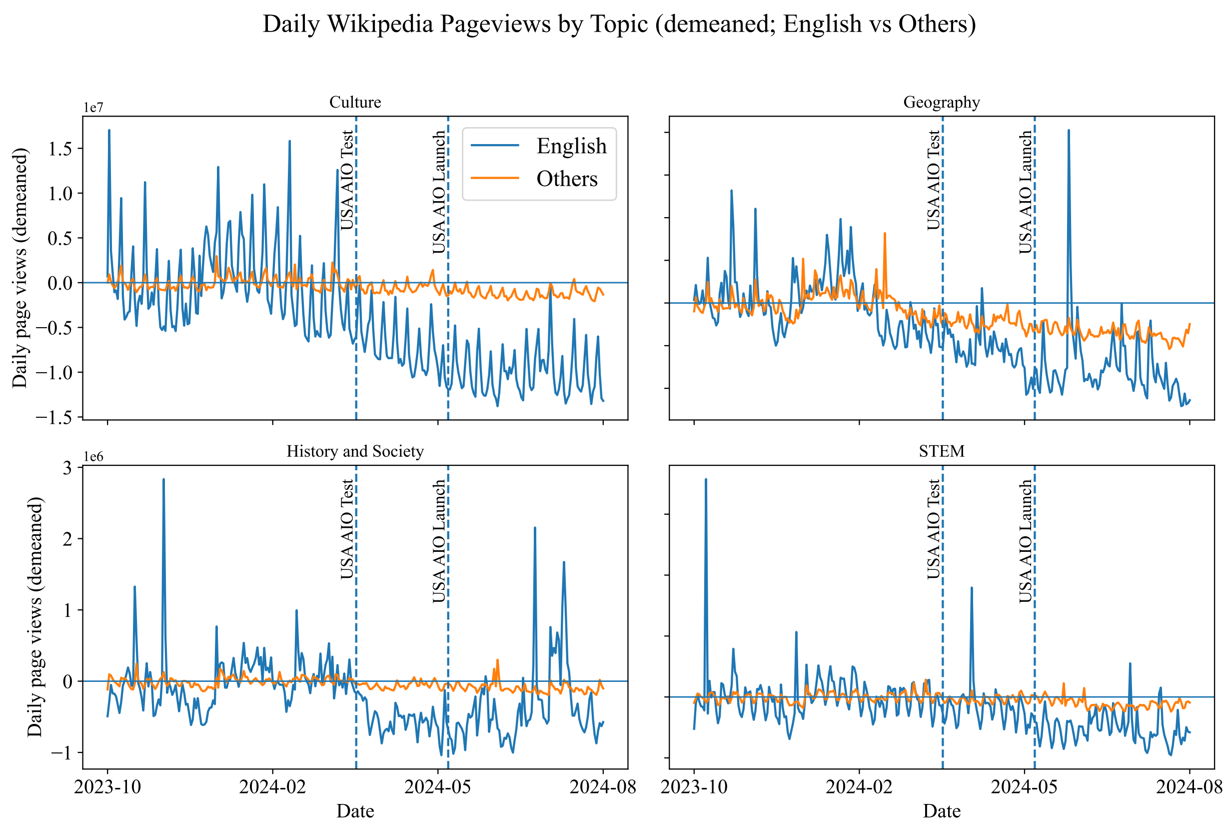

Figure 2: Plot of the page views evolution for English vs. the control group for different topics. The vertical lines identify the start of the Google AIO's test and launch periods.

for Culture articles (-19.6%), while the decline is smallest for STEM articles (-7.4%). Geography (-16.6%) and History & Society (-9.9%) fall in between.[6] Taken together, these results indicate that AIO is associated with a broad-based reduction in English Wikipedia traffic, but that the effect is more pronounced for some topics, such as culture.

3.4 Robustness Checks

We also conduct a series of robustness checks to assess the robustness and stability of our main findings. We briefly summarize them below and refer readers to Web Appendices §B and §C for details.

- **Excluding the test period.** In our main specification, we define the post-treatment period as beginning March 22, 2024, when AIO testing began in the U.S. To check whether the results are significantly influenced by the inclusion of the test period, we re-estimate the model excluding this test window and

[6] We use the results from regression without fixed-effects in Web Appendix §C.4 to calculate the percentage changes compared to the baseline.

Table 5: Topic-specific effect of Google AIO on Wikipedia page views (separate regressions by topic)

	Culture	Geography	History & Society	STEM
$\text{English}_l \times \text{Post}_t$	-324.3735^{***}	-105.6416^{***}	-143.5114^{***}	-170.1528^{***}
	(24.7378)	(17.8534)	(31.1137)	(40.3879)
Constant	655.9673^{***}	400.5817^{***}	518.3416^{***}	744.4963^{***}
	(3.8963)	(3.2098)	(4.1066)	(4.7139)
Observations	22,875,713	15,673,940	2,808,950	5,000,138
Article – language FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
Clusters (article and date)	25,542 and 292	19,346 and 292	2,745 and 292	4,408 and 292
R^2	0.2013	0.1841	0.3899	0.9840

Robust standard errors (in parentheses) are two-way clustered by article and date.
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

define the post-period as beginning on May 15, 2024, the date of full U.S. launch. The results remain statistically significant and of similar magnitude (see Web Appendix §C.2).

- **Alternative Model Specifications (PPML and Weighted Log).** We also re-estimate our main effect using alternative model specifications. First, we consider a Poisson pseudo-maximum likelihood (PPML) specification, which models proportional changes. This model is often used with skewed, nonnegative outcomes like page views, when the magnitudes of the treatment and control groups are significantly different (instead of logging the outcome variable); see McConnell (2024). The estimated effect is a 3.5% decline in English Wikipedia views relative to control languages (see Web Appendix §C.1.1).[7] Second, we estimate a weighted log specification. This model uses a log transformation of the views, $\log(view_{i,l,t}+1)$, as outcome and weighs the articles in the OLS based on their pre-treatment traffic share. It reflects proportional changes in the aggregate traffic rather than an average article by downweighting the low-traffic observations (Solon et al. 2015; Autor et al. 2020). Consistent with the previous results, the effect on the English articles is negative (-8.1%) and significant (see Web Appendix §C.1.2).
- **Restricting to high-traffic articles common across languages.** We repeat our analysis using a more restrictive sample of articles that appear in both English and at least one control-language top-1000 list during the pre-period. Although this reduces the article sample by 90%, the estimated treatment effect remains negative and statistically significant—and is even larger in magnitude (see Web Appendix §C.3).

[7]This effect is smaller in magnitude than the levels-based estimate because PPML down-weights high-traffic outliers and captures average *proportional* changes rather than absolute losses. This makes it more conservative when the treatment disproportionately affects high-volume articles.

- **Topic-level interaction model.** In addition to estimating separate regressions by topic, we also run a pooled model with interaction terms between treatment and topic indicators. The relative ranking of effects is consistent with our main estimates (see Web Appendix §C.4).
- **Weekly aggregation.** We aggregate article-language page views to the weekly level and re-estimate the difference-in-differences model. The estimated effect is statistically significant and consistent with the daily results, with English pages receiving roughly 1,504 fewer weekly views per article or a 14.8% decline (see Web Appendix §C.5).
- **Parallel trends validation.** We perform linear pre-trend tests and event-study diagnostics, and find that there are no significant and systematic deviations from the parallel trends assumption. Additionally, we also run an HonestDiD sensitivity analysis and find that our main conclusions are robust to plausible violations of parallel trends (see Web Appendix §B).

