Can Advertising Fund Free AI? Evidence from ChatGPT's International Advertising Rollout

Maximilian Kaiser* Christian Schulze†

October 2026

Updates, teaching slides, and press kit:
https://ecommerceresearch.org/chatgpt-ads/

Abstract

Advertising could keep artificial intelligence (AI) assistants free and make ChatGPT a Google Search competitor if advertisements do not devalue the product and justify their price. This study tests both conditions during ChatGPT's staggered international rollout, with the European Union untreated, using data from 1,898 online shops selling $16 billion annually. The introduction of advertisements moderately reduces the volume but not the conversion of organic ChatGPT referrals, a proxy for product value. However, paid ChatGPT visits perform worst among major channels, offering less than a sixth of the value for money of paid Google Search. Current performance is not sufficient to finance free access costs. Only more valuable advertisements can bridge this gap, potentially making the product less useful. A Google competitor is not yet in sight.

Keywords:
generative AI, advertising, AI monetization, economics of AI, digital platforms, digital economy, marketing performance, channel analytics

*Vienna University of Economics and Business, Grips Intelligence, research@maximiliankaiser.org; LinkedIn
†Frankfurt School of Finance & Management; email@christian-schulze.de; LinkedIn

Acknowledgements. The authors thank Grips Intelligence for providing the extensive data used in this study free of charge. Grips Intelligence had no role in formulating the research questions, designing the study, selecting analytical methods, interpreting results, or drafting the manuscript.

1 The Advertising Bargain

ChatGPT, OpenAI's AI assistant, serves more than 1 billion users each week (OpenAI 2026g), yet roughly 95% pay nothing (O'Brien 2026, Malik 2026). To finance that free access, OpenAI turned to the very business model its CEO, Sam Altman, had deemed "a last resort for us" less than two years earlier: advertising (Altman 2024). Ads launched in the United States in February 2026, followed by a staggered rollout to eight additional countries (OpenAI 2026d,h) within our observation period. Less than 200 days after launch, OpenAI was already reporting $1 billion in projected annualized advertising revenue (OpenAI 2026g).

OpenAI's advertising bet faces two revenue hurdles of very different size. The first is covering the cost of free access. The announced $1 billion corresponds to about $1 per free user per year, a fast start for a new advertising product but not yet enough to cover free users' inference costs, which would require roughly $5 according to OpenAI's own figures (Muppidi and Palazzolo 2026). The second is the growth investors expect: OpenAI's projections anticipate $102 billion in annual advertising revenue by 2030 (Muppidi 2026), about $40 for each of the 2.52 billion free users it expects by then (Fried 2026, Muppidi 2025).

Whether advertising can meet these revenue expectations matters beyond OpenAI's planned initial public offering (Wall Street Journal 2026). The remedies a United States court imposed in the antitrust case against Google's search monopoly rest partly on the expectation that providers of AI assistants will become advertising-financed competitors for Google's commercial queries and shift the power balance in the digital economy (United States v. Google LLC 2025, slip op. at 1–2, 45, 128). On a societal level, the outcome will help determine whether advertising can do for frontier AI assistants what it has done for many services of the internet era: keep access free.

The ad-supported business model rests on a two-sided bargain (Moshary 2025). **Condition A requires a valuable product**: advertising must not substantially diminish the value of the service to users. **Condition B requires valuable ads**: advertisements must deliver enough value to advertisers to support the prices the platform needs to charge.

Free access to AI assistants is mainly a test of Condition B, as advertisers must find enough

value in the ads to pay about $5 per free user per year. ChatGPT as a viable business and competitor to Google also tests Condition A, because it presumes that users stay with the ad-supported product; a drop in total users would mean even more ambitious ad revenue targets for each user who stays. Google Search has maintained this bargain for more than two decades (Ghose and Yang 2009, Blake, Nosko, and Tadelis 2015, Simonov, Nosko, and Rao 2018). Whether AI assistants can do the same is unknown.

Condition A may be particularly difficult for ChatGPT. Search engines present lists in which advertisements are visibly separated from organic (i.e., unpaid) results. ChatGPT instead produces a curated conversational response that can resemble personalized advice. If advertising causes users to interpret organic recommendations as commercially influenced, they may become less likely to ask ChatGPT for shopping guidance, click on those recommendations, or trust its advice enough to make a purchase (Hermann, Puntoni, and Schweidel 2026). In experiments, users often fail to notice promotion woven into a chat answer, but discount the answer once they learn of it (Salvi, Cuevas, and Horta Ribeiro 2026, Tang et al. 2025). OpenAI's initial ad format appears designed to limit this risk: sponsored recommendations are labeled and displayed separately from the organic answer (Figure 1). This separation may protect the organic product, but it may also reduce the effectiveness of the advertisements (Aribarg and Schwartz 2020).

Condition B depends on whether ChatGPT ads justify their cost to advertisers. Ads create value directly, when a consumer clicks and buys, or indirectly, when exposure prompts a later purchase through another channel (Li and Kannan 2014, Dinner, Van Heerde, and Neslin 2014). These combined returns must cover the cost of advertising and be competitive with those from established channels such as Google Search.

ChatGPT's staggered rollout across nine countries, with the European Union (EU) untreated, and the traffic and sales data of 1,898 websites let us test both conditions from outside the platform. We observe how the free organic recommendations fared once ads arrived in a market, a proxy for product value, and what paid clicks were worth to the websites that bought them, a direct measure of ad value. Both are normally hidden from view: OpenAI does not publish them, and advertisers

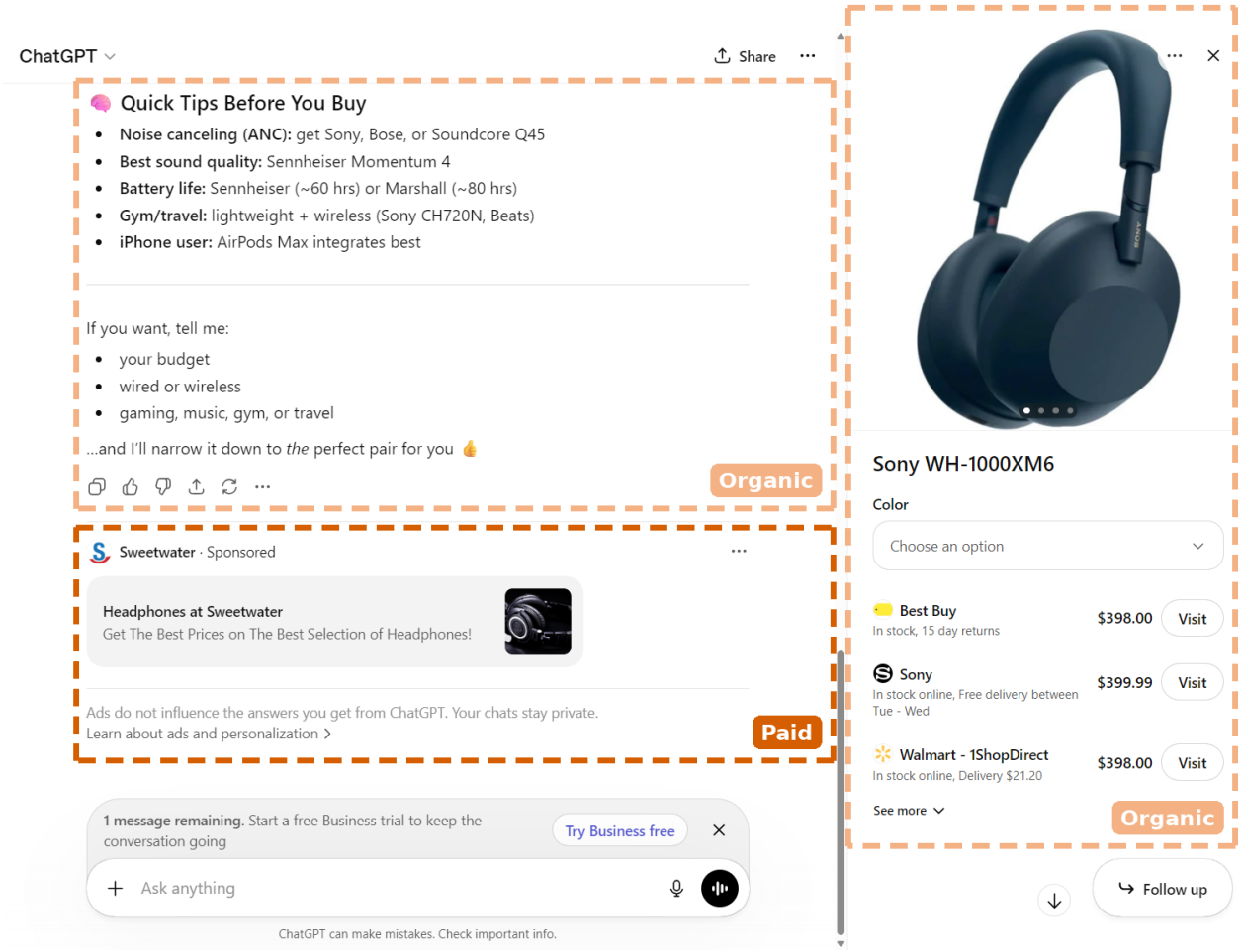


Figure 1: **Ads displayed in ChatGPT (March 2026).** A sponsored product recommendation appears as a labeled card below the conversational answer for free-tier and Go-tier users, visually separated from the organic response. Dashed frames mark organic content (the conversational answer and the product panel with its merchant links) versus paid content (the sponsored card). Screenshot by the authors, March 2026, reproduced to document the ad format.

rarely disclose what their clicks returned.

2 Rollout, Data, and Outcomes

Ads launched in the United States on February 9, 2026; in Australia, Canada, and New Zealand on April 16; in the United Kingdom on June 4; in Japan and Korea on June 19; and in Brazil and Mexico on August 4 and 6, respectively (OpenAI 2026d, Korea JoongAng Daily 2026, Japan Times 2026, Sacchitiello 2026, Rodrigo 2026). The EU remained untreated throughout: OpenAI

announced its inclusion on August 15 (PPC Land 2026), and EU ads went live only after our window closed, on August 24 (OpenAI 2026c, Scanlon 2026). The opportunity for companies to buy ads expanded from an invitation-only launch to self-service access without a minimum spend in May 2026 and to conversion-optimized campaigns in June (Donovan 2026, Adegbola 2026).

Treatment is market-level ad availability. Unclicked ad impressions are not observed, so estimates measure the effect of making ads available in a market, not the effect on an individual known to have seen one. Ads appear on the free and low-priced Go tiers only (OpenAI 2026h). Independent sampling found ads on about one quarter of commercial US prompts in July (Deda 2026a).

The data covers 60 weeks, from June 29, 2025, to August 22, 2026, and nearly 10 billion visitor sessions, aggregated by website, visitor country, device, marketing channel, and week, with transactions, revenue, and engagement measures. The organic referral panel used to analyze Condition A contains 1,898 internationally active websites. The paid referral panel used to analyze Condition B contains 387 websites with referrals from paid ChatGPT ads; 102 of them, the financial sample, have complete financial tracking, and for 31 of the 102, the cost sample, linked Google Ads records provide advertising costs (Web Appendices A to C). Referral-source information lets us distinguish the 22.4 million organic ChatGPT referrals and the 0.45 million paid ChatGPT referrals from established channels such as organic and paid search. The data captures the last click that brought a visitor to a website; it does not reveal prompts, recommendation sets, ad impressions, or subscription tiers.

Organic ChatGPT accounted for 0.23% of all visits to the organic referral panel over the full window and 0.40% over the window's final four weeks. Over the full window it converted at 1.99%; before the rollout, its conversion rate approximately matched organic search's (Figure 2), meaning it had closed the gap previously documented for the channel's first year (Kaiser and Schulze 2026). Ad delivery intensified as well: on the paid referral panel, paid ChatGPT sessions accounted for 0.007% of all visits over the full window and 0.030% over its final four weeks (Table C.1).

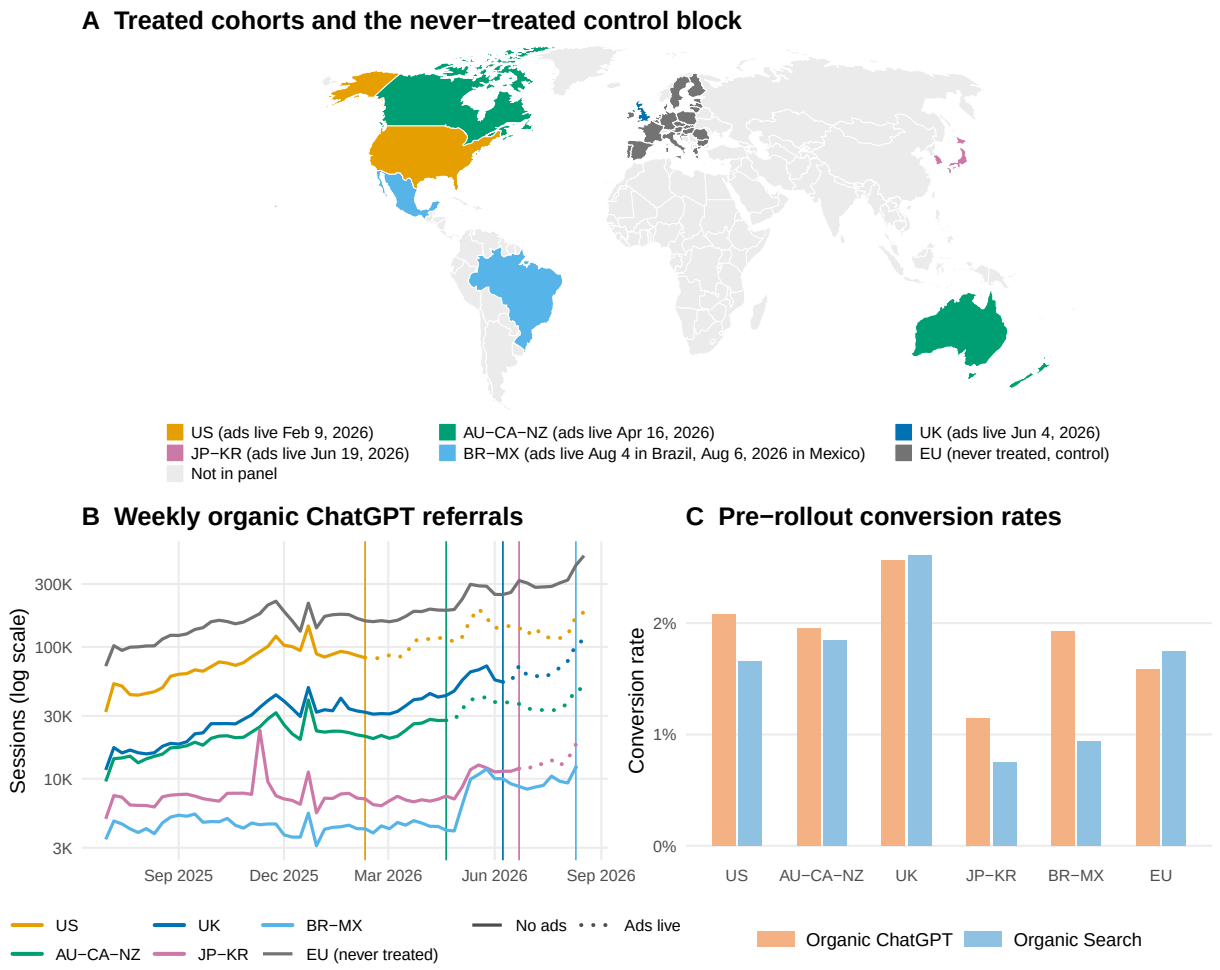


Figure 2: **The rollout and the organic ChatGPT channel.** (A) Map of treated cohorts and the untreated EU control block. (B) Weekly organic ChatGPT referral sessions to the organic referral panel by region group, on a log scale; vertical lines mark launch weeks and the segments after each launch are dotted. (C) Conversion rates of organic ChatGPT and organic search referrals over each region's pre-rollout period; for the EU, the period before February 9, 2026.