Together, these robustness checks confirm that our main results are not driven by pre-existing trends, choice of outcome scale, specific sample construction, or temporal aggregation.

4 Economic Impact and Implications

Our estimates imply that the introduction of AI Overviews (AIO) reallocates a nontrivial amount of user attention from publishers (Wikipedia in this case) to the search results page. We now use our difference-in-differences estimates to perform a simple back-of-the envelope calculation quantifying how this shift in user attention can change the distribution of ad dollars from publishers to ad platforms/search engines.

Recall that the deployment of AIO is associated with $\hat{\beta} = -220.54$ fewer daily views for the English version of an article (on average). So, the implied change in total daily views for our sample (of 52262 english-language articles) is:

$$\Delta\widehat{PV}_{day} \;=\; \hat{\beta} \times 52262 \;\approx\; -11.53 \text{ million page views/day}, \tag{2}$$

which results in about 4.21 billion fewer page views per year. Because AIO exposure is staggered and English Wikipedia traffic is global, these quantities should be interpreted as lower bounds on the traffic reallocation that would occur under full AIO exposure for English queries. Further, such traffic/attention changes can feed back into content supply, even when content is not directly monetized. Indeed prior work on Wikipedia has shown that changes to its audience/community size can affect contributors' incentive to edit/produce content (Zhang and Zhu 2011).

Although Wikipedia does not monetize traffic through advertising/subscriptions, for many information publishers, page views map directly into advertising revenue. We therefore compute a revenue-equivalent effect by applying plausible ad-market prices to the implied change in page views. Most display advertising is transacted on a CPM (cost per thousand impressions) basis, and publishers typically retain 70% of the CPM as commission. While the number of display ads shown and monetized per page varies across publishers and topics, we assume the average is 3.7 and use this in our calculations. [8] Then, the implied publisher revenue per thousand page views (page RPM) is approximately:

$$\text{RPM} \approx 3.7 \times 0.70 \times \text{CPM} = 2.59 \times \text{CPM}. \tag{3}$$

Average CPMs for programmatic display ads usually fall between \$3.28 and \$9.37 (Association of National Advertisers (ANA) 2025).[9] This corresponds to a RPM $\approx$ \$8.50–\$24.28 per 1,000 page views. Applying these benchmarks to our estimate of $|\Delta\widehat{PV}_{day}| \approx 11.53$ million lost page views per day yields a daily revenue impact of

$$\Delta\widehat{\text{Rev}}_{day} = \frac{|\Delta\widehat{PV}_{day}|}{1000} \times \text{RPM} \approx \text{\$0.098M to \$0.279M per day},$$

That is equivalent to an annual loss of income of $365 \times \Delta\widehat{\text{Rev}}_{day} \approx$ \$35.71M to \$102.12M USD for a publisher like Wikipedia. To the extent that the traffic and ad revenues lost by publishers like Wikipedia are transferred to AIOs (and ads shown within the AIOs), these numbers can interpreted as a transfer in ad revenue from publishers to search engines.

Beyond traffic and advertising, we note a few other implications for publishers, platforms, and policy-makers. First, publishers that monetize through subscriptions, lead generation, or commerce were also likely to have been impacted by AIOs, as fewer search referrals reduce the traffic to the top of the funnel. Beyond traffic and advertising, AIO can also affect publishers that monetize via subscriptions, lead generation, or

[8] Mainstream news article pages display about 5.22 ads per page, on average (Zeng et al. 2020). Further, industry reports suggest that 71% of display impressions are typically viewed per page (DoubleVerify 2024). Together, this implies about $5.22 \times 0.71 \approx 3.7$ monetizable display impressions per page view.

[9] We note that there can be significant variations in these numbers based on the publisher reputation, topic of the page, etc. As such, these numbers are meant to denote representative averages.

commerce by reducing search referrals at the top of the funnel. More broadly, as discovery and conversion shift "within the platform," behavioral data and measurement feedback concentrate at the search and ad platform, increasing publishers' reliance on platform-controlled targeting, auctions, and attribution. This raises policy concerns around – (i) market and gatekeeper power: UI changes can shift traffic and value with limited visibility or recourse (Greenstein 2010) and the broader need for algorithmic accountability (Ukanwa et al. 2023); (ii) data ownership and copyright: if AIOs and similar AI-based summarizations use publisher content for generating AIOs and monetizing them without sharing the revenues with publishers, this raises concerns around fair attribution for content used to generate answers. Indeed, this issue is now a source of a variety of lawsuits in the AI space (Reuters 2026). See Ye and Yoganarasimhan (2025) for recently proposed solutions in this space, and (iii) data governance: greater concentration of interaction data can advantage platforms and intermediaries, strengthening the case for standards around measurement and interoperability. We note that many of these issues have been the focus of research and policy debate even before the evolution of AI; see Goldfarb and Tucker (2019) for a broader discussion. The growth and impact of AI and AI-powered technology further call for additional research and policy on these issues.

5 Conclusion

Google's AIO shift search toward an answer-first interface, raising the possibility that users satisfy informational intent on the results page rather than clicking through to publisher sites. Using Wikipedia as a transparent measurement environment and exploiting the staged rollout of AIO, we estimate a within-article, cross-language difference-in-differences design comparing English Wikipedia traffic (earlier exposed) to versions of the same articles in Hindi, Indonesian, Japanese, and Portuguese (later exposed).

We find clear evidence of traffic displacement. After the introduction of Google's AIO feature in US, English Wikipedia page views fall by about 15% relative to the comparison languages. Effects are heterogeneous across topics: declines are largest for Culture and smallest for STEM, consistent with stronger substitution when a short synthesized answer can resolve the query. Because exposure is imperfect in a language-based design (global English readership and gradual rollout), we interpret these estimates as conservative lower bounds. Our findings imply that generative-answer features such as AIO in search engines can materially reallocate attention away from publisher pages, with direct consequences for monetization via advertising and the top of the funnel for subscriptions and commerce.

We close with a few caveats and suggestions for future research. First, our research focuses on one large publisher (Wikipedia); it is possible that the magnitude and direction of the effect of AIO on other publishers is different. In particular, we note that Wikipedia is unique in two ways. On the one hand, it is the canonical answer repository for informational queries, for which AIOs serve as a direct substitute. On the other hand, it is also one of the websites that tends to be heavily cited in AIOs (unlike many smaller publishers), which may have mitigated the impact on Wikipedia. Thus, it is hard to extrapolate the effect to other types of websites. It would be interesting to see how AIOs and generative summaries in search engines affect traffic to smaller publishers and/or sites that are largely transactional or e-commerce based. Second, while the focus here was on Google AIO, our findings are also relevant to other traditional search engines such as Bing, as well as AI-based answer engines or summarization tools; for example, OpenAI's ChatGPT Search (OpenAI 2024) and Perplexity AI (Perplexity AI 2025). Thus, the overall effect on publishers can be amplified even further. Finally, AIOs and LLM-based summarization features are still evolving and their long-term impact on the digital economy is still unclear. One interesting direction would be to see how publisher quality and incentives to contribute to repositories such as Wikipedia changes, and how this in turn affects AIOs themselves.

Funding and Competing Interests Declaration

Author(s) have no competing interests to declare.

References

J. D. Angrist and J.-S. Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 2009.

Association of National Advertisers (ANA). Programmatic transparency benchmark report (q1 2025), 2025. URL `tinyurl.com/s3422r3x`.

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

D. Autor, D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen. The fall of the labor share and the rise of superstar firms. *The Quarterly journal of economics*, 135(2):645–709, 2020.

M. Bertrand, E. Duflo, and S. Mullainathan. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275, 2004.

Bing. Introducing bing generative search. Article, July 2024. URL `tinyurl.com/4epze5jb`. Accessed: 2026-01-19.

H. Budaraju. New ways to connect to the web with AI overviews. Google Product Blog, August 2024. URL `tinyurl.com/mr2zme39`. Accessed: 2025-12-19.

J. Calzada and R. Gil. What do news aggregators do? evidence from google news in spain and germany. *Marketing Science*, 39(1):134–167, 2020.

A. C. Cameron, J. B. Gelbach, and D. L. Miller. Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249, 2011.

A. Chapekis and A. Lieb. Google users are less likely to click on links when an AI summary appears in the results. *Pew Research Center*, July 2025. URL `tinyurl.com/yk4x87kr`.

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

J. Davies. Google ai overviews linked to 25URL `tinyurl.com/4t7bfs99`.

DoubleVerify. Dv global insights: 2024 trends report. Report, July 2024. URL `tinyurl.com/3y3dzfwf`. Accessed: 2026-01-19.

R. Fishkin. 2024 zero-click search study: For every 1,000 EU google searches, only 374 clicks go to the open web. in the US, it's 360. SparkToro Blog, July 2024. URL `tinyurl.com/2jx8tw6n`.

A. Goldfarb and C. Tucker. Digital economics. *Journal of economic literature*, 57(1):3–43, 2019.

Google. More opportunities for your business on google search. `tinyurl.com/2snbhp7x`, May 2025. Accessed: 2026-01-15.

S. Greenstein. Gatekeeping economics. *IEEE Micro*, 30(5):102–104, 2010.

S. Greenstein and F. Zhu. Is wikipedia biased? *American Economic Review*, 102(3):343–348, 2012.

L. Harsel, A. Drozdov, and C. Skopec. The most-cited domains in ai: A 3-month study. `tinyurl.com/ym2ss4p3`, 2025. Semrush blog post, published November 10, 2025.

Interactive Advertising Bureau (IAB). Digital ad revenue surges 15% yoy in 2024, climbing to $259b, according to iab. IAB news release (conducted by PwC), April 2025. URL `tinyurl.com/ym25uh64`. Accessed 2025-12-20.

B. McConnell. Can't see the forest for the logs: on the perils of using difference-in-differences with a log-dependent variable, 2024.

C. McMahon, I. Johnson, and B. Hecht. The substantial interdependence of wikipedia and google: A case study on the relationship between peer production communities and information technologies. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11, pages 142–151, 2017.

M. Miller. New user trends on wikipedia, October 2025. URL `tinyurl.com/3r8k9ysa`.

R. K. Nielsen and R. Fletcher. Concentration of online news traffic and publishers' reliance on platform referrals: Evidence from passive tracking data in the uk. *Journal of Quantitative Description: Digital Media*, 2, 2022.

OpenAI. Introducing ChatGPT search, 2024. Link. Accessed on May, 2025.

Perplexity AI, 2025. Link. Accessed on May, 2025.

E. Reid. Supercharging search with generative AI, May 2023. URL `tinyurl.com/yc5xkbm3`.

E. Reid. Generative AI in search: Let Google do the searching for you. Google Blog, May 2024. URL `tinyurl.com/4y3ehssc`.

E. Reid. AI in search: Driving more queries and higher quality clicks, August 2025. URL `tinyurl.com/ykk2ucf2`.

Reuters. Google defends ai search summaries against publishers' lawsuit. `https://www.reuters.com/legal/litigation/google-defends-ai-search-summaries-rolling-stone-publishers-lawsuit-2026-01-1` Jan. 2026. Reuters Legal, accessed January 2026.

B. Schwartz. Google starts testing ai overviews from sge in main google search interface, March 2024. URL `tinyurl.com/bddcsh4r`.

SemRush. Wikipedia traffic stats, 2025. URL `tinyurl.com/2wzzfxw7`.

SimilarWeb. The impact of generative AI: Publishers, July 2025. URL `tinyurl.com/nhb5ajm2`.

Similarweb. Wikipedia website performance, 2025. URL `tinyurl.com/54wfb8vz`.

G. Solon, S. J. Haider, and J. M. Wooldridge. What are we weighting for? *Journal of Human resources*, 50(2):301–316, 2015.

Statista. Market share of leading search engines worldwide from january 2015 to december 2025. Article, 2025. URL `https://tinyurl.com/f46smwst`. Accessed: 2026-01-19.

E. J. Tabuena. New ahrefs data outline how to optimize for ai overviews. `tinyurl.com/2kwfy95z`, 2025. DesignRush News. Accessed: 2026-01-19.

K. Ukanwa, W. Rand, and P. P. Zubcsek. Why firms should want algorithmic accountability. *USC Marshall School of Business Research Paper Sponsored by iORB*, 2023.

S. Venkatachary. AI overviews in search are coming to more places around the world, Octobor 2024. URL `tinyurl.com/mrx9xfkc`.

N. Vincent and B. Hecht. A deeper investigation of the importance of wikipedia links to search engine results. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1):1–15, 2021.

Wikimedia Foundation. Page view analytics — Wikimedia analytics API. Wikimedia Documentation, June 2024a. URL `tinyurl.com/2mpr99bw`. Accessed: 2025-12-18.

Wikimedia Foundation. Getting started — wikimedia analytics API. Online documentation, 2024b. URL `tinyurl.com/4wt78pzb`. Accessed: 2025-12-19.

Wikimedia Foundation. Project metrics — wikimedia analytics API. Online documentation (examples), 2024c. URL `tinyurl.com/4taukjcs`. Accessed: 2025-12-19.

Wikimedia Foundation. Page metrics — wikimedia analytics API. Online documentation (examples), 2025. URL `tinyurl.com/426msm4c`. Accessed: 2025-12-19.

Wikipedia. Languages used on the internet, 2024. URL `tinyurl.com/ynwy2ajc`.

Wikipedia. List of most-visited websites. `tinyurl.com/4css9t5x`, 2025a. Accessed: 22 January 2026.

Wikipedia. Wikipedia:statistics. `tinyurl.com/y32j75p6`, 2025b. Accessed: 22 January 2026.

Z. Ye and H. Yoganarasimhan. Fair document valuation in llm summaries via shapley values. *arXiv preprint arXiv:2505.23842*, 2025.

E. Zeng, T. Kohno, and F. Roesner. Bad news: Clickbait and deceptive ads on news and misinformation websites. In *Workshop on Technology and Consumer Protection (ConPro '20)*, 2020. URL `tinyurl.com/2dksmpz4`. Accessed: 2026-01-19.

X. Zhang and F. Zhu. Group size and incentives to contribute: A natural experiment at chinese wikipedia. *American Economic Review*, 101(4):1601–1615, 2011.

Appendix

A Search Results Pages with and without AIO

In this section, we present two examples of Google's search engine results (SERPs) without and with AIO. As we can see in Figures A1 and A2, AIO pushes the top search results (including Wikipedia in this example) down in the SERP.