3 Condition A: Moderately Fewer Organic Referrals, Unchanged Conversion

Condition A asks whether ads compromised the value of ChatGPT to its users. We observe that value indirectly, in the traffic and sales data of online shops, so damage would appear on either of two margins: fewer referred visitors, or visitors who buy less often. We estimate both with the same design. Within each website, market, device, and week, we compare organic ChatGPT referrals

with Google's organic search. Factors that affect an online shop or market as a whole thus cancel out: directly in the conversion comparison, and through country-, channel-, and device-by-week effects in the volume model. We evaluate every rollout cohort against the never-treated EU, with the Sun–Abraham estimator keeping already-treated cohorts from serving as controls (Sun and Abraham 2021). Same-week comparisons across countries remove common shocks from model releases, competing products, and publicity. Volume is modeled as session counts with a Poisson specification, so that weeks without ChatGPT referrals count as zeros rather than dropping out; conversion is modeled as the rate difference between the two channels, weighted at pre-launch levels. Estimates are expressed against the pooled pre-launch average of the cohorts rather than a single reference week, and standard errors are clustered by website throughout (Web Appendices D and E).

After the introduction of ads, organic referral volume fell by 6 to 8%. Organic referral volume changed by −5.9%, averaged over the first 13 post-launch weeks, with a 95% interval from −12.6% to +1.2% ($p = .103$; Figure 3A; Table F.1). Declines at least this large arose in 10 of 100 placebo rollouts among untreated European countries (Table F.2), so chance alone produces a decline of this size about one time in ten. Both the interval and the placebo test indicate a decline in volume after the introduction of ads, but neither establishes it with confidence.

However, these estimates would understate the decline if total ChatGPT usage in the treated markets had grown faster than in Europe without ads. OpenAI's published regional adoption index suggests exactly that: North America outgrew Europe by about 6% in the year before the launch and by about 12% per year over our window through June 2026 (OpenAI 2026e). Including the 12% (6%) differential in the model moves the estimate to −8.4% (−7.2%), with 95% intervals from −14.9% to −1.5% (−13.8% to −0.2%) and $p = .019$ ($p = .045$) (Web Appendix G). Declines of that size arose in 4 (8) of the 100 placebo rollouts, and the interval excludes zero once the growth differential exceeds about 5% per year (Tables F.1 and F.2). In summary, the analyses suggest that the introduction of ads moderately reduced organic referrals, by about 6 to 8% relative to untreated

markets.

Visitors who still arrived converted as before. Organic ChatGPT conversion changed by −0.02 percentage points from a 2.21% pre-launch baseline, with a standard error of 0.05 percentage points ($p = .648$; Figure 3C; Table F.3), a decline of about 1% whose interval comfortably includes zero. Declines at least this large arose in 187 of 500 placebo rollouts (Table F.2), so chance alone produces a decline of this size about two times in five; the 95% interval and the placebo distribution rule out an immediate average decline larger than 6 and 10% of baseline, respectively (Table H.1). Changes in the customer base have little room to explain this stability: because referral volume fell at most moderately, conversion is measured on nearly the same visitors as before, not on a small, unusually determined remainder.

Per eligible user, the bounds are wider. Users on the Plus and Pro tiers see no ads. These paying users are only about 5% of users (O'Brien 2026) but use ChatGPT three (Plus) to eleven (Pro) times as much as free users (Friar 2026); if they average five (ten) times as much, users eligible for ads account for about 79% (66%) of referral sessions. Dividing the market-level bounds by that share gives the bounds for eligible users: at most a 7 to 12% (9 to 15%) decline in conversion and 16 to 19% (19 to 23%) in volume (Figure 3, B and D; Table H.1; Web Appendix H).

Pre-trends and samples leave the estimates in place. Treated and control paths move similarly before treatment for both outcomes (joint tests, $p = .458$ for conversion and $p = .109$ for volume; Table I.1). The cohorts' own volume paths reject in four of five cohorts (United States, $p < .001$), and the pooled test itself rejects in 49 of 100 placebo rollouts without any ads (Table F.2); the test on the United Kingdom's conversion path alone gives $p < .001$, but removing that country moves the pooled estimate by only a hundredth of a percentage point. These tests have little power against a smooth growth differential of the size discussed above. Dropping the United States cohort leaves volume at −5.8% and conversion at −0.003 percentage points; dropping any other single country moves conversion by at most 0.02 percentage points; and dropping the largest websites, individually

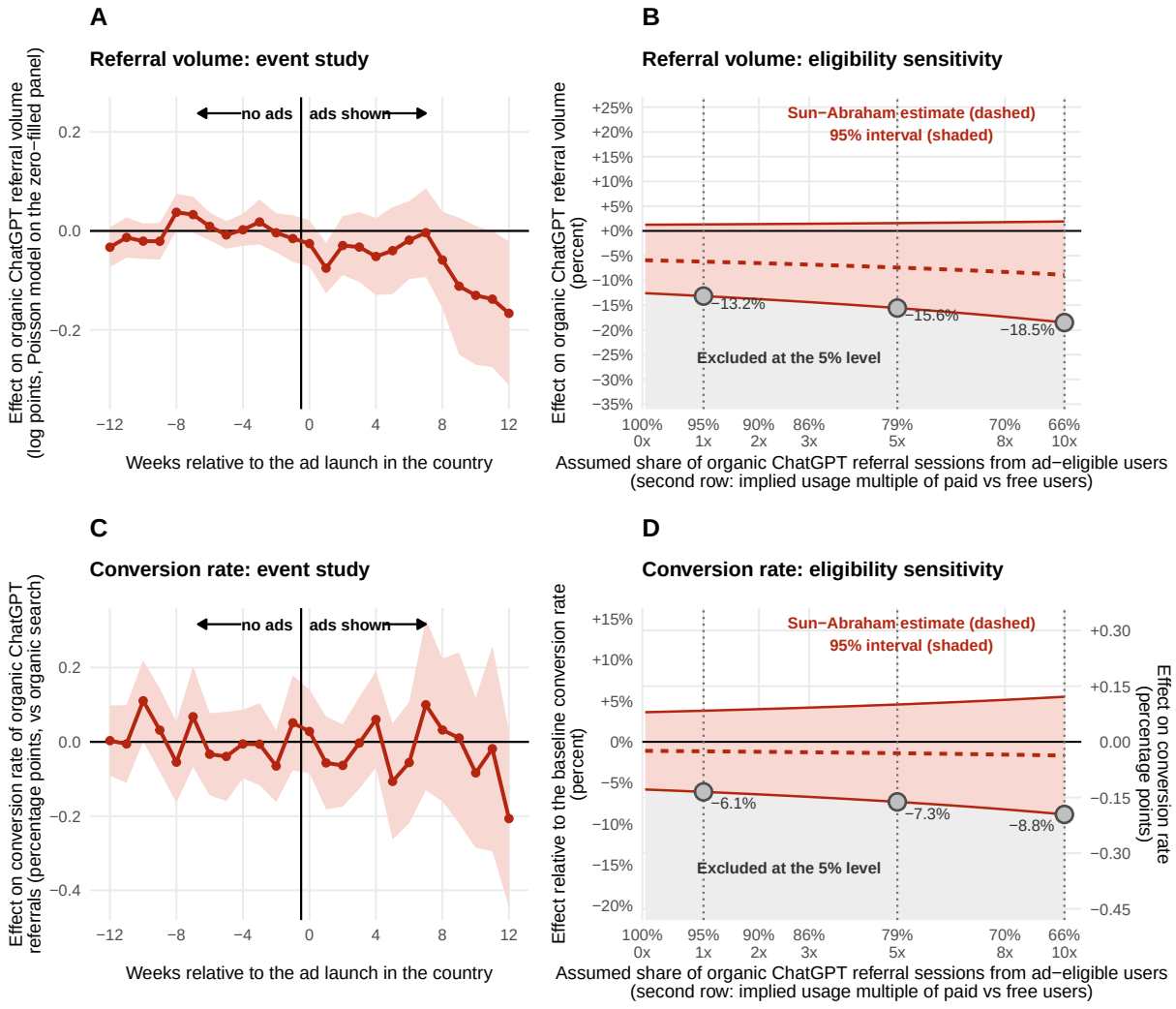


Figure 3: **Organic performance at the ad launches and its eligibility sensitivity.** (A) Organic ChatGPT referral volume, a Poisson model of session counts, and (C) conversion rate, weighted at pre-launch levels: Sun–Abraham event studies, re-centered on the pooled pre-launch average, with website-clustered 95% intervals. (B, D) The same estimates as the assumed share of referral sessions from ad-eligible users falls from 100% to 66%, equivalently as paying users are assumed to use ChatGPT up to ten times as much as free users; markers give the ruled-out decline at one, five, and ten times the free users' usage; the conversion panel carries percentage points on its right axis, and its marks are shares of the 2.21% pre-launch baseline.

or the ten largest together, keeps volume between −5.1% and −8.8% (Tables F.1 and F.3). Because observed ad activity concentrated in the United States and United Kingdom, the pooled estimates describe the early rollout rather than five equally intense rollout cohorts.

Advertising has not substantially diminished the value of the product. In the early rollout, the observed damage was moderate: ads cost ChatGPT about 6 to 8% of its organic referrals, and the visitors who still arrived converted as before. If conversion is read as a proxy for trust in ChatGPT's advice, the introduction of ads left that trust intact.

4 Condition B: Advertiser Value Remains Unproven

Condition B asks whether the ads are worth their price. We continue with our paid referral panel of 387 advertiser websites, 102 with complete financial tracking. Advertiser entry is self-chosen, not randomized, so this evidence is descriptive rather than causal. A channel is good or bad only relative to what an advertiser could buy instead. We therefore benchmark as an advertiser would, against paid Google Search: the channel these same websites already buy for similar commercial intent, at prices the market has tested for decades.

Paid ChatGPT visits generated less value for advertisers than paid search. In US traffic, where ad delivery concentrated, a paid ChatGPT visit generates $0.89 in revenue per session against $2.45 for a paid Google Search visit, a 64% shortfall. Within the same shops, both margins contribute: paid ChatGPT visitors buy less often (conversion rates of 0.37% versus 1.17%; adjusted difference -1.87 log-odds, $p = .005$; Table J.1), and their orders are smaller, by $74 ($p < .001$; both adjusted for website, device, and week; Web Appendix J). The shortfall is robust: among the 87 of the 102 advertisers buying both paid channels in US traffic, the dual-channel advertisers, 97% earn less per ChatGPT visit and 98% convert worse. Most websites make no sale from ChatGPT ads at all: 84% record no paid ChatGPT transaction despite 18,726 paid ChatGPT sessions, and the remaining 14 websites carry all 570 transactions (Table K.1).

Paid ChatGPT trails direct traffic on every metric. Figure 4 places both ChatGPT channels among the websites' established channels (Kaiser and Schulze 2026), each measured against direct traffic, adjusted for website-by-device and device-by-week effects. Paid ChatGPT is the only channel

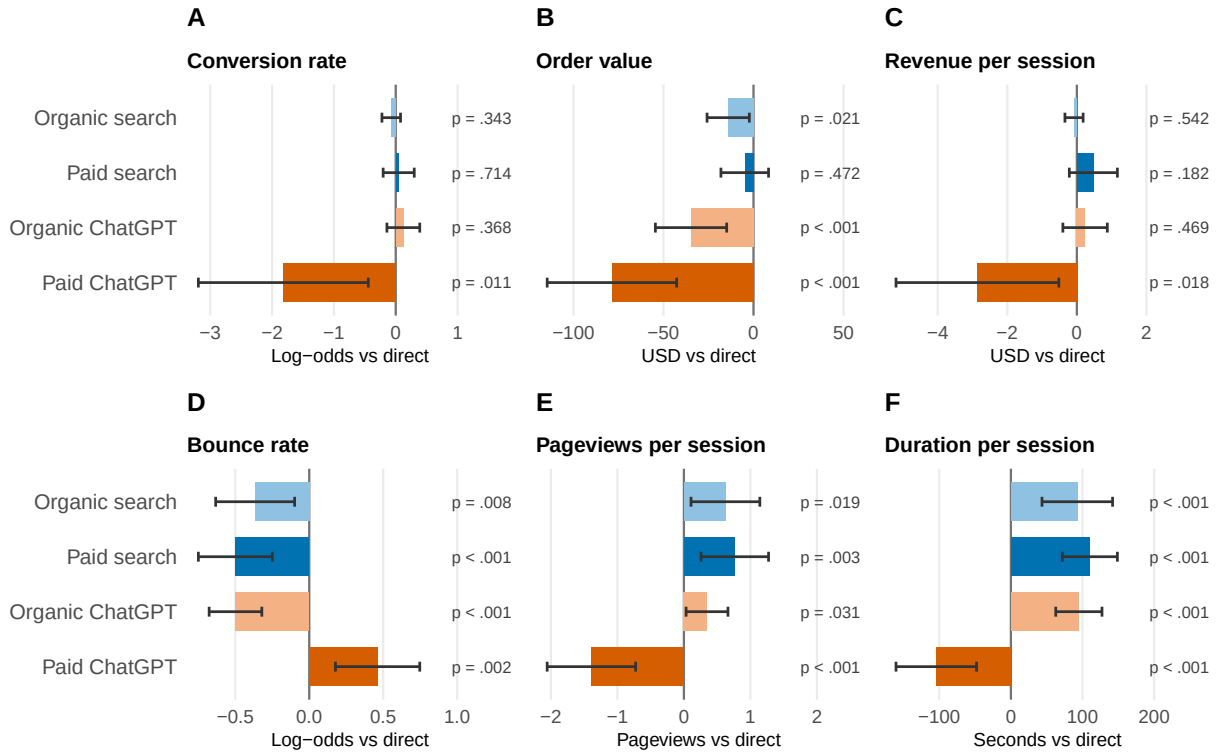


Figure 4: **Performance of paid ChatGPT versus established channels.** (A) Conversion rate (log-odds), (B) average order value (USD), and (C) revenue per session (USD) for the 102 advertisers of the financial sample. (D) Bounce rate (log-odds), (E) pageviews per session, and (F) duration per session (seconds) for the paid referral panel of 387 websites. All estimates are differences versus direct traffic, net of website, device and week effects (Web Appendix J). Whiskers show 95% intervals from website-clustered standard errors; the exact two-sided p-value of each difference is printed at the right of its row (estimates and standard errors in Tables J.1 and K.2).

that falls short of direct traffic on all six metrics. Among the 102 advertisers of the financial sample, its visitors convert far less often (log-odds −1.82, $p = .011$), place orders \$79 smaller ($p < .001$), and generate \$2.86 less revenue per session ($p = .018$). Google's paid search, by contrast, does not differ detectably from direct traffic on any of the three ($p = .714$, .472, and .182; panels A to C). On the larger paid referral panel, paid ChatGPT visitors show weaker engagement. They bounce more often (log-odds +0.46, $p = .002$), view 1.4 fewer pages ($p < .001$), and stay 104 seconds less ($p < .001$) than direct visitors. Paid search, organic search, and organic ChatGPT visitors all bounce less, view more pages, and stay longer (panels D to F; largest $p = .031$).