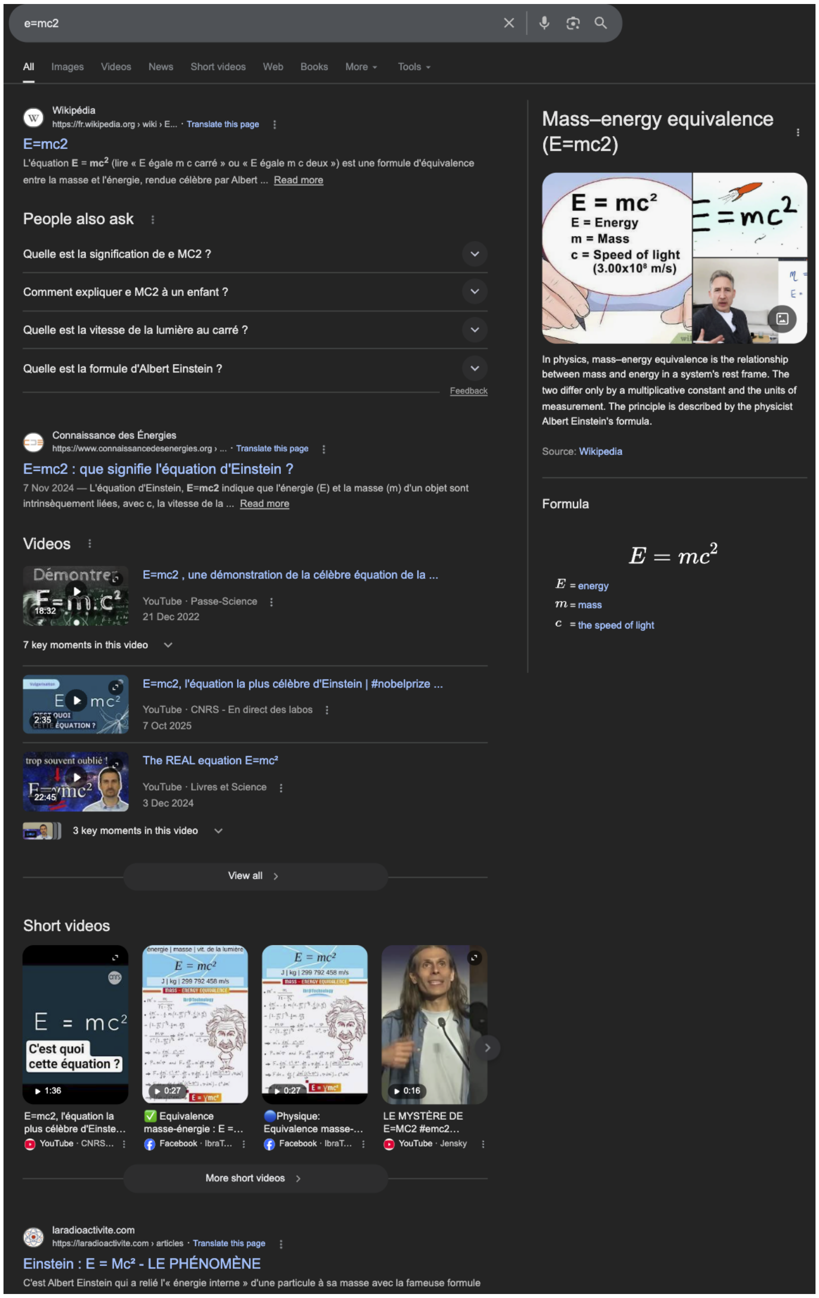

Figure A1: A Google search result page *without* AIO for the query "e=mc2."

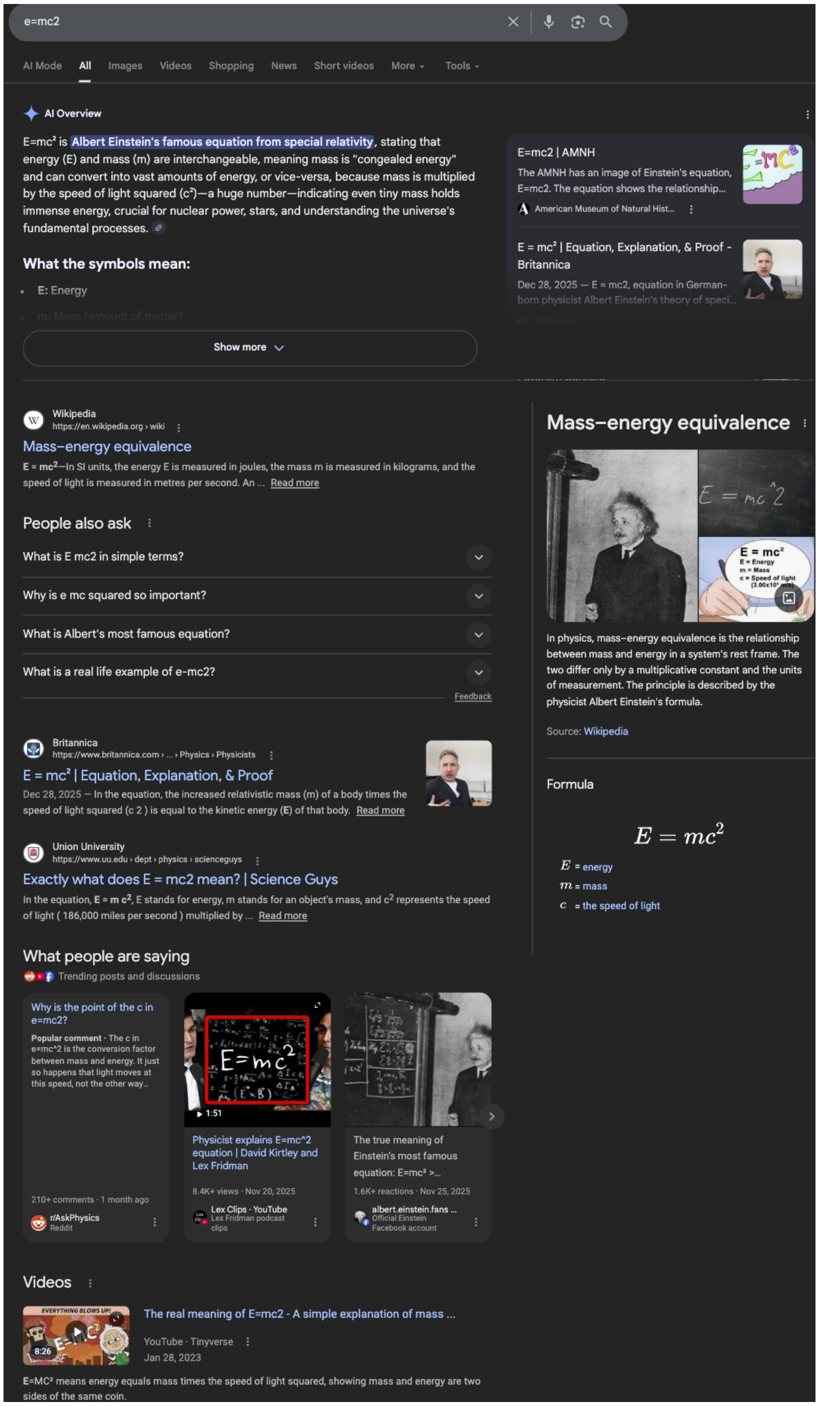

Figure A2: A Google search result page *with* AIO for the query "e=mc2."