Actual prices are more than six times the parity price. The search benchmark yields a parity price, i.e., the price at which a ChatGPT click returns as much per dollar as a search click. A paid-search click returns $2.45 on average and costs the US-focused advertisers of our cost sample a median of $0.72. A paid ChatGPT visit returns $0.89 and is therefore worth about $0.26 at the same rate of return (between $0.05 and $0.57 when advertisers are resampled; Web Appendix L).

Actual market prices are much higher: the largest publicly known account paid $1.72 per ChatGPT click (Opascope 2026), more than six times the parity price. The comparison is conservative, as that actual price sits toward the low end of what practitioners report and below OpenAI's recommended starting bid of $3 to $5 per click (Deda 2026a, OpenAI 2026b). Two caveats apply: the parity price assumes similar profit margins in both channels, and it counts only the tracked visit. However, untracked later business raises ChatGPT's parity price substantially *only* if it adds proportionally more to a ChatGPT click's return than to a search click's. Matched comparisons around each advertiser's first paid week, an observational stand-in for the experimental lift measurement of Johnson, Lewis, and Nubbemeyer (2017), find no detectable lift in the channels the campaigns did not buy (Table N.1).

No single advertiser drives the shortfall. Dropping each of the 102 advertisers in turn leaves the conversion and revenue deficits in place in every case; across the 102 re-estimates, the largest p-value is .015 for conversion and .007 for revenue per session. The largest advertiser, with 66% of all paid ChatGPT purchases, makes the channel look better rather than worse: dropping it deepens both deficits. Nor is the share of advertisers without a single paid ChatGPT purchase a small-numbers artifact: had their ChatGPT visits converted at their own paid-search rates, 42 of the 87 dual-channel advertisers would show no purchase, whereas 73 do (Table K.1). Three limitations remain: campaigns that never produced a click are invisible, first-click dates can lag true campaign starts, and the cross-channel intervals rule out only large indirect effects.

Advertising has not yet delivered enough value to advertisers to justify its current price. In the first 28 weeks of advertising, the shortfall was large: a paid ChatGPT visit returned about a

third of a paid-search visit's value at about 2.4 times its price; taken together, ChatGPT ads offer more than six times worse value for money than Google Search ads. The best-known public success story (Opascope 2026) is the exception, not the rule: in our data, only one (small) advertiser in 102 matches the public account's conversion and revenue per visit (Table N.2).

5 Discussion

The introduction of ads in ChatGPT cost the product a moderate share of its organic referrals but showed no detectable average change in conversion rates (Condition A). The ads themselves return far less than they cost: at current performance, the parity price is $0.26 per click (Condition B).

What would it take for ads of that value to pay for free access to AI assistants? A free user sends about seven prompts a day, in one or more conversations (Friar 2026), a fifth of which carry commercial intent (Adams 2026): some 511 commercial prompts a year. Today about one in four of these carries an ad (Deda 2026a). At the parity price (Condition B) and a click-through rate of 0.5%, the best third-party estimate (Mordecai 2026), covering the $5 cost of free access would take more than seven ads on each commercial prompt. The $40 per free user that would meet OpenAI's 2030 projections and make ChatGPT a viable business and competitor to Google would take about 60 ads per commercial prompt (Table P.1). That is not a realistic scenario, so the value per click must improve.

Raising the value of a click, however, works against volume (Yan, Miller, and Skiera 2022). One reading of our results is that the clicks are worth little because the sponsored cards sit apart from the answer, where ChatGPT's value lies, and that the same separation is why the organic product was largely spared in our data (Condition A). Moving ads into the answer could raise their value: in a preregistered experiment, a model arguing for a sponsored product won a majority of choices, largely unnoticed even when labeled, whereas a neutrally described sponsored item moved choices no more than a search placement (Salvi, Cuevas, and Horta Ribeiro 2026). But users who become aware of ads embedded in an answer feel manipulated and voice concerns about bias (Tang

et al. 2025), and unlike search, ChatGPT has little room to place ads without altering the product's logic (Hermann, Puntoni, and Schweidel 2026). Integrating ads into the answer would then buy value per click with users' trust, potentially improving Condition B while risking Condition A.

For society, our analyses of OpenAI's ad rollout leave open whether advertising can keep frontier AI assistants free. The projected revenue covers only one fifth of the cost of free access, and far less if prices settle at what the clicks are worth in our data; only more valuable ads can close that gap. For the industry, the advertising-financed competitor to Google that the search-case remedies count on is not yet in sight (United States v. Google LLC 2025, slip op. at 1–2, 45, 128), and more advertisers and clicks alone will not bring it, because unfavorable unit economics do not scale away. The central challenge for AI assistants is whether advertising can become valuable enough to finance the product without making the product less useful.

Funding, Author Affiliations and Competing Interests

Funding. This research was supported by Grips Intelligence through the provision of data and computing resources. No direct financial support was provided to the authors.

Author Affiliations and Competing Interests. Maximilian Kaiser was previously employed by Grips Intelligence and now serves as its scientific advisor; Grips Intelligence is the provider of the data analyzed in this study. The data was provided free of charge under a research agreement whose only substantive restriction is that the identities of the panel websites remain confidential; the agreement grants Grips Intelligence no right to review, approve, or delay publication, and imposes no restriction on the analyses performed, the results reported, or their interpretation. Grips Intelligence sells market-intelligence services on e-commerce traffic generally; it has no commercial relationship with OpenAI, does not sell ChatGPT advertising or advertising-measurement services tied to it, and has no financial stake in whether ChatGPT advertising is found to be effective or ineffective. Beyond Maximilian Kaiser's salary and virtual company shares, the authors received no payment, equity, or other consideration from Grips Intelligence. The company had no role in formulating the research questions, designing the study, selecting analytical methods, interpreting results, or drafting the manuscript. Christian Schulze declares no competing interests.

References

Adams, Peter (2026), "What Defines ChatGPT Ads? 'Super Intentional' Users, OpenAI Tells Cannes," Marketing Dive (June 23), https://www.marketingdive.com/news/how-openai-positioning-chatgpt-ads-cannes-lions-debut/823476/.

Adegbola, Anu (2026), "OpenAI Confirms Conversion-Focused Ads Are Coming to ChatGPT," Search Engine Land (May 27), https://searchengineland.com/openai-confirms-conversion-focused-ads-are-coming-to-chatgpt-478843.

Altman, Sam (2024), "A Fireside Chat with Sam Altman, OpenAI CEO, at Harvard University," Harvard Business School, YouTube (talk of May 1, 2024; posted October 16), https://www.youtube.com/watch?v=FVRHTWWEIz4&t=2240.

Aribarg, Anocha and Eric M. Schwartz (2020), "Native Advertising in Online News: Trade-Offs Among Clicks, Brand Recognition, and Website Trustworthiness," *Journal of Marketing Research*, 57 (1), 20–34.

Blake, Thomas, Chris Nosko, and Steven Tadelis (2015), "Consumer Heterogeneity and Paid Search Effectiveness: A Large-Scale Field Experiment," *Econometrica*, 83 (1), 155–174.

CNBC (2026a), "CNBC Transcript: OpenAI CFO Sarah Friar Speaks with CNBC's Jim Cramer on 'Mad Money' Today," CNBC (September 15), https://www.cnbc.com/2026/09/15/cnbc-transcript-openai-cfo-sarah-friar-speaks-with-cnbcs-jim-cramer-on-mad-money-today.html.

CNBC (2026b), "OpenAI Revamps Shopping Experience in ChatGPT After Struggling with Instant Checkout Offering," CNBC (March 24), https://www.cnbc.com/2026/03/24/openai-revamps-shopping-experience-in-chatgpt-after-instant-checkout.html.

Deda, Yulia (2026a), "ChatGPT Shows Ads for 1 in 4 Commercial Prompts," SE Ranking Blog (August 10), https://seranking.com/blog/chatgpt-ads-study/.

Deda, Yulia (2026b), "How to Advertise on ChatGPT: A Step-by-Step Guide," SE Ranking Blog (September 10), https://seranking.com/blog/how-to-advertise-on-chatgpt/.

Dinner, Isaac M., Harald J. Van Heerde, and Scott A. Neslin (2014), "Driving Online and Offline Sales:

The Cross-Channel Effects of Traditional, Online Display, and Paid Search Advertising," *Journal of Marketing Research*, 51 (5), 527–545.

Donovan, Ben (2026), "Rufus, ChatGPT, and the Agentic Ad Convergence," Marketplace Pulse (May 7), https://www.marketplacepulse.com/articles/rufus-chatgpt-and-the-agentic-ad-convergence.

Fletcher, Ashley (2026), "Google AI Overviews & ChatGPT Ads: Paid Search Impact and Protecting Visibility," Adthena Blog (May 1), https://www.adthena.com/resources/blog/google-ai-overviews-paid-search-impact/.

Friar, Sarah (2026), "OpenAI CFO Sarah Friar on IPO, AI Rivalries, New Device, and Spending $100B+ on Compute," All-In Podcast (June 2), https://youtu.be/TjrShuj_Zsg.

Fried, Ina (2026), "Scoop: OpenAI Projects $100 Billion in Ad Revenue by 2030," Axios (April 9), https://www.axios.com/2026/04/09/openai-100-billion-in-ad-revenue.

Ghose, Anindya and Sha Yang (2009), "An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets," *Management Science*, 55 (10), 1605–1622.

Hermann, Erik, Stefano Puntoni, and David A. Schweidel (2026), "From Advice to Advertising: Conversational AI Advertising, Platform Monetization, and Consumer Welfare," SSRN working paper (posted October 21, 2025; revised June 6, 2026), https://doi.org/10.2139/ssrn.5634270.

Horwitz, Joel (2026), "ChatGPT Ads: What 130,000 Real Impressions Taught Us in 9 Days," Synter, Research & Articles (June 11), https://synterai.com/blog/chatgpt-ads-real-data.

Japan Times (2026), "OpenAI to Introduce Ads to ChatGPT in Japan," The Japan Times (June 19), https://www.japantimes.co.jp/business/2026/06/19/companies/openai-chatgpt-advertisements/.

Johnson, Garrett A., Randall A. Lewis, and Elmar I. Nubbemeyer (2017), "Ghost Ads: Improving the Economics of Measuring Online Ad Effectiveness," *Journal of Marketing Research*, 54 (6), 867–884.

Kaiser, Maximilian and Christian Schulze (2026), "Frontiers: ChatGPT Referrals to E-Commerce Websites: How Do LLMs Compare Against Traditional Channels?," *Marketing Science*, 45 (4), 699–715.

Korea JoongAng Daily (2026), "OpenAI Starts Showing Ads to Lower-Tier ChatGPT Users in Korea," Korea

JoongAng Daily (June 19), https://www.koreajoongangdaily.com/business/openai-starts-showing-ads-to-lowertier-chatgpt-users-in-korea/12731460.

Korea Times (2025), "Kakao Launches ChatGPT Service on KakaoTalk," The Korea Times (October 28), https://www.koreatimes.co.kr/business/companies/20251028/kakao-launches-chatgpt-service-on-kakaotalk.

Li, Hongshuang and P. K. Kannan (2014), "Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment," *Journal of Marketing Research*, 51 (1), 40–56.

Malik, Aisha (2026), "ChatGPT Reaches 900M Weekly Active Users," TechCrunch (February 27), https://techcrunch.com/2026/02/27/chatgpt-reaches-900m-weekly-active-users/.

Mordecai, Darrell (2026), "How ChatGPT Ads Are Reshaping Paid Search," Similarweb AI Search Blog (September 3), https://aisearch.similarweb.com/blog/chatgpt-ads-reshaping-paid-search/.

Moshary, Sarah (2025), "Does Sponsored Search Advertising Augment Organic Search? Evidence from an E-Commerce Platform," *Management Science*, 71 (11), 9687–9709.

Muppidi, Sri (2025), "OpenAI Forecasts Nearly as Many ChatGPT Subscribers as Spotify by 2030," The Information (November 25), https://www.theinformation.com/articles/openai-projected-least-220-million-people-will-pay-chatgpt-2030.

Muppidi, Sri (2026), "OpenAI Forecasts Advertising to Hit $102 Billion by 2030," The Information (April 9), https://www.theinformation.com/articles/openai-forecasts-advertising-hit-102-billion-2030.

Muppidi, Sri and Stephanie Palazzolo (2026), "OpenAI Boosts Revenue Forecasts, Predicts $111 Billion More Cash Burn Through 2030," The Information (February 20), https://www.theinformation.com/articles/openai-boost-revenue-forecasts-predicts-112-billion-cash-burn-2030.

O'Brien, Matt (2026), "ChatGPT Maker OpenAI Shifts Its Focus to Business Users amid Anthropic Pressure," AP News (April 16), https://apnews.com/article/openai-chatgpt-spud-sam-altman-anthropic-mythos-3c2674f5cdf67ac6d88eedb207de117c.

Opascope (2026), "ChatGPT Ads Benchmarks: ROAS, CPC, and Cost From 15 Days of Real Spend," Opascope Insights (June 15), https://opascope.com/insights/chatgpt-ads-benchmarks/.

OpenAI (2026a), "Ads in ChatGPT," OpenAI Help Center, https://help.openai.com/en/articles/20001047-ads-in-chatgpt (accessed September 22, 2026).

OpenAI (2026b), "Ads in ChatGPT: The Basics," OpenAI Help Center, https://help.openai.com/en/articles/20001207-ads-in-chatgpt-the-basics (accessed August 31, 2026).

OpenAI (2026c), "ChatGPT Ads Expands Across Europe," OpenAI Blog (August 18; updated August 31), https://openai.com/index/chatgpt-ads-expands-across-europe/.

OpenAI (2026d), "ChatGPT Release Notes," OpenAI Help Center (entries of February 9, April 16, and June 4), https://help.openai.com/en/articles/6825453-chatgpt-release-notes.

OpenAI (2026e), "How ChatGPT Adoption Has Expanded," OpenAI Signals (June 30), https://openai.com/index/how-chatgpt-adoption-has-expanded/.

OpenAI (2026f), "Introducing ChatGPT Go, Now Available Worldwide," OpenAI Blog (January 16), https://openai.com/index/introducing-chatgpt-go/.

OpenAI (2026g), "A Milestone in Expanding Access to AI," OpenAI Blog (August 31), https://openai.com/index/expanding-access-to-ai-with-chatgpt-ads/.

OpenAI (2026h), "Testing Ads in ChatGPT," OpenAI Blog (February 9; updated March 26, May 7, and August 11), https://openai.com/index/testing-ads-in-chatgpt/.

PPC Land (2026), "ChatGPT Free and Go Users in Europe Face Ads from Later This Month," PPC Land (August 15), https://ppc.land/chatgpt-free-and-go-users-in-europe-face-ads-from-later-this-month/.

Rijo, Luis (2026), "OpenAI Gains 2 Latin American Markets for ChatGPT Ads, Reaching 9 Total," PPC Land (August 17), https://ppc.land/openai-gains-2-latin-american-markets-for-chatgpt-ads-reaching-9-total/.

Rodrigo, Michel (2026), "ChatGPT Inicia su Piloto de Anuncios en México," Marketing4eCommerce México (August 6), https://marketing4ecommerce.mx/chatgpt-inicia-su-piloto-de-anuncios-en-mexico/.

Sacchitiello, Bárbara (2026), "ChatGPT Começa a Exibir Anúncios aos Usuários do Brasil," Meio &

Mensagem (August 4), https://www.meioemensagem.com.br/midia/chatgpt-comeca-a-exibir-anuncios-aos-usuarios-do-brasil.

Salvi, Francesco, Alejandro Cuevas, and Manoel Horta Ribeiro (2026), "Commercial Persuasion in AI-Mediated Conversations," arXiv preprint arXiv:2604.04263 (April 5), https://arxiv.org/abs/2604.04263.

Scanlon, Krystal (2026), "OpenAI's Ads Business Hits Europe at the Six Month Mark," Digiday (August 19), https://digiday.com/marketing/openais-ads-business-hits-europe-at-the-six-month-mark/.

Simonov, Andrey, Chris Nosko, and Justin M. Rao (2018), "Competition and Crowd-Out for Brand Keywords in Sponsored Search," *Marketing Science*, 37 (2), 200–215.

Sun, Liyang and Sarah Abraham (2021), "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 225 (2), 175–199.

Tang, Brian Jay, Kaiwen Sun, Noah T. Curran, Florian Schaub, and Kang G. Shin (2025), "Ads that Talk Back: Implications and Perceptions of Injecting Personalized Advertising into LLM Chatbots," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 9 (4), Article 213.

United States v. Google LLC (2025), "Memorandum Opinion (Remedies), No. 1:20-cv-03010 (APM), ECF No. 1436 (D.D.C. September 2, 2025)," U.S. District Court for the District of Columbia (September 2), https://storage.courtlistener.com/recap/gov.uscourts.dcd.223205/gov.uscourts.dcd.223205.1436.0.pdf.

Wall Street Journal (2026), "OpenAI Plans Fourth-Quarter IPO in Race to Beat Anthropic to Market," The Wall Street Journal (January 29), https://www.wsj.com/tech/ai/openai-ipo-anthropic-race-69f06a42.

Yan, Shunyao, Klaus M. Miller, and Bernd Skiera (2022), "How Does the Adoption of Ad Blockers Affect News Consumption?," *Journal of Marketing Research*, 59 (5), 1002–1018.

Can Advertising Fund Free AI? Evidence from ChatGPT's International Advertising Rollout

Web Appendix

Web Appendix A:	Data Construction and Channel Tagging	21
Web Appendix B:	Samples and Treatment Assignment	23
Web Appendix C:	Descriptive Statistics	26
Web Appendix D:	Estimation of the Organic-Channel Models	32
Web Appendix E:	Inference and Sensitivity	37
Web Appendix F:	Primary Estimates and Estimator Choice	39
Web Appendix G:	Usage-Growth Scenarios	46
Web Appendix H:	What the Estimates Rule Out	48
Web Appendix I:	Validity and Robustness	51
Web Appendix J:	Estimation of Advertiser Value	53
Web Appendix K:	Channel Performance	58
Web Appendix L:	Prices and the Parity Price	62
Web Appendix M:	Advertiser Heterogeneity	63
Web Appendix N:	Matched Comparisons at Advertiser Entry	64
Web Appendix O:	What We Estimate and What the Data Can Show	67
Web Appendix P:	Financing Arithmetic	69

Web Appendix A Data Construction and Channel Tagging

The data comes from the Google Analytics 4 e-commerce tracking of a panel of websites, provided by a third-party market-intelligence company. Each daily record gives, for one website, one visitor country, one device category (desktop or mobile; tablets are classified as mobile at extraction), and one traffic source, the number of sessions, transactions, revenue, bounces, total session duration, and pageviews, as recorded by the website's own tracking; a transaction is a completed order, revenue is the order value the site reports including taxes but excluding shipping, and a bounce is a session in which the visitor leaves without a further action on the site (a non-engaged session in Google Analytics 4). Days without activity have no record and count as zero. Faulty daily records with more transactions than sessions or with negative revenue are dropped before aggregation. For our analyses, we aggregate the records to weeks that run Sunday to Saturday and are dated by the Monday they contain, and convert revenue to US dollars at the average exchange rates of the day. The estimation panels keep the device dimension; sample filters and descriptive statistics sum across devices (Web Appendix D).

Traffic sources are identified from the source and medium strings that Google Analytics records for each session. Organic ChatGPT combines the six ChatGPT referral strings (`chatgpt.com` with the mediums `(not set)`, `referral`, `organic`, `listing`, and `ai-assistant`, and `chatgpt / ai-assistant`). Paid ChatGPT combines 28 paid-tagged ChatGPT, OpenAI, and vendor variants carrying cpc, cpm, ppc, display, paid, sponsored, or paid-search tags, of which `chatgpt.com / cpc` carries almost all sessions. The same definitions determine which websites count as advertisers and how large their campaigns are, so every analysis works from the same channel definition.

In June 2026, Google Analytics changed the referral tag classification to `ai-assistant`; counting both tag generations keeps the series continuous, and after June 21, 2026, 98.0% of recognized organic ChatGPT sessions carry the new tag. During the three weeks dated May 25, June 1, and June 8, 2026, one large retailer's website records twice its usual US organic ChatGPT sessions with unchanged transactions, a pattern consistent with double-counted sessions; we halve its sessions, bounces, pageviews, and session duration in exactly those weeks. Ads were served

only in the nine treated countries from their launch dates, so paid-tagged ChatGPT sessions from any other country, and from a treated country before the delivery week that contains its launch date, cannot have been served and are removed from every panel; they total in the order of ten thousand sessions.

Advertising costs come from the same platform's Google Ads data: daily cost, clicks, and impressions by advertising network. Linked records exist for 31 of the 102 advertisers of the financial sample, the cost sample; 29 of them record search-network cost inside the window. Costs are recorded per account, not per visitor country. The parity calculation therefore uses the 16 of these 29 accounts whose paid-search traffic in the panel comes at least 80% from the United States. For each, the click price is its search-network cost divided by its search clicks over the full window; the median across the 16 is \$0.72. It is close to the \$1.01 median account price on Google's search network across 40,183 accounts in the months before the ad launch (the data provider's approximate account median; the two windows of the feed cannot be combined into one full-window median). The cost data enters only the paid-search click price used for the parity price in the Condition B section of the main text, and through it the financing arithmetic of the Discussion; no estimation sample or dependent variable uses it.

The growth of ChatGPT usage in the treated markets relative to Europe, which calibrates the growth scenarios of Web Appendix D, is taken from OpenAI's own published regional adoption index, weekly active users by region and month (OpenAI 2026e). Alternative outside series (Google Trends, Statcounter, available Similarweb and Semrush estimates) were examined and proved too unstable or too cost-prohibitive to serve as a control; they enter no estimate.

Web Appendix B Samples and Treatment Assignment

The panel covers 60 weeks, from the week dated June 30, 2025 to the week dated August 17, 2026, whose last day is Saturday, August 22, 2026. OpenAI announced ads for the European Economic Area on August 15 (PPC Land 2026) and ads went live on August 24 (OpenAI 2026c, Scanlon 2026), two days after the window closed, so the EU is untreated throughout.

Two samples are used. The organic referral panel serves Condition A. It starts from the websites in the data provider's feed, those with more than ten transactions under at least one ChatGPT source/medium label, organic or paid, summed over the delivery from June 2025 to the data pull, and takes all their traffic in the nine treated countries and the EU-27. A website's records in a region group, the five rollout cohorts and the EU, are kept when their totals over the window show plausible retail activity: more than ten transactions, an average order value consistently between \$5 and \$2,000, a conversion rate between 0.1% and 30%, and records in at least 55 of the 60 weeks. A website is dropped from every region group when more than 5% of its weekly website-by-region-by-channel cells with a transaction show an order value outside \$5 to \$2,000; this removes 225 of the 2,136 eligible websites. A website enters the panel if it passes in at least two region groups, which keeps the comparisons within internationally active websites, and only its passing region groups are used. The panel comprises the 1,898 websites the frozen-weight conversion model retains: those contributing at least one estimation unit, a website, visitor country, and device with organic ChatGPT and organic search sessions in the same week, inside the estimation window, after cells that are the only observation of their fixed-effect group are removed (1,902 websites before that step). Weeks in which a unit records no organic ChatGPT session are genuine zeros and are kept in the volume model; the conversion rate is undefined in such weeks and they drop out of the conversion model (Web Appendix D). A website need not have EU cells of its own.

The paid referral panel serves Condition B. It comprises the 387 websites of the two advertiser feeds that record at least one paid ChatGPT session inside the window, with no requirement of international activity. Advertisers whose first paid ChatGPT week falls after the window are dropped when the feeds are read; the feeds then hold 571 properties, and the 184 without a paid

ChatGPT session inside the window enter no Condition B sample. All 387 enter the engagement comparisons; the 102 in the e-commerce feed, which records transactions and revenue, with more than ten transactions over the window, summed over all traffic sources and visitor countries, form the financial sample for conversion, order value, and revenue per session. A website counts as an advertiser from its first week with paid-tagged ChatGPT sessions, so campaigns that never produce a click are invisible, and entry can be later than the campaign start.

Treatment is assigned at the country level and is intention-to-treat: every organic ChatGPT referral from a treated country counts as treated from the launch week on, whether or not the referring user saw an ad. The launch week of a country is the data week dated by the first Monday on or after its launch date, and a cohort is the set of countries with the same launch week: the United States, February 9, 2026, ads went live that Monday; Australia, Canada, and New Zealand, April 20, after the Thursday, April 16 launch; the United Kingdom, June 8, after the launch on the night of Thursday, June 4; Japan and Korea, June 22, after the Friday, June 19 launch; Brazil and Mexico, August 10, after the pilot launches on Tuesday, August 4 in Brazil and Thursday, August 6 in Mexico (OpenAI 2026d, Korea JoongAng Daily 2026, Japan Times 2026, Sacchitiello 2026, Rodrigo 2026). In Brazil and Mexico, the pilot launch, when ads began showing to users on the free and Go tiers, and the commercial opening to advertisers are two events: OpenAI listed both countries as launched in its August 11 update and announced commercial availability on August 17 (OpenAI 2026h, Rijo 2026); the cohort week rests on the pilot dates, when ads first reached users. Launch days before the first Monday fall into the pre-launch period, which if anything pulls the estimates toward zero. The EU-27 countries are the only control group; a country's own pre-launch weeks provide its baseline, and the estimator of Web Appendix D never uses an already-treated cohort as a control for a later one. The Brazil and Mexico cohort has two post-launch weeks inside the window. Changes common to all countries, such as model releases, global price changes, and international publicity, are absorbed by the week effects. OpenAI's country-specific changes inside the window were few: the low-cost Go tier reached countries in stages between August 2025 and January 2026, Instant Checkout was offered to US users from September 2025 to March 2026,

and ChatGPT was embedded in KakaoTalk in Korea from October 2025 (OpenAI 2026f, CNBC 2026b, Korea Times 2025). None coincides with a cohort's launch week, although the withdrawal of Instant Checkout falls in the US post-launch window. Italy's 2023 local suspension of ChatGPT predates the window.

Web Appendix C Descriptive Statistics

The paper uses two samples, defined in Web Appendix B. The organic referral panel of 1,898 internationally active websites serves Condition A; its totals cover each website's eligible region-group countries. The paid referral panel of 387 websites with paid ChatGPT referrals serves Condition B; its totals cover all visitor countries. Of the 387, the 102 with financial tracking and more than ten transactions over the window form the financial sample for conversion, order value, and revenue per session, 31 of the 102, the cost sample, have linked Google Ads cost records, and 87 of the 102 buy both paid channels in US traffic, the base of the breadth statistics in Web Appendix J. Table C.1 gives descriptive statistics for both panels, including the traffic shares of the two ChatGPT channels, and Table C.2 the website categories.

Table C.1 counts sessions, transactions, and revenue over the five channels the models use, direct, organic search, paid search, organic ChatGPT, and paid ChatGPT, for the 60 weeks dated June 30, 2025 to August 17, 2026, the last ending on August 22. On the organic referral panel these channels carry 6.3 billion sessions and $18.1 billion of revenue, about $16 billion a year. Organic ChatGPT contributes 22.4 million sessions and 446,090 transactions and converts at 1.99% over the full window. On the paid referral panel, session counts refer to all 387 websites and financial figures to the 102 websites of the financial sample; the 0.45 million paid ChatGPT sessions quoted in the main text (446,405) are the panel's total.

The traffic shares quoted in the main text are shares of all sessions from every source, including email, social, referral, and untagged traffic; Table C.1 reports them for both panels and two windows, the full 60 weeks and the final four weeks, the weeks dated July 27 to August 17, 2026. The organic referral panel recorded 9.9 billion sessions from all sources, the nearly 10 billion visits of the main text. Organic ChatGPT rose from 0.23% of them over the full window to 0.40% over the final four weeks. On the paid referral panel, paid ChatGPT was 0.007% of all sessions over the full window across 387 websites and 0.030% over the final four weeks across the 378 websites recording sessions then.

Table C.2 assigns each website to one of fifteen categories from its hostname. The organic

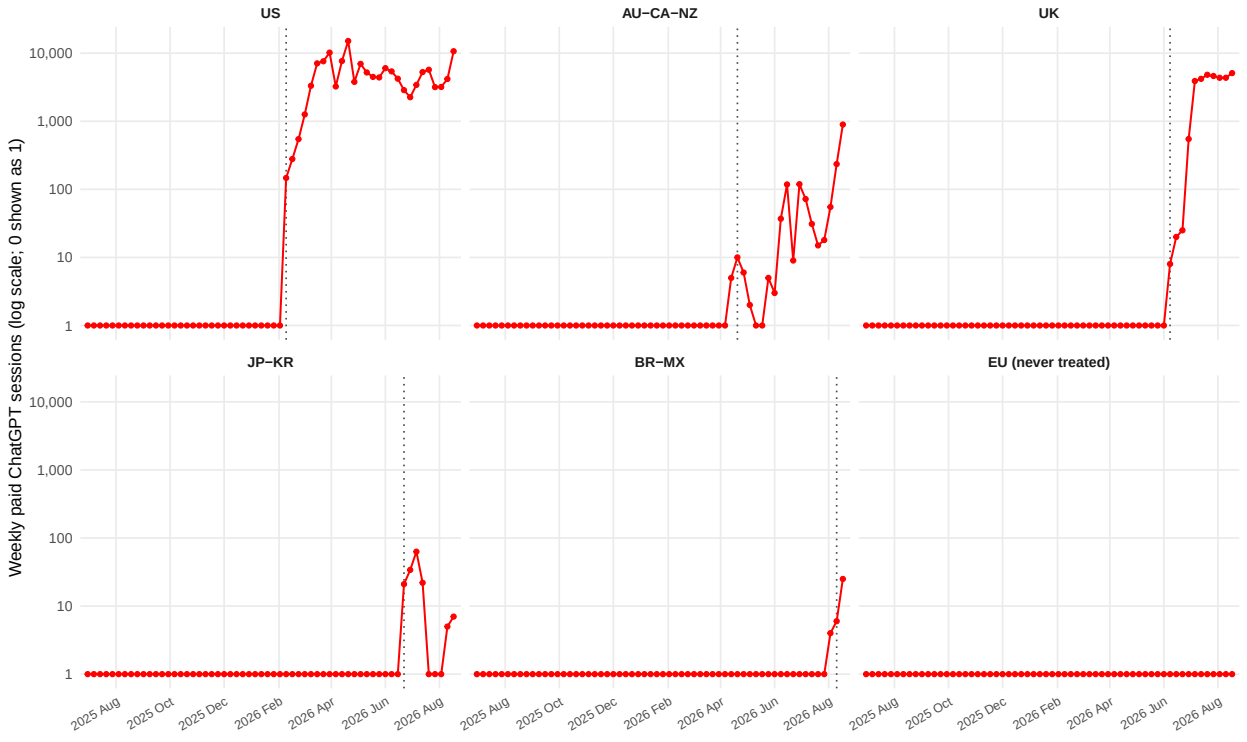


Figure C.1: **Paid ChatGPT referrals by rollout cohort.** Weekly paid-tagged ChatGPT sessions by cohort on a log scale; weeks without a paid session are plotted at 1. Dotted lines mark launch weeks. Paid-tagged sessions are kept only for treated countries from the delivery week containing their launch date (Web Appendix A). The series documents realized paid clicks after each launch, not ad impressions.

referral panel is predominantly retail: home and garden, fashion, electronics, health, and sports together hold 55% of its websites, with services, travel, and ticketing a minority. The paid referral panel is a different population: services, software, education, and finance make up half of its websites, and the five retail categories that lead the organic panel less than a third.