B Parallel Trends Test

The validity of our difference-in-differences design hinges on a parallel trends assumption: absent AIO, English and non-English editions of the same article would have exhibited similar evolution in page views over time. We probe this assumption in three complementary ways: (i) a linear (parametric) pre-trend test, (ii) a dynamic event-study specification, and (iii) a sensitivity analysis that quantifies how large violations of parallel trends would need to be to change our conclusions.

B.1 Linear pre-trend test

As a simple parametric check, we estimate a linear differential-trend model using only pre-treatment observations ($\text{Post}_t = 0$):

$$\text{views}_{i,l,t} \;=\; \alpha_{i,l} \;+\; \delta_t \;+\; \eta\big(\text{English}_l \times t\big) \;+\; \varepsilon_{i,l,t}, \tag{A1}$$

where t is a linear time index (days since the start of the sample), $\alpha_{i,l}$ are article–language fixed effects, and δ_t are day fixed effects. The coefficient of interest, η, captures whether English versions exhibit a systematically different linear trend than non-English ones during the pre-period. Under parallel trends, we expect $\eta \approx 0$.

According to the results in Table A1, the estimated differential trend is small in magnitude and statistically indistinguishable from zero (p-value $= 0.068$). This provides no evidence of a meaningful *linear* divergence between English and non-English editions prior to treatment.

Table A1: Pooled linear pre-treatment parallel trend test with unit and day fixed effects

	Page views	
English $\times$ t (η)	-0.4967	(0.2696)
Constant	591.7003***	(6.1960)
Observations	23,277,043	
Article – language FE	Yes	
Day FE	Yes	
Clusters	Article = 52,262;	date = 146
R^2	0.8869	

Sample restricted to pre-period observations ($\text{Post}_t = 0$). Dependent variable is views. Unit (article-language) and day fixed effects are included. Two-way cluster-robust standard errors by article and day. Linear time is measured as $t = \text{date} - \text{Oct. } 28, 2023$. Significance: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

B.2 Event study (dynamic pre-trends)

We also test for differential *dynamic* behavior by estimating a relative-time event-study specification:

$$\text{views}_{i,l,t} \;=\; \alpha_{i,l} \;+\; \delta_t \;+\; \sum_{\tau \in \mathscr{T} \setminus \{\tau_0\}} \beta_\tau\big(\text{English}_l \times \mathbf{1}\{t - t^* = \tau\}\big) \;+\; \varepsilon_{i,l,t}, \tag{A2}$$

where t^* denotes the event date and τ indexes relative time (in days). The omitted period τ_0 serves as the reference. Each β_τ captures the difference in mean views between English and non-English editions at relative time τ, minus the corresponding difference in the reference period.

Figure A3 plots the estimated β_τ coefficients. Some pre-treatment coefficients are statistically different from zero, suggesting evidence against strict parallel trends. However, with a long pre-treatment window and substantial day-to-day variation in traffic (e.g., seasonality, holidays, and platform-wide shocks), individual relative-time coefficients can be noisy. For this reason, we complement the event-study visualization with the linear pre-trend test in Table A1 and with a formal sensitivity analysis described next.

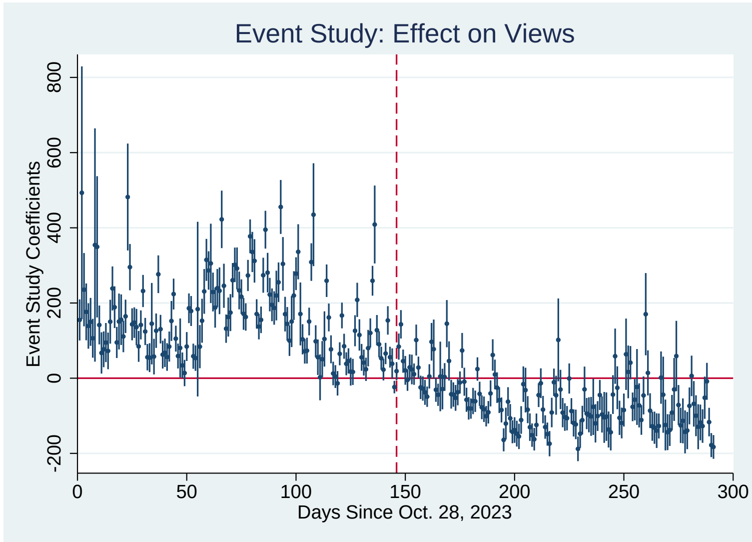

Figure A3: Event-study coefficients for the daily data (Oct. 28, 2023–Aug. 14, 2024). The vertical dashed line marks the event date t^* (March 22, 2024).

B.3 Sensitivity analysis with HonestDiD

Finally, rather than treating parallel trends as an all-or-nothing assumption, we quantify how inference changes when we allow for bounded deviations from parallel trends in the post-treatment period. The HonestDiD approach constructs confidence intervals that remain valid under researcher-specified restrictions on the evolution of the (unobserved) differential trend between treated and control groups (Rambachan and Roth 2023).

In our application, we use a smoothness restriction that bounds changes in the *slope* of the differential trend across consecutive periods:

$$\Delta^{SD}(M) := \{\lambda : |(\lambda_{t+1} - \lambda_t) - (\lambda_t - \lambda_{t-1})| \leq M, \forall t\}, \tag{A3}$$

where λ_t denotes the treated–control difference in counterfactual trends at time t, and M governs how far the post-treatment differential trend is allowed to deviate from a linear extrapolation of the pre-period. The special case $M = 0$ corresponds to an exactly linear differential trend. We report robust confidence intervals for our post-treatment estimand across a range of M values in Figure A4, making transparent how large departures from parallel trends would need to be to overturn the main conclusions. According to the results, a linear deviation from parallel trends still results in a negative, statistically significant effect at the 95% confidence level. This holds for the M values as large as 0.2.

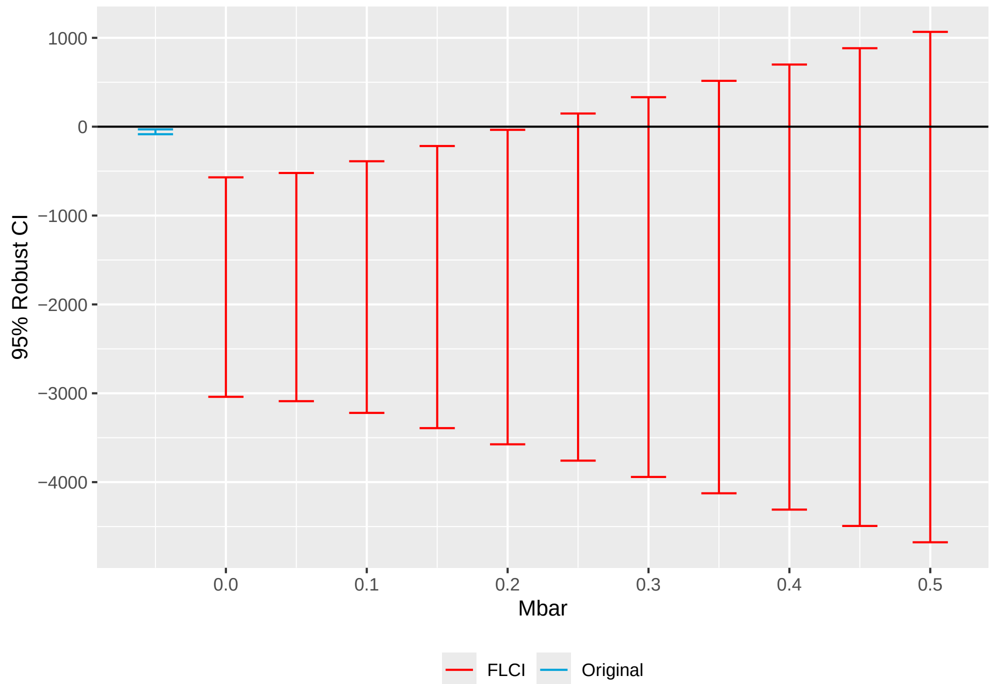

Figure A4: Daily HonestDiD estimation for page views.