Figure C.1 shows paid ChatGPT referrals after the rollout in each cohort, week by week, documenting realized paid-click activity; impressions are unobserved. Figure C.2 shows the descriptive weekly conversion rates of organic ChatGPT and the established channels for each cohort and for the EU, accompanying the estimates of the Condition A section of the main text.



Figure C.2: **Weekly conversion rates by channel and rollout cohort.** Descriptive weekly conversion rates, transactions per session, session-weighted across the organic referral panel, for organic ChatGPT, direct, organic search, and paid search traffic, in each treated cohort and the untreated EU. Dotted lines mark launch weeks.

Table C.1: Descriptive Statistics by Sample

	Condition A	Condition B	
	Organic referral panel *1,898 websites*	Paid referral panel *387 websites*	Financial sample *102 of the 387*
SAMPLE			
Weeks (June 30, 2025 to August 22, 2026)	60	60	60
Websites	1,898	387	102
Website–region–week–channel cells	1,535,742	452,161	133,559
Sessions from all sources, full window (millions)	9,936	6,863	2,330
Sessions from all sources, final four weeks (millions)	685	509	161
FIVE MODELED CHANNELS COMBINED			
Sessions, full window (millions)	6,330	4,574	1,487
Sessions, final four weeks (millions)	437	343	104
Transactions (millions)	131.3	—	17.2
Revenue ($ millions)	$18,086	—	$2,617
Conversion rate	2.07%	—	1.16%
Revenue per session	$2.86	—	$1.76
Average order value	$137.78	—	$152.00
ORGANIC CHATGPT REFERRALS			
Sessions, full window	22,408,828	20,272,944	5,706,551
Sessions, final four weeks	2,765,769	2,597,808	682,470
Share of all-source sessions, full window	0.226%	0.295%	0.245%
Share of all-source sessions, final four weeks	0.404%	0.511%	0.423%
Transactions	446,090	—	74,977
Revenue	$72,456,665	—	$14,161,260
Conversion rate	1.99%	—	1.31%
Revenue per session	$3.23	—	$2.48

Table C.1: Descriptive Statistics by Sample (continued)

	Condition A	Condition B	
	Organic referral panel *1,898 websites*	Paid referral panel *387 websites*	Financial sample *102 of the 387*
PAID CHATGPT REFERRALS			
Sessions, full window	not used	446,405	194,367
Sessions, final four weeks	not used	153,127	49,023
Share of all-source sessions, full window	not used	0.0065%	0.0083%
Share of all-source sessions, final four weeks	not used	0.0301%	0.0304%
Transactions	not used	—	603
Revenue	not used	—	$142,019
Conversion rate	not used	—	0.31%
Revenue per session	not used	—	$0.73
SESSIONS BY REGION GROUP, FIVE CHANNELS (MILLIONS)			
US	1,807	1,388	840
AU–CA–NZ	434	221	98
UK	771	1,165	312
JP–KR	94	136	4
BR–MX	183	881	77
EU (never treated)	3,039	268	72
Other countries	not in panel	515	82

NOTES: Each column refers to its own websites; the financial sample is a subset of the paid referral panel (samples defined in Web Appendix B). All sources = every source/medium the website records; five modeled channels = direct, organic search, paid search, organic ChatGPT, paid ChatGPT. Final four weeks = the weeks dated July 27 to August 17, 2026 (378 of the 387 paid-panel websites record sessions in them). Rates are ratios of window totals. A cell is one website, visitor region group, week, and channel with at least one session. — = no revenue tracking; not used = Condition A uses organic referrals only. Financial-sample figures pool all visitor countries; the US-only figures of the main text are in Table J.1, Panel B.

Table C.2: Website Categories by Sample

	Organic referral panel		Paid referral panel		Financial sample	
	% of websites	% of sessions	% of websites	% of sessions	% of websites	% of sessions
Home, garden & DIY	16.2	18.1	13.4	7.8	22.5	20.6
Fashion, footwear & jewelry	13.2	20.6	6.5	17.0	11.8	21.4
Services, software, education & finance	11.3	9.8	52.2	37.1	20.6	31.0
Electronics & appliances	9.9	11.4	3.1	2.9	5.9	5.4
Health, pharmacy & supplements	8.2	3.5	3.4	0.2	2.0	0.1
Sports & outdoor	8.0	7.2	1.8	1.1	6.9	3.1
Travel, tickets & leisure	7.6	3.9	7.2	11.4	6.9	3.0
Beauty & personal care	5.7	4.6	3.4	0.4	6.9	0.9
Food, drink & grocery	5.2	2.9	1.3	0.3	2.9	0.3
Books, media, hobby & gifts	4.7	3.9	2.1	3.7	4.9	10.4
Auto & motor	2.8	1.3	2.1	1.0	1.0	0.3
Toys, baby & kids	2.4	3.1	1.8	1.4	3.9	3.3
Pets	2.1	0.7	0.5	0.1	2.0	0.2
Other	1.4	0.4	0.5	0.0	2.0	0.1
General merchandise & marketplaces	1.3	8.6	0.8	15.6	0.0	0.0
Websites	1,898	1,898	387	387	102	102

NOTES: Shop-level categories assigned from each website's hostname by a large language model in one pass (15 categories; the authors reviewed the assignments), following the LLM classification of website categories in Kaiser and Schulze (2026). The organic referral panel is the Condition A panel (the websites the frozen-weight conversion model retains inside the event windows); the paid referral panel is every website of the e-commerce and the non-e-commerce advertiser feed with a paid ChatGPT session inside the window; the financial sample is the subset with complete revenue tracking and more than ten transactions over the window (Web Appendix B). Session shares weight each website by its sessions from all sources over the window. 'Services, software, education & finance' collects websites whose online conversions are bookings, sign-ups, licenses or leads rather than shop orders; non-e-commerce feed websites are classified from their best-matching hostname, unclassified websites (no hostname on record) are shown separately when present.

Web Appendix D Estimation of the Organic-Channel Models

All Condition A models compare, within a website, visitor country, and device (a unit), organic ChatGPT referrals with organic search referrals week by week, and follow that comparison across each cohort's launch against the untreated EU. Using organic search as the comparison channel assumes that ChatGPT's ads did not change organic search referrals to the same website; a spillover into organic search would bias the comparison, and Table F.1, Panel B, repeats the estimates with direct traffic as the comparison channel. Cohort-by-event-time effects are estimated as interactions with the never-treated EU as the reference group, together with the treated countries' organic-search cells in the volume model, the construction of Sun and Abraham (2021), so that each cohort's effects are measured only against genuine controls; they are then averaged with weights proportional to the number of estimation cells each cohort contributes at each event time, or, in the conversion model, to those cells' regression weights. A pooled two-way fixed-effects regression would instead mix in comparisons of later cohorts against already treated ones, whose own dynamics would masquerade as the counterfactual. Treated units enter from 12 weeks before to 12 weeks after their launch week; EU units enter for all 60 weeks; the common window keeps the cohorts' horizons comparable and gives each a pre-launch baseline of the same length. The headline effect is the average over the launch week and the twelve weeks after it, net of the pre-launch average defined below.

Referral volume is modeled as session counts by Poisson pseudo-maximum likelihood on the full zero-filled grid of unit, channel, and week, so weeks in which a unit received no organic ChatGPT referrals count as zeros rather than dropping out; in treated countries only 24% of unit-weeks before February 9, 2026, and 27% after have sessions in both channels, although those cells hold 99.9% of organic ChatGPT sessions, so a ratio of the two channels' counts would discard most of the grid. The two channels are stacked: organic ChatGPT cells of treated countries are the treated observations, organic search cells of the same countries and all EU cells are coded as never treated. The model carries fixed effects for unit by channel, country by week, channel by week, and device by week, and the cohort-by-event-time interactions. Shocks common to a country-week, a channel-week, or a device-week are absorbed; a shock to one website in one week is absorbed only

insofar as it is common to those groups. Coefficients are log points and are reported as percentage changes.

Notation. A unit i is a website, visitor country $k(i)$, and device $d(i)$ (one website reports through two analytics views, whose records enter separately); $c \in \{\text{ChatGPT}, \text{search}\}$ indexes the organic channel; w indexes weeks; g indexes cohorts with launch week L_g, and $r = w - L_g$ is event time; $G(i)$ is the cohort of the unit's country, none for EU units. S denotes sessions and CR the conversion rate in percent. The expected organic session count of unit i in channel c and week w is

$$\mathrm{E}[S_{icw}] = \exp\Big(\alpha_{ic} + \lambda_{k(i),w} + \mu_{cw} + \delta_{d(i),w} + \sum_{g}\sum_{r\neq -1} \beta_{gr}\, \mathbf{1}[c = \text{ChatGPT}]\, \mathbf{1}[G(i) = g]\, \mathbf{1}[w - L_g = r]\Big), \tag{1}$$

with α a unit-by-channel effect, λ a country-by-week effect, μ a channel-by-week effect, and δ a device-by-week effect. The indicators are one only for organic ChatGPT cells of treated countries at event time r, so organic search cells and all EU cells carry no treatment term. Reading: β_{gr} is the log change in cohort g's organic ChatGPT referrals at event time r relative to week -1, net of what happened to organic search in the same units and to ChatGPT referrals in the EU in the same week; $100(\exp(\beta_{gr}) - 1)$ is the percentage change.

Conversion is modeled as the weekly difference between a unit's organic ChatGPT and organic search conversion rates, in percentage points, with the same cohort-by-event-time interactions and with unit and device-by-week fixed effects. Because the two channels are differenced within a unit-week, shocks common to both channels of a website in a week cancel. Each unit-week is weighted so that units with little traffic in either channel do not dominate: the weight is the unit's mean, over the weeks before February 9, 2026 in which both channels have sessions, of the smaller of the two channels' weekly session counts, and it stays fixed after launch; units without such a week leave the panel. This follows one rule for both outcomes: nothing that advertising can change enters the weights. A unit-week in which either channel has no sessions has no defined rate and is not used, so the conversion sample can change composition after launch; the volume model, which keeps such weeks as zeros, is the guard against that.

The dependent variable is the weekly within-unit gap $y_{iw} = \mathrm{CR}_{iw}^{\mathrm{ChatGPT}} - \mathrm{CR}_{iw}^{\mathrm{search}}$, in percentage points, and

$$y_{iw} = \alpha_i + \delta_{d(i),w} + \sum_g \sum_{r \neq -1} \beta_{gr} \, \mathbf{1}[G(i) = g] \, \mathbf{1}[w - L_g = r] + \varepsilon_{iw}, \tag{2}$$

estimated by weighted least squares with the fixed weight ω_i, the unit's mean over the weeks before February 9, 2026 in which both channels have sessions of $\min(S_{iw}^{\mathrm{ChatGPT}}, S_{iw}^{\mathrm{search}})$. Reading: β_{gr} is the change in the ChatGPT-minus-search conversion gap of cohort g at event time r relative to week -1, compared with the same gap in EU units in the same week and net of device-specific week effects; the weight gives units with more traffic in both channels more say and does not move after launch.

Post-launch effects are summarized in two ways. The headline re-centers the post-launch average on the pre-launch average over weeks -12 to -2, both pooled over cohorts with the cell weights of their own period, computed as a linear contrast on the fitted coefficients with delta-method standard errors from the website-clustered covariance; this avoids anchoring on a single, possibly noisy reference week and leaves week -1 readable as a placebo deviation from the pre-period mean. Because each average weights the cohorts by their cells, or their cell weights, in that period, a cohort with few post-launch weeks counts less after launch than before; with one cohort weight in both terms, the post-launch shares, the volume headline is -3.9% (Table F.1). Cohort-specific effects are in Web Appendix F.

With n_{gr} the number of estimation cells of cohort g at event time r (for conversion, the sum of their weights), N_{post} their sum over $r = 0$ to 12 across cohorts, and N_{pre} their sum over $r = -12$ to -2, the effect reported in the main text is

$$\tau = \sum_g \sum_{r=0}^{12} \frac{n_{gr}}{N_{\mathrm{post}}} \beta_{gr} \; - \; \sum_g \sum_{r=-12}^{-2} \frac{n_{gr}}{N_{\mathrm{pre}}} \beta_{gr}, \tag{3}$$

the cell-weighted post-launch average minus the cell-weighted pre-launch average, pooled over cohorts. Reading: the average post-launch effect is measured against the average pre-launch deviation from week -1, so the comparison does not rest on a single reference week; referencing

each cohort to week -1 alone would drop the second term. Standard errors come from the delta method on the website-clustered covariance of the β's. For conversion, τ is in percentage points; for volume, τ is an average log effect, and the reported figure is $100(\exp(\tau) - 1)$; intervals are transformed endpoint by endpoint and standard errors are reported in log points.

The comparison with the EU assumes that, without advertising, organic ChatGPT referrals relative to organic search would have moved alike in the treated countries and in the EU. If ChatGPT was adopted faster in the treated countries, the estimate would be biased upward, that is, it would understate any decline. No outside series is precise enough to serve as a control, and post-launch usage must not enter the model at all, because advertising can change it. The growth differential therefore enters as a deterministic adjustment inside the identical estimator: the organic ChatGPT session counts of treated countries are scaled down by a factor of $(1 + g)^{-t/52}$, where t is the number of weeks since the start of the panel and g the assumed annual growth advantage of the treated countries over the EU, before and after launch; organic search counts and EU cells are untouched. The adjustment is a smooth pre-existing trend from a fixed anchor, not a kink at launch, and it assumes every additional user as productive as existing ones. The estimate shifts almost exactly linearly in $\log(1 + g)$ with unchanged standard errors, which yields the scenario grid and the threshold at which the interval first excludes zero; each threshold was verified by refitting the model at that value (Table F.1; Web Appendix G). The two values used in the main text, 6% and 12% per year, come from OpenAI's regional adoption index (Web Appendix A): North America's weekly active users outgrew Europe's by 5.7% between January 2025 and January 2026, and by 11.6% at an annual rate from July 2025 to June 2026.

The estimates are intention-to-treat on all organic referrals from treated countries, but only users on the free and Go tiers see ads. The rescaling in Figure 3, B and D, is based on eligibility, not on observed exposure: if a share s of referral sessions comes from users who could see ads, the remaining sessions are unaffected, and effects are homogeneous among the eligible, the effect per eligible user is the market-level effect divided by s. We compute s from the share of paying users, about 5%, and their usage relative to free users: at equal usage $s = 0.95$, at five times $s = 0.79$, and

at ten times $s = 0.66$, which brackets the CFO's account of Plus users at three times and Pro users at eleven times a free user's usage (Friar 2026). Treating all paying users as ad-free understates s, since Go subscribers see ads, so the rescaled bounds are if anything too wide. The figure divides the log effect and its standard error by s before converting to percentages; the exact relative-effect alternative, the percentage effect divided by s, differs by less than a percentage point at $s = 0.66$. The bounds quoted in the main text are the lower endpoints of the 95% intervals, in percentage points and as a share of the pre-launch organic ChatGPT conversion rate of the treated units under the model's weights, 2.21%, for the headline and for each alternative convention.

Web Appendix E Inference and Sensitivity

Standard errors for every Condition A regression are clustered by website, the level at which the panel's units are sampled, with 1,898 clusters in the organic referral panel. Treatment is assigned by country, which argues for clustering by country as well, and we examined two-way clustering by website and country. The panel has 36 country clusters, nine of them treated. On every Poisson fit the two-way covariance is singular: on the headline fit it has rank 51 among 103 cohort-by-event-time parameters, 52 zero eigenvalues and no negative one, whereas the website-clustered covariance has full rank. Its standard errors are smaller than the website-clustered ones on every estimate but the Brazil and Mexico cohort of the count model and the Australia, Canada, and New Zealand cohort of the conversion model, and the placebo exercise below shows that it is unusable: its 95% interval excludes zero in 40 of 100 rollouts in which no ads launched, and its joint pre-trend test returns $p < .05$ in 96 of them. The website-clustered error is closer to nominal but also over-rejects on this design: the 95% interval excludes zero in 13 of 100 placebo rollouts and the joint pre-trend test returns $p < .05$ in 49 of 100, which is why the main text reports the placebo count next to every interval and why the pre-trend evidence rests on the placebo yardstick as much as on the test (Table F.2).

The placebo exercise re-runs the headline estimator on rollouts that never happened. The real treated countries leave the panel; a pseudo-rollout keeps the real launch weeks and the real cohort sizes, one, three, one, two, and two countries, draws the pseudo-treated countries without replacement from the EU-27, and uses the remaining EU countries as the control group; estimator, fixed effects, event window and sample rules are otherwise identical to the headline, and each conversion draw uses the headline panel's frozen weights, computed once from the calendar weeks before February 9, 2026. For volume we use 100 draws each a Poisson refit on the full grid that also yields the two-way covariance and the joint pre-trend test; for conversion 500 draws, with website-clustered errors only. The main text reports the number of draws with a decline at least as large as the real estimate. Read as a share of the draws, this is a one-sided placebo p-value under the assumption that pseudo-rollouts among EU countries are comparable to the real rollout; the

two-sided version counts draws whose absolute estimate is at least as large as the real one, and both are reported in Table F.2 together with the band of the draws and the share of draws in which each nominal 95% interval excludes zero, the calibration check behind the clustering choice above.

Pre-launch parallel movement is tested with a joint Wald test that the event-time coefficients for weeks -12 to -2, relative to week -1, are zero, on the website-clustered covariance: once per outcome on the cohort-averaged path and once per cohort (Table I.1). The joint test has little power against a smooth growth differential of the size considered in Web Appendix D: a differential of g per year adds a pre-launch slope of $\log(1+g)/52$ per week, 0.0011 to 0.0022 at g of 6% to 12%, whereas the slope of the count model's pre-launch path is estimated at 0.0031 per week with a standard error of 0.0036, so the smallest slope it can detect at 80% power is 0.010 per week, about nine times the 6% differential; for conversion the detectable slope is 0.017 percentage points per week.

The Condition A estimates are re-run on alternative samples. By country: without the United States cohort, whose launch coincided with two publicized events involving a competing assistant, and, for conversion, without each other country in turn. By website: without the largest website, without each of the five largest in turn, and without the ten largest together, since a few websites carry much of the treated traffic. By comparison channel: with direct traffic instead of organic search (Table F.1, Panel B).

Web Appendix F Primary Estimates and Estimator Choice

This section reports the two primary estimates of Condition A. Table F.1 carries the referral-volume estimates by cohort and sample, with the growth-differential scenarios of Web Appendix G, and Table F.3 the conversion estimates by cohort and sample. The two headline estimates of the Condition A section of the paper are the first rows of these tables; the bounds and the pre-trend and placebo results that the main text also quotes follow in Web Appendices E, H, and I, and the event-study paths are Figure 3 of the paper. Inference throughout is on website-clustered standard errors (Web Appendix E); the two-way website-and-country covariance is kept in the export files only.

Referral volume. Averaged over the launch week and the twelve weeks after it and re-centered on the pooled pre-launch average, organic ChatGPT referrals changed by −5.9%, with a 95% interval from −12.6% to +1.2% (Table F.1, Panel A). By cohort, the United States shows −9.2%, the United Kingdom −6.8% over its eleven post-launch weeks, Australia, Canada, and New Zealand −2.6%, Japan and Korea +13.6% over nine weeks, and Brazil and Mexico −23.4% over two; with website clustering only the last is distinguishable from zero, and the three cohorts with fewer than thirteen post-launch weeks are averaged over the weeks they have. The count model weights cohorts by their estimation cells, so the pooled estimate is not an average of five equally weighted experiments, and the pooled contrast subtracts a pre-launch average in which Brazil and Mexico, the most negative cohort with two post-launch weeks, weigh more than in the post-launch average; with one cohort weight in both terms, the post-launch shares, the estimate is −3.9% (Panel A, second row; Web Appendix D). Panels B and D hold the sample checks quoted in the paper: without the United States cohort the estimate is −5.8%, and dropping the largest websites one at a time or the ten largest together moves it between −5.1% and −8.8%. Panel C carries the growth-differential scenarios of Web Appendix G.

Conversion. With weights frozen at pre-launch levels, the gap between the organic ChatGPT and the organic search conversion rate changed by −0.024 percentage points (standard error 0.053) re-centered on the pre-launch average (Table F.3, Panel A); the 95% interval reaches −0.13 points, the bound behind the 6% of baseline quoted in the paper, which Web Appendix H expresses as a share of the pre-launch conversion rate of the treated units under the model's weights, 2.21% (the full-window descriptive rate of Table C.1 is 1.99%). Without the United States cohort the estimate is −0.003 (0.054), and without the United Kingdom, the one cohort whose pre-launch path rejects (Web Appendix I), −0.034 (Panel B). The cohort-specific effects (Panel C) lie between −0.18 and +0.01 points, none distinguishable from zero; the United States carries 61% of the post-launch weight, the United Kingdom 18%, Australia, Canada, and New Zealand 16%, Japan and Korea 4%, and Brazil and Mexico 0.5%. Pre-trends and placebo rollouts are in Web Appendix I; what the intervals rule out is in Web Appendix H.

Table F.1: Referral Volume: Sun–Abraham PPML on the Zero-Filled Panel and Its Robustness

	Estimate (SE website)	p	Implied	95% CI website	Cells	Websites
Panel A: headline, cohorts and the cohort-consistent contrast						
Post-launch effect, re-centered on the pooled pre-launch average (weeks 0–12)	−0.061 (0.037)	.103	−5.9%	[−12.6, 1.2]	9,624,538	1,898
cohort-consistent contrast (post-launch cohort shares in both terms)	−0.040 (0.035)	.261	−3.9%	[−10.4, 3.0]	9,624,538	1,898
United States (13 post-launch weeks)	−0.096 (0.080)	.231	−9.2%	[−22.4, 6.3]	9,624,538	1,898
Australia, Canada, New Zealand (13 post-launch weeks)	−0.026 (0.053)	.622	−2.6%	[−12.2, 8.1]	9,624,538	1,898
United Kingdom (11 post-launch weeks)	−0.070 (0.059)	.236	−6.8%	[−16.9, 4.7]	9,624,538	1,898
Japan, Korea (9 post-launch weeks)	0.128 (0.097)	.189	13.6%	[−6.1, 37.5]	9,624,538	1,898
Brazil, Mexico (2 post-launch weeks)	−0.267 (0.074)	<.001	−23.4%	[−33.7, −11.5]	9,624,538	1,898
Panel B: samples and comparison channel						
Without the United States cohort	−0.059 (0.044)	.178	−5.8%	[−13.6, 2.7]	9,482,988	1,898
Direct traffic instead of organic search as the comparison channel	−0.087 (0.057)	.124	−8.4%	[−18.0, 2.4]	9,573,588	1,898
Panel C: usage-growth differential removed before estimation						
Growth differential of 6%/yr removed	−0.075 (0.037)	.045	−7.2%	[−13.8, −0.2]	9,624,538	1,898
Growth differential of 12%/yr removed	−0.088 (0.037)	.019	−8.4%	[−14.9, −1.5]	9,624,538	1,898
Growth differential of 18%/yr removed	−0.100 (0.037)	.007	−9.6%	[−15.9, −2.7]	9,624,538	1,898
threshold: the interval first excludes zero at a differential of 5.3%/yr (model re-estimated at that value: −7.1% [−13.6, 0.0])						

Table F.1: Referral Volume: Sun–Abraham PPML on the Zero-Filled Panel and Its Robustness (continued)

	Estimate (SE website)	p	Implied	95% CI website	Cells	Websites
Panel D: leave the largest websites out						
Without the largest website (13.5% of treated post-launch sessions)	−0.058 (0.039)	.135	−5.6%	[−12.6, 1.8]	9,617,325	1,897
Without the five largest websites	−0.070 (0.039)	.075	−6.8%	[−13.7, 0.7]	9,596,571	1,893
Without the ten largest websites	−0.092 (0.041)	.024	−8.8%	[−15.8, −1.2]	9,560,311	1,888
Without the website ranked 2 (7.6% of treated post-launch sessions)	−0.065 (0.038)	.092	−6.3%	[−13.1, 1.1]	9,618,968	1,897
Without the website ranked 3 (4.6% of treated post-launch sessions)	−0.053 (0.037)	.156	−5.1%	[−11.8, 2.0]	9,623,858	1,897
Without the website ranked 4 (4.1% of treated post-launch sessions)	−0.072 (0.037)	.048	−7.0%	[−13.4, −0.1]	9,617,286	1,897
Without the website ranked 5 (2.9% of treated post-launch sessions)	−0.059 (0.039)	.126	−5.7%	[−12.6, 1.7]	9,617,286	1,897

NOTES: Poisson pseudo-maximum likelihood with Sun–Abraham cohort×event-time interactions on the zero-filled website×country×device×channel×week session panel; treated units enter from 12 weeks before to 12 weeks after their launch week and EU units for all weeks; the organic ChatGPT cells of treated countries are the treated units, the organic-search cells of the same countries and all EU cells are never treated; fixed effects unit×channel, country×week, channel×week, device×week; cohort×week effects aggregated with equal weight per cell and re-centered to the pooled pre-launch average (weeks −12 to −2); cohorts with fewer than thirteen post-launch weeks are averaged over the weeks they have. The cohort-consistent row uses one cohort weight (the post-launch share) in both the post- and the pre-launch average (Web Appendix D). Implied effect = $\exp(\hat{\beta}) - 1$: the proportional change in organic ChatGPT referrals to the panel's treated cells. Growth rows remove a differential trend of g percent per year in ChatGPT use from treated cells before estimation (every additional user as productive as existing ones); the threshold is solved by linear interpolation in $\log(1 + g)$ and confirmed by re-estimating the model at the threshold. Leave-out rows drop the websites with the largest post-launch organic ChatGPT volume. Standard errors clustered by website; p is the exact two-sided p-value of the estimate against zero on that standard error; the two-way (website and country) covariance is reported in Web Appendix E. Pre-trend joint Wald test on weeks −12 to −2 (website-clustered): $\chi^2(11) = 17.0$, $p = .109$ (Table I.1). Cohort-preserving placebo rollouts on the EU (100 draws): 10 draws at or below the headline (Table F.2).

Table F.2: Placebo Rollouts in the Never-Treated EU: Volume and Conversion

	Referral volume (PPML, log points)	Conversion rate (IW, frozen weights, pp)
Actual estimate, re-centered post-launch average	-0.061 (-5.9%)	-0.024
Pseudo-rollouts drawn / converged	100 / 100	500 / 500
Distribution of the placebo estimates		
Median	+0.007	+0.001
5th–95th percentile band	[-0.082, +0.081]	[-0.154, +0.159]
Standard deviation	0.053	0.094

Table F.2: Placebo Rollouts in the Never-Treated EU: Volume and Conversion (continued)

	Referral volume (PPML, log points)	Conversion rate (IW, frozen weights, pp)
Tail counts against the actual estimate (share in parentheses)		
Draws at least as negative (one-sided p)	10 of 100 (.100)	187 of 500 (.374)
Draws at least as large in absolute value (two-sided p)	26 of 100 (.260)	378 of 500 (.756)
Draws at least as negative as the 6%/yr growth-adjusted estimate (-0.075)	8 of 100 (.080)	—
Draws at least as negative as the 12%/yr growth-adjusted estimate (-0.088)	4 of 100 (.040)	—
Calibration of the nominal inference across draws		
Nominal 95% interval excludes zero, website-clustered SE	13 of 100 (.130)	51 of 500 (.102)
Nominal 95% interval excludes zero, two-way SE	40 of 100 (.400)	not computed
Joint pre-trend test rejects at 5%, website-clustered SE	49 of 100 (.490)	not computed
Joint pre-trend test rejects at 5%, two-way SE	96 of 100 (.960)	not computed

NOTES: Each pseudo-rollout removes the real treated countries, keeps the five real launch weeks and the real cohort sizes (one, three, one, two, and two countries), draws the pseudo-treated countries without replacement from the EU-27, and re-estimates the headline model of the column; the estimate is the same re-centered post-launch average (weeks 0 to 12 against the pre-launch average over weeks -12 to -2) as the actual one. Volume: 100 draws of the Sun–Abraham Poisson model on the zero-filled grid (seed 20260918), with the website-clustered and the two-way (website and country) covariance and the joint pre-trend Wald test per draw. Conversion: 500 draws of the interaction-weighted conversion model with calendar-frozen weights (seed 20260729), website-clustered covariance per draw. Because no ads were served in the EU, the spread of the placebo estimates measures ordinary cross-market fluctuation for the estimator and averaging window; the shares are permutation calibrations under exchangeability of countries, not design-based randomization p-values. The growth-adjusted rows count draws at least as negative as the actual estimate after the treated cohorts' organic ChatGPT sessions are deflated by a 6 or 12 percent-per-year growth differential (usage-growth scenarios, Web Appendix G). The calibration rows count draws whose nominal 95 percent interval excludes zero, which would be about 5 of 100 under correct coverage, and draws whose joint pre-trend test rejects at the 5 percent level.

Table F.3: Conversion Rate: Sun–Abraham Estimates with Frozen Weights, by Sample and Cohort

	Estimate (SE website)	p	95% CI website	Cells	Websites
Panel A: weights frozen at pre-launch levels (headline)					
Post-launch effect, re-centered (weeks 0–12)	−0.024 (0.053)	.648	[−0.13, 0.08]	733,142	1,898
Panel B: samples					
Without the United States cohort	−0.003 (0.054)	.951	[−0.11, 0.10]	696,666	1,869
Without the United Kingdom	−0.034 (0.062)	.590	[−0.16, 0.09]	710,104	1,892

Panel C: cohort-specific effects of the headline model, re-centered (weeks 0–12)

	Estimate (SE website)	p	Post-launch weeks	Share of weight
United States	−0.032 (0.083)	.703	13	60.8%
Australia, Canada, New Zealand	0.013 (0.068)	.845	13	16.3%
United Kingdom	0.006 (0.081)	.941	11	18.3%
Japan, Korea	−0.180 (0.122)	.140	9	4.1%
Brazil, Mexico	−0.021 (0.212)	.922	2	0.5%

NOTES: Dependent variable: the weekly within-unit gap between the organic ChatGPT and the organic search conversion rate, in percentage points (Web Appendix D). Sun–Abraham cohort-by-event-time interactions with the never-treated EU as the reference group, unit and device-by-week fixed effects, weighted least squares with each unit's weight frozen at its pre-launch mean of the smaller of the two channels' weekly session counts. The post-launch effect is the cell-weighted average over weeks 0 to 12 minus the cell-weighted average over weeks −12 to −2, pooled over cohorts, with delta-method standard errors on the website-clustered covariance; p is the exact two-sided p-value of the estimate against zero on that standard error. Panel B re-estimates the model without the United States cohort and without the United Kingdom, the one cohort whose pre-launch path rejects the joint test (Table I.1). Panel C: each cohort's own post-launch average minus its pre-launch average, the number of post-launch weeks the cohort has inside the window, and its share of the post-launch weight in the pooled effect. Websites: Panel A uses the 1,898 websites the frozen-weight model retains, the Condition A panel of every exhibit; the Panel B samples are subsets of it.