C Appendix to Robustness Checks

C.1 Alternative Model Specifications

To assess the sensitivity of our findings to functional-form and distributional assumptions, we re-estimate our baseline two-way fixed effects difference-in-difference under PPML and Weighted Log specification.

C.1.1 PPML Specification

As a robustness check to our baseline two-way fixed-effects difference-in-differences model in levels (Equation 1), we also estimate a multiplicative specification using Poisson pseudo-maximum likelihood (PPML). PPML is a natural alternative in our setting because daily page views are nonnegative and highly right-skewed. Moreover, log-linearizing the outcome can induce bias under heteroskedasticity due to Jensen's inequality (McConnell 2024). PPML directly estimates the conditional mean in levels while allowing for a log-link (exponential mean) and remains consistent under standard quasi-maximum likelihood conditions even when the data are not literally Poisson (Wooldridge 2010). Formally, we assume an exponential conditional mean for article-language-day views:

$$\mathbb{E}[\text{views}_{i,l,t} \mid \alpha_{i,l}, \delta_t] = \exp\big(\alpha_{i,l} + \delta_t + \beta\left(\text{English}_l \times \text{Post}_t\right)\big)\,, \tag{A4}$$

where $\alpha_{i,l}$ are article$\times$language fixed effects and δ_t are date fixed effects, as in the main specification. We estimate Equation (A4) by PPML and report two-way cluster-robust standard errors by article and date.

A key advantage of the PPML formulation is interpretability: β is a semi-elasticity of the conditional mean, so the implied proportional effect is

$$\%\Delta \approx 100 \times (\exp(\beta) - 1)\,.$$

Table A2 reports the results. The estimated DiD coefficient is negative ($\hat{\beta} = -0.0353$) and statistically significant. Interpreting the coefficient through the exponential link implies that AIO is associated with an approximate $100 \times (\exp(-0.0353) - 1) \approx -3.47\%$ change in expected English page views relative to control-language editions.

The PPML estimate is smaller in magnitude than the main levels estimate, which is expected because the two models target different objects: the baseline OLS specification estimates an average additive change in views, while PPML estimates an average proportional change in the conditional mean (McConnell 2024). Importantly, both approaches yield the same qualitative conclusion: English Wikipedia page views decline after the onset of AIO exposure relative to later-exposed language editions.

Table A2: Main effect of Google AIO launch on the English articles' views (PPML).

	Page views (PPML)	
post $\times$ English (DiD)	-0.0353^{**}	(0.0131)
Constant	8.2886^{***}	(0.0047)
Observations	46,534,093	
Article – language FE	Yes	
Date FE	Yes	
Clusters	Article = 52,262;	date = 292

Outcome variable: views. Estimator: PPML with article$\times$language and date fixed effects.
Robust standard errors in parentheses are two-way clustered by article and date.
Significance: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

C.1.2 Weighted Log Specification

As an additional robustness check, we also consider a weighted log specification. Unlike the basic log-specification that can lead to incorrect estimates due to Jensen's inequality when the magnitude of the baseline outcomes across treated and control groups varies systematically (McConnell 2024), this specification accounts for systematic differences in baseline scale by reweighting observations with predetermined pre-period traffic (e.g., an article's pre-period view share). The goal is to prevent low-traffic units from receiving the same influence as high-traffic units in the log regression. As a result, we get a better view of the proportional changes in aggregate (Solon et al. 2015; Autor et al. 2020). Formally, we estimate:

$$\ln(views_{i,l,t} + 1) = \alpha_{i,l} + \delta_t + \beta\,(English_l \times Post_t) + \varepsilon_{i,l,t}, \tag{A5}$$

where $\alpha_{i,l}$ are article$\times$language fixed effects and δ_t are date fixed effects, as in the main specification. The $\ln(views + 1)$ transformation accommodates zero counts.

We estimate this equation by weighted least squares using analytic weights that depend only on pre-treatment traffic. Let $V^{\text{pre}}_{i,\text{en}} = \sum_{t\in\text{Pre}} views_{i,\text{en},t}$ denote total pre-period English views for article i, and define the article-level view-share weight $w_i = V^{\text{pre}}_{i,\text{en}} / \sum_j V^{\text{pre}}_{j,\text{en}}$. To ensure that an article does not mechanically receive more total control weight simply because it exists in more comparison languages, we use language-balanced weights

$$w_{i,l} = \begin{cases} w_i, & l = \text{en}, \\ w_i / N_{i,\text{non-en}}, & l \neq \text{en}, \end{cases}$$

where $N_{i,\text{non-en}}$ is the number of distinct non-English control editions of article i in our estimation sample. As in the baseline, we report two-way cluster-robust standard errors by article and date. In this model, β is reported in log points, and the exact proportional effect is $\exp(\beta)-1$. According to the estimation results in Table A3, the effect is negative (an 8.1% decline in page views) and statistically significant, which is directionally consistent with previous results.

Table A3: Main effect of Google AIO launch on the English articles' views (weighted log OLS).

	Log page views (weighted OLS)	
post × English (DiD)	-0.0848^{***}	(0.0083)
Constant	6.3408^{***}	(0.0020)
Observations	46,534,093	
Article – language FE	Yes	
Date FE	Yes	
Clusters	Article = 52,262; date = 292	

Outcome variable: $\ln(\text{views}+1)$. Estimator: weighted OLS with article×language and date fixed effects.
Weights are proportional to pre-period views (analytic weights).
Robust standard errors in parentheses are two-way clustered by article and date.
Significance: $^{***}p<0.001$, $^{**}p<0.01$, $^{*}p<0.05$.

C.2 Test Period Data Exclusion

As we mentioned in §2, Google began the AIO test on March 22nd, 2024, but officially announced the launch on May 15th, 2024. Given this test period, we drop the corresponding data and run the main-effect estimation as a robustness check. As we can see in Table A4, results are consistent with §3, representing a statistically significant negative effect.

Table A4: Main effect of Google AIO launch on the English articles' views.

	Page views – Without FE		Page views – With FE	
English	1313.303^{***}	(5.334)		
post × English (DiD)	-269.967^{***}	(7.542)	-243.735^{***}	(16.968)
Constant	144.427^{***}	(2.557)	570.297^{***}	(2.072)
Observations	38,086,648		38,086,648	
Article – language FE	No		Yes	
Date FE	No		Yes	
R^2	0.0019		0.9079	

Outcome variable: views.
Columns 1–2: OLS regression of views on English and post×English without fixed effects.
Columns 3–4: HDFE OLS regression of views on an English × Post indicator ("DiD") with article×language and date fixed effects.
Standard errors in parentheses. Columns 3–4 use two-way clustering by article and date (as in the Stata output).
*** $p<0.001$, ** $p<0.01$, * $p<0.05$.