Web Appendix G Usage-Growth Scenarios

If ChatGPT usage grew faster in the treated countries than in the EU for reasons unrelated to advertising, the headline understates the decline; Web Appendix D defines how an assumed annual growth advantage g of the treated countries is removed from their organic ChatGPT session counts before estimation. Table F.1, Panel C, reports the estimates for the two values of g taken from OpenAI's regional adoption index, an 18% stress case beyond the index, and the value of g at which the interval first excludes zero.

The two scenarios. OpenAI's index of weekly active users by continent (OpenAI 2026e) shows North America outgrowing Europe by 5.7% between January 2025 and January 2026 and by 11.6% at an annual rate from July 2025 to June 2026. The scenarios use 6% and 12% per year and apply one value to all treated countries, since the index resolves continents, not countries. With a 6% differential removed, organic ChatGPT referrals changed by −7.2% after launch, with a 95% interval from −13.8% to −0.2%; with 12% removed, by −8.4%, from −14.9% to −1.5%; an 18% differential, beyond anything the index shows, gives −9.6% (standard error 0.037). The range of roughly 6 to 8% quoted in the Condition A section of the paper spans the headline estimate and the 12% scenario; it is a range across assumptions, not an interval. On the coefficient scale the adjustment shifts the estimate by −0.24 log points per unit of $\log(1+g)$ and leaves the standard errors unchanged to the fourth decimal, so intermediate values follow by linear interpolation in $\log(1+g)$; the threshold below was solved on that interpolation and verified by refitting the model at that value.

Where the inference changes. The website-clustered 95% interval touches zero at a differential of 5.3% per year, a nominal threshold given the over-rejection reported in Web Appendix I; the refit at that value gives −7.1% with an interval from −13.6% to 0.0% (Table F.1, Panel C). Of the 100 cohort-preserving EU pseudo-rollouts of Web Appendix I, 10 produce a post-launch effect at or below the headline estimate, 8 at or below the estimate with 6% removed, and 4 at or below the

estimate with 12% removed, the counts quoted in the paper.

Web Appendix H What the Estimates Rule Out

This section converts the intervals of Web Appendix F into the declines they exclude and rescales them to the users who can see ads (Table H.1).

Under parallel trends. The headline conversion interval reaches −0.13 percentage points, so declines larger than 5.8% of the 2.21% pre-launch rate lie outside the 95% interval (Table H.1, Panel A); referenced to week −1, the least favorable of the two reference conventions, the excluded decline is 10.7%. Against the placebo distribution of Web Appendix I instead of the model's standard error, the excluded conversion decline is 0.22 points, 9.8% of the baseline, and 37 of the 500 no-ads rollouts fall below the model's bound of 0.13 points; for volume the placebo yardstick is wider, 15.7% against the model's 12.6%, although none of the 100 volume draws falls below the model's bound (Panel D). For referral volume the headline interval reaches −12.6% (Panel B); the growth scenarios of Web Appendix G move it to −13.8% and −14.9%.

Per user who can see ads. Only users on the free and Go tiers see ads. Panel C divides the headline bounds by the share s of referral sessions from those users, computed from a 5% paying-user share and their usage relative to free users (Web Appendix D). At paying users using ChatGPT five times as much as free users, $s = 0.79$, the excluded conversion decline is 7.3% and the excluded volume decline 15.6%, the 7% and 16% quoted in the paper; at ten times, $s = 0.66$, they are 8.8% and 18.5%, the 9% and 19% quoted in parentheses. Panel D's last two rows divide the placebo-yardstick bounds by the same shares: 12.3% and 19.4% at $s = 0.79$, and 14.9% and 23.0% at $s = 0.66$, the upper ends of the ranges quoted in the paper.

Table H.1: What the Intervals Exclude Under Parallel Trends, and per User Who Can See Ads

Panel A: conversion rate, percentage points			
	Estimate (SE)	95% interval	Excluded decline
Headline, re-centered on the pooled pre-launch average	−0.024 (0.053)	[−0.128, 0.080]	0.128 pp (5.8% of 2.21%)
referenced to week −1	−0.075 (0.083)	[−0.237, 0.086]	0.237 pp (10.7% of 2.21%)
Panel B: referral volume, count model			
	Estimate (SE), log points	95% interval	Excluded decline
Headline, no growth differential	−0.061 (0.037)	[−12.6%, +1.2%]	12.6%
growth differential of 6%/yr removed	−0.075 (0.037)	[−13.8%, −0.2%]	13.8%
growth differential of 12%/yr removed	−0.088 (0.037)	[−14.9%, −1.5%]	14.9%

Table H.1: What the Intervals Exclude Under Parallel Trends, and per User Who Can See Ads (continued)

Panel C: excluded decline per user who can see ads, headline bounds divided by the eligible share s

	Conversion	Volume
s = 1.00 (all referral sessions from users who can see ads)	5.8%	12.6%
s = 0.95 (paying users use ChatGPT as much as free users)	6.1%	13.2%
s = 0.79 (five times as much)	7.3%	15.6%
s = 0.66 (ten times as much)	8.8%	18.5%

Panel D: the headline bounds against the placebo distribution (95th percentile of the absolute placebo estimates)

	Placebo yardstick	Excluded decline	Placebo draws below the model's bound
Conversion rate, pp	0.192	0.216 pp (9.8% of 2.21%)	37 of 500
Referral volume, count model	0.110 log points	15.7%	0 of 100
per user who can see ads, s = 0.79		0.273 pp (12.3% of 2.21%) / 19.4%	
per user who can see ads, s = 0.66		0.330 pp (14.9% of 2.21%) / 23.0%	

NOTES: Estimates and website-clustered standard errors from Tables F.1 and F.3. Excluded decline: the lower limit of the two-sided 95% interval; declines larger than it lie outside the interval. Conversion is expressed as a share of the pre-launch organic ChatGPT conversion rate of the treated units under the frozen weights, 2.21%; volume as $100[1 - \exp(\text{lower limit})]$. Panel C: s is the share of referral sessions from users who can see ads, computed from a 5% paying-user share and their usage relative to free users, unrounded; the log effect (volume) or the effect in points (conversion) and its standard error are divided by s before conversion, the rescaling of Figure 3 of the paper (Web Appendix D); dividing the percentage effect by s instead gives 19.2% in the last row. Panel D: the estimate minus the 95th percentile of the absolute placebo estimates of Table F.2 (500 conversion draws, 100 volume draws), and the number of draws at or below the model-based bound; the last two rows divide these placebo-yardstick bounds by the eligible shares of Panel C at five and ten times as much use, conversion then volume, the per-user figures the main text quotes next to the model-based ones.

Web Appendix I Validity and Robustness

This section reports the checks behind the pre-trends and samples paragraph of the Condition A section of the paper: pre-launch trends, placebo rollouts, and the samples without one cohort. Unless stated otherwise, conversion refers to the frozen-weight model of Table F.3 and volume to the count model of Table F.1; the sensitivity to the largest websites is in Table F.1, Panel D.

Pre-launch trends. The joint test of the eleven pre-launch coefficients gives $p = .458$ for conversion (Wald 10.8 on 11 degrees of freedom) and $p = .109$ for volume (Wald 17.0), and p-values between .144 and .789 for every cohort's conversion path except the United Kingdom's ($p < .001$); removing the United Kingdom moves the pooled conversion estimate by 0.01 percentage points (Table I.1; Table F.3, Panel B). The cohorts' own volume paths depart from zero more often: the United States at $p < .001$, the United Kingdom at $p = .002$, Japan and Korea at $p = .021$, Brazil and Mexico at $p = .016$, and Australia, Canada, and New Zealand at $p = .185$, while the pooled path gives $p = .109$. The test has little power against a smooth growth differential of the size considered in Web Appendix G (Web Appendix E), and its calibration with five treated cohorts is imperfect: for volume it returns $p < .05$ in 49 of 100 EU pseudo-rollouts.

Placebo rollouts. Table F.2 re-assigns the five launch weeks and the real cohort sizes to countries drawn from the untreated EU and re-estimates both headline models. Of 100 volume pseudo-rollouts, 10 produce a decline at least as large as the actual −5.9% and 26 an effect at least as large in absolute value; of 500 conversion pseudo-rollouts, 187 produce a decline at least as large as the actual −0.024 points and 378 an effect at least as large in absolute value. Read as shares of the draws, these are placebo p-values under the assumption that pseudo-rollouts among EU countries are comparable to the real rollout (Web Appendix E). The 5th to 95th percentile bands are −0.082 to +0.081 log points and −0.15 to +0.16 points.

Without one cohort. Dropping the United States cohort moves volume to −5.8% (Table F.1, Panel B) and moves conversion to −0.003 points (Table F.3, Panel B); dropping any other single

Table I.1: Joint Pre-Trend Tests, Pooled and by Cohort

	Conversion rate	Referral volume, count model
	Wald χ^2 (df) [p]	Wald χ^2 (df) [p]
Pooled (all cohorts)	10.82 (11) [.458]	16.95 (11) [.109]
US	7.12 (11) [.789]	36.19 (11) [<.001]
AU–CA–NZ	7.85 (11) [.727]	14.95 (11) [.185]
UK	34.79 (11) [<.001]	29.08 (11) [.002]
JP–KR	15.91 (11) [.144]	22.48 (11) [.021]
BR–MX	8.65 (11) [.654]	23.28 (11) [.016]
Slope of the cohort-averaged pre-launch path, weeks −12 to −2 (per week)		
	Percentage points	Log points
Estimated slope (SE)	−0.0074 (0.0061)	0.0031 (0.0036)
Smallest detectable slope (80% power, 5% size)	0.0170	0.0101
Slope added by a growth differential of 6% / 12% per year	—	0.0011 / 0.0022

NOTES: Joint Wald tests that the eleven pre-launch coefficients (event weeks −12 to −2; week −1 is the reference) of the event studies are zero, website-clustered covariance. Conversion: the frozen-weight model of Table F.3, on the cohort-averaged path and on each cohort's own path. Volume: the count model of Table F.1, cohort-averaged and per cohort. The test has little power against a smooth growth differential of the size considered in Web Appendix G (Web Appendix E): the slope rows give the least-squares slope of the cohort-averaged pre-launch path with a delta-method standard error, the smallest slope the test would detect at 80% power, and the slope a differential of 6% or 12% per year adds, $\log(1+g)/52$; with five treated cohorts the test also over-rejects: for volume it rejects in 49 of 100 EU pseudo-rollouts.

country moves conversion by at most 0.02 percentage points, the largest change, for the Netherlands, being 0.016 points (the leave-one-country run is not tabulated).

Comparison channel. With direct traffic instead of organic search as the comparison channel, the count model gives −0.087 (0.057), an implied −8.4% with a 95% interval from −18.0% to +2.4% (Table F.1, Panel B), a decline of the same order as the headline and no more precise; a spillover of the ads into organic search would have to move the two comparisons apart, and it does not.

Web Appendix J Estimation of Advertiser Value

The channel comparison behind Figure 4 and the contrasts in the Condition B section is estimated on the financial sample of 102 websites. The design follows the channel comparison of Kaiser and Schulze (2026), which measured organic ChatGPT referrals against the established channels with the same outcome models. This paper tightens the fixed effects to website-by-device and device-by-week and weights the models by sessions. A cell is one website (one analytics view), one visitor-country scope (the United States against all other countries), one device, one week, and one channel, and enters when it records at least one session; scope defines the cells but enters no fixed effect, so each channel coefficient pools US and non-US cells within a website and device. Each model carries website-by-device and device-by-week fixed effects, so it adjusts for persistent differences between websites and devices and for device-specific calendar-week changes, and every coefficient is a difference from direct traffic, the omitted channel, for the five channels organic ChatGPT, paid ChatGPT, direct, organic search, and paid search. Conversion is modeled as a quasi-binomial logit of transactions per session weighted by sessions; its coefficients are log-odds, the model's own comparison scale, and not percentage-point differences. Revenue per session is modeled by least squares weighted by sessions, on values winsorized at the 1st and 99th percentiles within channel. Average order value is revenue per transaction on cells with at least one transaction, winsorized within channel and weighted by transactions. Conversion and order value help interpret the revenue gap; their adjusted estimates are separate models and do not reconstruct it exactly. Standard errors are clustered by website. All 102 websites contribute paid ChatGPT observations and 101 contribute paid search observations. The paid ChatGPT against paid search differences quoted in the main text, for conversion, order value, and revenue per session, are contrasts of the two channel coefficients from the same fitted model with delta-method standard errors. The level figures, \$0.89 and \$2.45 per session and 0.37% and 1.17% conversion, are pooled totals of revenue, transactions, and sessions in US traffic across the same sample, channel by channel; because advertisers differ in basket size, the pooled and the within-website order-value comparisons can differ in sign, and Web Appendix K shows both.

Bounce rate, pageviews per session, and session duration use the full paid referral panel of 387 websites with the same weights and clustering and with the device-by-week effects interacted with the data feed, with or without financial tracking; bounce rate is a session-weighted logit, the other two are session-weighted least squares on per-session means. Coefficients are relative to direct traffic (Figure 4, D to F).

The breadth statistics of the Condition B section describe the same shortfall advertiser by advertiser. Because ads ran in the treated countries and paid ChatGPT traffic concentrates in the United States, 154,000 of 194,000 sessions, they are computed in US traffic over the full window, on the same basis as the level figures and the parity price. Each advertiser's paid ChatGPT and paid search totals are compared within the website, which requires both channels in US traffic: 87 of the 102 advertisers qualify, and for them we report the share earning less per paid ChatGPT visit than per paid search visit, the share converting worse, and the share with no paid ChatGPT transaction. The concentration of paid ChatGPT purchases is the dominant advertiser's share among the 88 websites with US paid ChatGPT traffic; across all visitor countries, over all 102, it is 66% instead of 70%. The zero-purchase calibration asks how many of the 87 would show no paid ChatGPT purchase if their paid ChatGPT visits had converted at their own US paid-search rate: treating purchases as independent, the probability that an advertiser with n paid ChatGPT sessions and paid-search conversion rate p records none is $(1 - p)^n$; the expected count is the sum of these probabilities over the 87 advertisers, and its standard deviation follows from the same independent draws, conditional on the observed rates (Table K.1, Panel C). The shares are not a US artifact: computed across all visitor countries for the 101 advertisers with both channels anywhere, they are 96%, 97%, and 82% against 97%, 98%, and 84% (Table K.1, Panel A). Because one advertiser carries most paid ChatGPT transactions, the three paid ChatGPT against paid search contrasts, which come from the channel models fitted on all 102 websites and all markets, are re-estimated dropping one of the 102 advertisers at a time.

The parity price is the price at which a paid ChatGPT click would return as much revenue per dollar as a paid search click. It multiplies the median paid-search cost per click of the 16

US-dominant cost-linked accounts (Web Appendix A), taken as the benchmark price for the whole sample, by the ratio of pooled revenue per session in US traffic, paid ChatGPT over paid search, from the level figures above: \$0.72 times \$0.89 divided by \$2.45 gives \$0.26. In symbols,

$$p^{*} = p_{\text{search}} \times \frac{\text{RPS}_{\text{ChatGPT}}}{\text{RPS}_{\text{search}}} = 0.721 \times \frac{0.888}{2.449} = 0.262, \tag{4}$$

rounded to \$0.26, where p_{search} is the median paid-search cost per click of the 16 US-dominant cost-linked accounts and RPS is pooled revenue per session in US traffic for each paid channel; the going price of \$1.72 is 6.57 times p^{*}, more than six. The parity price assumes similar profit margins in both channels and counts only the tracked visit. Advertiser-level uncertainty comes from 2,000 resamples, with replacement, of the 102 financial-sample websites that record paid ChatGPT or paid search sessions in US traffic, each website's records kept together; each resample recomputes the pooled revenue ratio, and the click price is held at its median, giving \$0.05 to \$0.57.