C.3 Shared Articles in Top-1000 Lists

As another robustness check, we change one step in our article collection in §2. Accordingly, we choose articles that appear in English *and* one other language's top-1000 list, at least once (previously, we considered appearing only in one language as the criterion). The new, more limiting rule helps to be more confident about the pre-treatment parallel trend, while keeping high-traffic articles. This results in a final list of 5,187 of unique articles. Although the number of articles is one tenth of the original list, the mean of the total daily pre-treatment page views is around 40 million (more than 40% of the original list). Tables A5 and A6 report the difference-in-difference results and the linear pre-treatment parallel trends test. As we can see, the results are in the same direction as our main results in §3, though showing a larger negative effect.

Table A5: Main effect of Google AIO launch on the English articles' views.

	Page views – Without FE		Page views – With FE	
Treated (English)	6513.447^{***}	(881.3345)		
post × Treated (DiD)	-1407.621^{***}	(51.6169)	-1364.472^{***}	(118.5931)
Constant	256.778^{***}	(42.8859)	1939.642^{***}	(14.9972)
Observations	5,838,612		5,838,612	
Article – language FE	No		Yes	
Date FE	No		Yes	
R^2	0.0052		0.9361	

Outcome variable: daily article–language views.
Robust standard errors (in parentheses) are two-way clustered by article and date.
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Table A6: Pooled linear pre-treatment parallel trend test with unit and day fixed effects

	Page views	
Treated × t (η)	-3.8351	(2.2116)
Constant	2034.827^{***}	(40.3768)
Observations	2,899,956	
Article – language FE	Yes	
Day FE	Yes	
Clusters	Article = 5,187;	date = 146
R^2	0.9115	

Sample restricted to pre-period observations ($\text{Post}_t = 0$). Dependent variable is views. Panel and day fixed effects are included. Two-way cluster-robust standard errors by article and day. Linear time is measured as $t = \text{date} - \text{Oct. 28, 2023}$.
Significance: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

C.4 Additional Topic-Level Analysis

In addition to the topic-level estimations with fixed-effects in Table 5, we provide regression results without fixed effects in Table A7. This helps us get a clearer view of changes in page views compared to the baseline.

As a complement to the topic-by-topic estimates in Table 5, we also estimate heterogeneity within a

Table A7: Topic-specific effect of Google AIO on Wikipedia page views (separate regressions by topic; without fixed effects).

	Culture	Geography	History & Society	STEM
English_l	1606.7370^{***}	683.6938^{***}	1459.6860^{***}	2320.6650^{*}
	(29.9805)	(18.0849)	(70.1971)	(1028.3650)
$\text{English}_l \times \text{Post}_t$	-341.3871^{***}	-140.0190^{***}	-154.9408^{***}	-182.5085^{***}
	(25.6387)	(18.3578)	(30.5132)	(41.4442)
Constant	136.7376^{***}	161.4687^{***}	103.7929^{***}	149.3170^{**}
	(1.9062)	(2.3468)	(4.9752)	(49.6521)
Observations	22,875,713	15,673,940	2,808,950	5,000,138
Article – language FE	No	No	No	No
Date FE	No	No	No	No
Clusters (article and date)	25,542 and 292	19,346 and 292	2,745 and 292	4,408 and 292
R^2	0.0225	0.0085	0.0339	0.0007

Robust standard errors (in parentheses) are two-way clustered by article and date.
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

single pooled regression by interacting the treatment indicator with topic dummies. Let

$$\mathscr{S} = \{\text{Culture}, \text{Geography}, \text{History \& Society}, \text{STEM}\},$$

denote mutually exclusive topic categories, and let $\text{Topic}_i \in \mathscr{S}$ denote the topic of article i. We estimate:

$$\text{views}_{i,l,t} = \alpha_{i,l} + \delta_t + \beta_0\big(\text{English}_l \times \text{Post}_t\big) + \sum_{k \in \mathscr{S} \setminus \{\text{Culture}\}} \beta_k \mathbf{1}\{\text{Topic}_i = k\} \times \big(\text{English}_l \times \text{Post}_t\big) + \varepsilon_{i,l,t}, \tag{A6}$$

where Culture is the omitted category. The coefficient β_0 captures the AIO-associated change in English views for Culture articles relative to controls, and each β_k captures how the effect for topic k differs from Culture; the implied effect for topic k is $\beta_0 + \beta_k$. The specification includes the same article×language and date fixed effects as in the main analysis and uses two-way clustering by article and date (Bertrand et al. 2004; Cameron et al. 2011).

Table A8 reports the results. The baseline (Culture) estimate is negative and precisely estimated, and the interaction terms are positive for Geography, History & Society, and STEM, implying smaller declines in those topics than in Culture. The implied topic-specific effects are closely aligned with the separate-regression estimates in Table 5, providing reassurance that our conclusions about heterogeneity do not hinge on whether topics are estimated jointly or separately.

To assess whether the topic-specific DiD estimates rest on a plausible parallel-trends assumption, we also conduct linear pre-treatment linear trend tests within each topic. We restrict the sample to pre-period observations and estimate

$$\text{views}_{i,l,t} \; = \; \alpha_{i,l}^{(k)} \; + \; \gamma_k t \; + \; \theta_k\big(\text{English}_l \times t\big) \; + \; \varepsilon_{i,l,t}^{(k)}, \tag{A7}$$

where t is a linear time index, $\alpha_{i,l}^{(k)}$ are article–language fixed effects, and θ_k captures any differential linear pre-trend for English relative to the comparison languages within topic k.

Table A8: Heterogeneous effect of Google AIO on Wikipedia page views, by article topic

	Page views	
$\text{English}_l \times \text{Post}_t$ (baseline: Culture)	-319.4724^{***}	(25.1046)
Geography	201.1256^{***}	(29.2991)
History & Society	185.7277^{***}	(41.7964)
STEM	158.0455^{**}	(49.4003)
Constant	571.0033^{***}	(2.4147)
Observations	46,358,741	
Article – language FE	Yes	
Date FE	Yes	
Clusters (article and date)	52,041 and 292	
R^2	0.9166	

Outcome variable: daily article–language views.
Robust standard errors (in parentheses) are two-way clustered by article and date.
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

C.5 Week-Level Analysis

Similar to the day-level data, we run a suite of analyses using the data aggregated on a weekly level.

C.5.1 Week-Level Page Views Plot

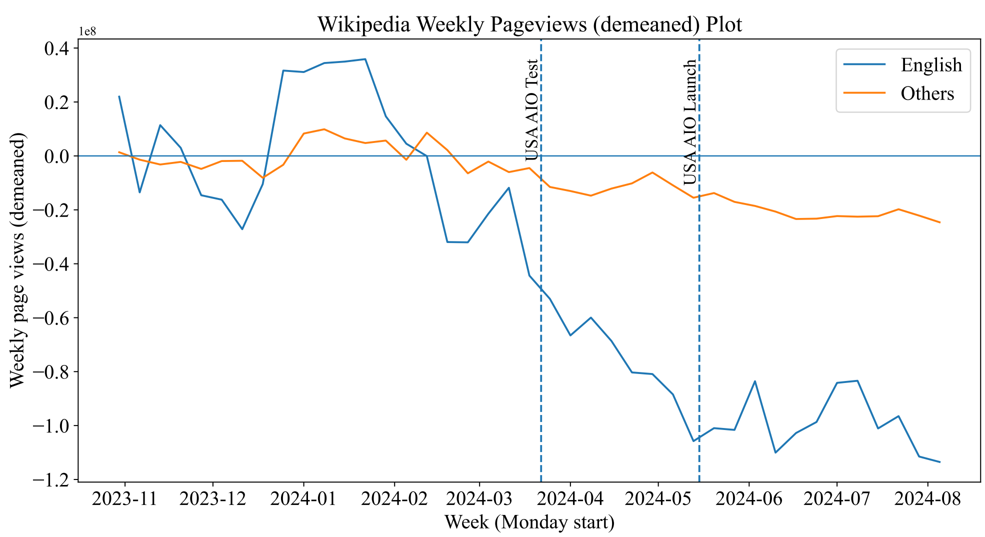

Figure A5: Plot of the weekly page views evolution for English vs. the control group. The vertical lines identify the start of the Google AIO's test and launch periods.

We start our week-level analysis by plotting the page views for English vs. non-English versions of the selected articles. Similar to the daily data, we can observe a growing divide between the two classes of articles after the Google AIO test date.

C.5.2 Week-Level Main Effect

The weekly DiD estimate is negative and highly significant in both the simple specification and the fixed-effects specification. In the preferred model with panel (article×language) FE and week FE, the post×English coefficient is -1503.74 weekly views per article-week.

Table A9: Weekly main effect of Google AIO launch on the English articles' views.

	Page views – Without FE		Page views – With FE	
English	9177.508^{***}	(606.7241)		
post × English (DiD)	-1649.339^{***}	(27.3312)	-1503.747^{***}	(141.0460)
Constant	988.7012^{***}	(43.7363)	3936.008^{***}	(23.3124)
Observations	6,455,239		6,455,239	
Article – language FE	No		Yes	
Week FE	No		Yes	
R^2	0.0020		0.9659	

Outcome variable: views.
Columns 1–2: regression of views on treated (English) and post×treated (DiD) without fixed effects.
Columns 3–4: HDFE regression with panel_id and week fixed effects.
Standard errors in parentheses. Columns 3–4 use two-way clustering by article_id and week (as in the Stata output).
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

C.5.3 Week-Level Parallel Trend

The estimated differential pre-trend slope (English×t) is small and statistically indistinguishable from zero. This provides week-level support for the parallel-trends assumption (at least in the linear trend sense), mirroring the daily analysis.

Table A10: Pooled linear pre-treatment parallel trend test with unit and week fixed effects

	Page views	
English × t (η)	-16.1108	(15.2692)
Constant	4067.6050^{***}	(49.8990)
Observations	3,066,253	
Article – language FE	Yes	
Week FE	Yes	
Clusters	Article = 52,262;	week = 19
R^2	0.9608	

Sample restricted to pre-period observations ($\text{Post}_t = 0$). Dependent variable is views. Unit (article–language) and week fixed effects are included; the main effects of `treated` (English) and t are absorbed by the fixed effects and therefore omitted. Two-way cluster-robust standard errors by article and week. Linear time is measured as t = week start date − Oct 28, 2023.
Significance: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

The event-study coefficients are near zero (with some fluctuations) in the pre-period, and then turn consistently negative after the event week, remaining negative thereafter. This matches the daily event-study takeaway: the divergence appears around the rollout window rather than being a pre-existing trend.

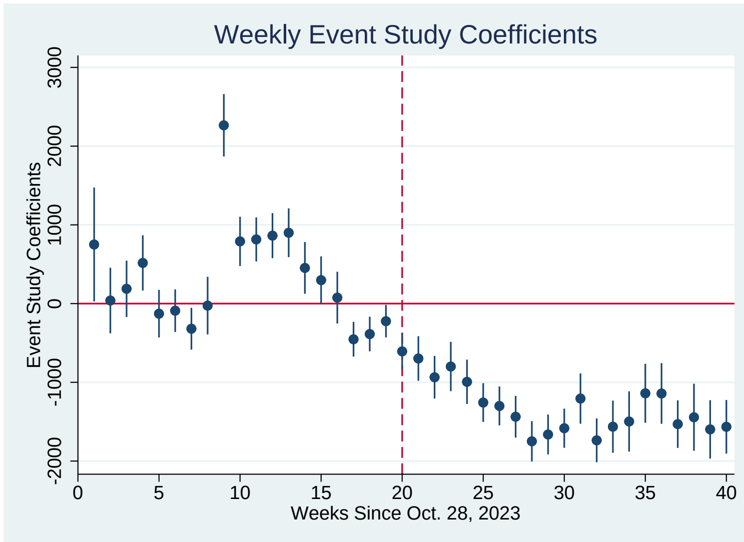

Figure A6: Event study coefficients for the weekly data (start: Oct. 28th, 2023 - end: Aug. 14th, 2024). The vertical dashed line represents the event week, including March 22nd, 2024

We present the robust confidence intervals of the HonestDiD estimation in Figure A7. Similar to the daily results in Figure A4, for modest departures, the intervals remain centered well below zero at 95% confidence. This reinforces that the negative weekly effect is not fragile to violations of the identifying assumption, especially for the case of linear violation of the pre-treatment parallel trends assumption.

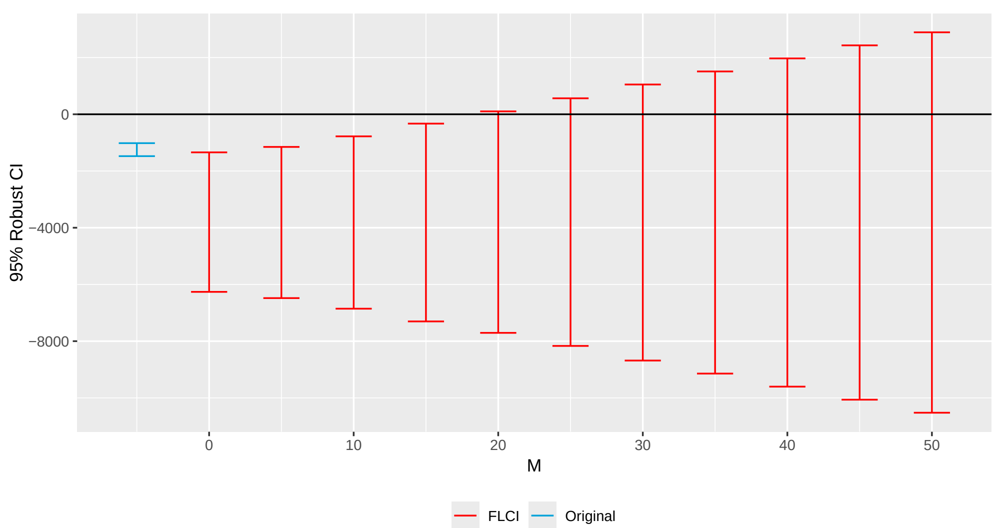

Figure A7: Weekly HonestDiD estimation for page views.

References

D. Autor, D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen. The fall of the labor share and the rise of superstar firms. *The Quarterly journal of economics*, 135(2):645–709, 2020.

M. Bertrand, E. Duflo, and S. Mullainathan. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275, 2004.

A. C. Cameron, J. B. Gelbach, and D. L. Miller. Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249, 2011.

B. McConnell. Can't see the forest for the logs: on the perils of using difference-in-differences with a log-dependent variable, 2024.

A. Rambachan and J. Roth. A more credible approach to parallel trends. *Review of Economic Studies*, 90(5):2555–2591, 2023.

G. Solon, S. J. Haider, and J. M. Wooldridge. What are we weighting for? *Journal of Human resources*, 50(2):301–316, 2015.

J. M. Wooldridge. *Econometric analysis of cross section and panel data*. MIT press, 2010.