The matched comparison cited in the Condition B section asks whether a paid ChatGPT click brings value through channels the campaign did not buy. It pairs each advertiser with a similar website that does not advertise on ChatGPT and compares the two websites' traffic before and after the advertiser's first paid ChatGPT referral. All traffic is US traffic. Controls come from the full feed after excluding every website that ever received paid ChatGPT traffic and every website linked to an advertiser through a shared analytics view or brand, which leaves 3,575 candidate websites. Websites are matched on two features of their US organic ChatGPT traffic in the twelve weeks before a reference week: its level (log sessions) and its growth (log change from the first to the second six weeks). Each advertiser is paired with the closest unused candidate that has the same dominant visitor country, without replacement and only if the standardized distance is at most one; advertisers without such a candidate stay unmatched. Matching runs twice. At the US launch week (February 9, 2026), 324 of the 387 advertisers find a match; only these enter the design. Each of them is then matched again at its own first paid-referral week, with the features measured over the twelve weeks before that week and standardized among the advertisers and candidates of that week;

an advertiser without organic ChatGPT traffic in those twelve weeks is dropped. This second match succeeds for 305 of the 324 advertisers. For each pair, weekly sessions from twelve weeks before to twelve weeks after the advertiser's first paid referral are stacked, with weeks without sessions set to zero, and a Poisson model compares the advertiser's change with its control's change. Fixed effects for website by device and for pair by calendar week by device absorb level differences and common shocks within each pair; standard errors are clustered by website. The model is run separately for the four channels the campaign did not buy: organic ChatGPT, direct, organic search, and paid search (Table N.1). Growth in these channels would be consistent with paid clicks bringing additional visits through other routes. Because advertisers choose when to start, the design cannot rule out that the decision to advertise and the traffic change share a cause; it can show patterns consistent with additional traffic but cannot establish it.

The one public advertiser account (Opascope 2026) is placed in the panel's distribution by computing for each advertiser the same conversion rate and revenue per paid ChatGPT session and comparing them with the account's reported 2.35% conversion rate and its implied $2.56 per click (1.49 dollars of revenue per dollar of spend at a $1.72 click price); a website matches when it meets both values, and the table also reports the matches with at least 100 paid ChatGPT sessions. The reconciliation is in Web Appendix L.

The financing arithmetic of the Discussion, revenue per free user, the cost of free access, and the ad load required at a given click-through rate and the parity price, is set out with its sources in Web Appendix P; beyond the parity price it uses no estimate from this study.

For Condition B, standard errors are clustered by website with 102 clusters in the financial models and 387 in the engagement models; intervals are pointwise and no adjustment is made for the number of channels or outcomes compared. The leave-one-advertiser-out estimates and the advertiser bootstrap above describe sampling variability across advertisers; they do not address advertiser self-selection.

Table J.1: Channel Comparison: Conversion Rate, Average Order Value, and Revenue per Session

	Conversion rate	Average order value	Revenue per session
Panel A: adjusted differences from direct traffic within website, device, and week			
	(log-odds)	(USD per transaction)	(USD per session)
Organic ChatGPT	0.123 (0.135)	-34.7 (10.1)	0.24 (0.33)
	$p = .368$	$p < .001$	$p = .469$
Paid ChatGPT	-1.818 (0.701)	-78.7 (18.3)	-2.86 (1.19)
	$p = .011$	$p < .001$	$p = .018$
Organic search	-0.073 (0.077)	-14.1 (6.0)	-0.08 (0.13)
	$p = .343$	$p = .021$	$p = .542$
Paid search	0.047 (0.128)	-4.9 (6.8)	0.47 (0.35)
	$p = .714$	$p = .472$	$p = .182$
Paid ChatGPT minus paid search	-1.866 (0.644)	-73.8 (16.9)	-3.33 (1.11)
exact p-value of the contrast	$p = .005$	$p < .001$	$p = .003$
Observations (cells)	86,056	50,460	86,056

Panel B: pooled levels, ratios of totals

	Conversion rate		Average order value		Revenue per session	
	US	All	US	All	US	All
Direct	0.86%	0.99%	$197	$166	$1.70	$1.65
Organic ChatGPT	1.22%	1.31%	$268	$189	$3.27	$2.48
Paid ChatGPT	0.37%	0.31%	$240	$236	$0.89	$0.73
Organic search	0.68%	1.08%	$185	$136	$1.25	$1.46
Paid search	1.17%	1.33%	$209	$154	$2.45	$2.05

NOTES:

Panel A: website × visitor-country scope × device × week cells; every coefficient is the difference from direct traffic, the omitted channel, with website × device and device × week fixed effects. Conversion is a session-weighted quasi-binomial logit; average order value is transaction-weighted least squares on cells with at least one transaction, winsorized at the 1st and 99th percentiles within channel; revenue per session is session-weighted least squares, winsorized within channel. The paid ChatGPT minus paid search row is the contrast of the two channel coefficients from the same model with delta-method standard errors. Standard errors clustered by website in parentheses; the line below each estimate gives its exact two-sided p-value.

Panel B: pooled totals across the same 102 websites, channel by channel, without adjustment. Because advertisers differ in basket size, the pooled and the within-website order-value comparisons can differ in sign: in raw US totals the paid ChatGPT basket is larger than the paid search basket, within the same website it is smaller.

Web Appendix K Channel Performance

Table J.1, Panel A, gives the differences from direct traffic adjusted for website-by-device and device-by-week effects for the 102 websites of the financial sample, and the contrast between the two paid channels that the Condition B section of the paper quotes. Panel B gives the pooled levels in US traffic behind the paper's \$0.89 against \$2.45 per session and 0.37% against 1.17% conversion. The pooled order values, \$240 for paid ChatGPT against \$209 for paid search, run the other way from the within-website contrast of −\$74; the two summaries estimate different quantities, one across advertisers and one within them, and the paper quotes the within-website one. Table K.1 carries the breadth statistics of the paper: of the 87 advertisers buying both paid channels in US traffic, 84 earn less per paid ChatGPT visit than per paid search visit, 85 convert paid ChatGPT visits at a lower rate, and 73 record no paid ChatGPT transaction at all; the 14 that do carry all 570 transactions, and the 73 without one had 18,726 paid ChatGPT sessions between them. Across all visitor countries the picture is the same: of the 101 advertisers with both channels anywhere, 97 earn less, 98 convert worse, and 83 record no paid ChatGPT purchase. Panel C of the same table collects the sparsity counts, the dominant-advertiser check, the zero-purchase calibration, and the parity interval (Web Appendices L and M). Table K.2 gives the engagement estimates of Figure 4, D to F, on the paid referral panel of 387 websites.

Table K.1: Breadth of the Shortfall Across Advertisers, the Dominant Advertiser, and the Derived Price Ratios

Panel A: advertisers buying both paid channels, compared website by website

	US traffic	All visitor countries
Advertisers with both paid channels	87	101
Earn less per paid ChatGPT visit than per paid search visit	84 (97%)	97 (96%)
Convert paid ChatGPT visits at a lower rate than paid search visits	85 (98%)	98 (97%)
No paid ChatGPT transaction over the window	73 (84%)	83 (82%)
their paid ChatGPT sessions	18,726	23,885
At least one paid ChatGPT transaction	14 (16%)	18 (18%)
their paid ChatGPT transactions	570	603
Paid ChatGPT sessions of these advertisers	154,022	194,362

Panel B: derived ratios quoted in the paper

	Value	
Revenue per session, paid ChatGPT over paid search (US)	\$0.89 / \$2.45 = 0.36	a 64% shortfall
Going price over the median paid-search click price	\$1.72 / \$0.72 = 2.38	
Going price over the parity price	\$1.72 / \$0.262 = 6.57	more than six

Table K.1: Breadth of the Shortfall Across Advertisers, the Dominant Advertiser, and the Derived Price Ratios (continued)

Panel C: sparsity, the dominant advertiser, the zero-purchase calibration, and the parity interval (financial sample, 102 websites)

	All countries	US traffic
Advertisers with paid ChatGPT sessions / with a purchase	102 / 18	88 / 14
Dominant advertiser's share of paid ChatGPT purchases / revenue	66.0% / 73.2%	69.8% / 76.0%
	Full sample	Dominant advertiser dropped
Paid ChatGPT − paid search, conversion (log-odds)	−1.87 (0.64)	−2.83 (0.31)
Paid ChatGPT − paid search, revenue per session (USD)	−3.33 (1.11)	−4.31 (1.15)
Paid ChatGPT − paid search, order value (USD)	−73.8 (16.9)	−33.9 (18.5)
Largest p-value over the 102 single-advertiser drops: conversion / revenue / order value	.015 / .007 / .070	
Zero-purchase advertisers among the 87, observed / expected at own paid-search rates	73 / 42.3 (SD 2.4)	
Parity price, 2,000 advertiser resamples, click price fixed at \$0.72	\$0.26 [\$0.05, \$0.57]	

NOTES: Panel A: each advertiser's paid ChatGPT and paid search totals over the window, compared website by website, in US traffic (the paper's basis, where ads ran longest and the level figures and the parity price are computed) and across all visitor countries. Panel B: the level figures of Table J.1, Panel B, the median paid-search click price of the 16 US-dominant cost-linked accounts, and the going price of the largest disclosed account. Panel C: counts over the financial sample; the contrasts are the paid ChatGPT minus paid search differences of the channel model of Table J.1 (website-by-device and device-by-week fixed effects, session-weighted for conversion and revenue per session and transaction-weighted for order value, website-clustered standard errors in parentheses), on the full sample and with the dominant advertiser removed; the largest p-value row gives, for each contrast, the maximum of its exact two-sided p-values over the leave-one-advertiser-out re-estimates. Zero-purchase calibration: the expected number of the 87 US dual-channel advertisers with no paid ChatGPT purchase if each converted its paid ChatGPT sessions at its own paid-search rate, purchases treated as independent (Web Appendix J). Parity interval: 95% percentile interval of the recomputed revenue ratio over 2,000 resamples, with replacement, of the 102 financial-sample websites that record paid ChatGPT or paid search sessions in US traffic, each website's records kept together, times the fixed click price.

Table K.2: Regression Results: Engagement Outcomes (Paid Referral Panel)

	Dependent variables		
	Bounce Rate	Pageviews/Session	Duration/Session
Organic ChatGPT	−0.499 (0.091)	0.348 (0.161)	94.9 (16.5)
	$p < .001$	$p = .031$	$p < .001$
Paid ChatGPT	0.462 (0.145)	−1.391 (0.339)	−104.2 (28.7)
	$p = .002$	$p < .001$	$p < .001$
Organic search	−0.366 (0.136)	0.624 (0.264)	92.7 (25.1)
	$p = .008$	$p = .019$	$p < .001$
Paid search	−0.500 (0.128)	0.765 (0.259)	110.2 (19.6)
	$p < .001$	$p = .003$	$p < .001$
Fixed Effects: Site×Device, Device×Week×Feed	Yes	Yes	Yes
R^2 (least-squares models)	—	0.786	0.835
Observations	302,261	302,261	302,261
Websites	387	387	387

NOTES: Paid referral panel of 387 websites, with and without financial tracking in their analytics. All coefficients relative to direct traffic. Bounce rate: quasi-binomial logit (log-odds). Pageviews/session and duration/session: session-weighted WLS, outcome winsorized at the 1st/99th percentile within channel. Websites without financial tracking carry no financial metrics, so conversion and revenue per session are not estimable in this sample. Week FE are interacted with the data feed (with vs. without financial tracking) to absorb feed-specific weekly shocks; the feed main effect is absorbed by the site FE. Cluster-robust SEs at site level (site = hostname for financially tracked websites, analytics view ID otherwise), in parentheses; the line below each estimate gives its exact two-sided p-value. Pseudo-R^2 not suitable for model comparison in quasi-likelihood frameworks.

Web Appendix L Prices and the Parity Price

The median paid-search click price of the 16 US-dominant accounts of the cost sample is $0.72 over the window, close to the $1.01 median account price on Google's search network across 40,183 accounts before the ad launch (Web Appendix A). The parity price multiplies the $0.72 by the ratio of revenue per session in US traffic, 0.36, giving $0.26; the going price of $1.72 is 2.38 times the search price and 6.57 times the parity price (Table K.1, Panel B). Resampling advertisers with replacement, 2,000 draws with the click price held at its median, gives $0.05 to $0.57 (Panel C); one advertiser carries 76% of US paid ChatGPT revenue, so resamples with and without it differ widely. The resampling draws, with replacement, from the 102 websites of the financial sample with paid ChatGPT or paid search sessions in US traffic. Table N.2 places the largest disclosed account, 2.35% conversion and about $2.56 per visit at $1.72 per click, in the panel: of the 102 websites with paid ChatGPT sessions, 18 record any transaction, the median website records none, the 90th percentiles are 0.14% conversion and $0.25 per visit, and one website meets both thresholds, on fewer than 100 paid ChatGPT visits. The largest paid account, with 66% of all paid ChatGPT transactions, converts 0.76% and earns $1.99 per visit.

Web Appendix M Advertiser Heterogeneity

Paid ChatGPT traffic is sparse and concentrated (Table K.1, Panel C): all 102 websites of the financial sample have paid ChatGPT sessions and 18 have a purchase; in US traffic, 88 and 14, and the dominant advertiser holds 66% of paid ChatGPT purchases, 70% in US traffic, and 76% of US paid ChatGPT revenue. Dropping each of the 102 advertisers in turn leaves the conversion and revenue contrasts against paid search in place in every case; across the 102 re-estimates the largest p-values are .015 for conversion, .007 for revenue per session, and .070 for order value. Dropping the dominant advertiser deepens the first two, to −2.83 log-odds ($p < .001$) and −\$4.31 per session ($p < .001$), and halves the order-value gap to −\$33.9 (standard error 18.5, $p = .070$); the exact p-values behind the main text's paid ChatGPT estimates are printed in Table J.1. Had every dual-channel advertiser's paid ChatGPT visits converted at its own paid-search rate, 42.3 of the 87 US dual-channel advertisers would show no purchase, with a standard deviation of 2.4; 73 do.

Web Appendix N Matched Comparisons at Advertiser Entry

Table N.1 reports the one-to-one matched design on the traffic of the four channels the campaigns did not buy: each advertiser is matched at its first paid-referral week to a website without paid ChatGPT referrals, and weekly US sessions are compared over the twelve weeks before and after entry, with website-by-device and pair-by-calendar-week-by-device fixed effects and website-clustered standard errors, on 305 entry-matched pairs. The organic ChatGPT estimate is +8% and not distinguishable from zero; direct, organic search, and paid search traffic show no detectable change. The intervals remain compatible with increases of 43% in organic ChatGPT, 28% in direct, 27% in organic search, and 23% in paid search traffic: they bound modeled traffic associations, not all indirect merchant value, and the design does not remove the advertiser's own decision to advertise. The design thus gives no indication that buying ChatGPT ads raises a website's organic ChatGPT referrals, the counterpart of the separation between paid and organic results in search advertising; with a smallest detectable increase of about 50%, however, it can rule out only large lifts, not moderate ones.

Table N.1: Entry-Matched Advertiser Associations Around the First Paid Referral

	Organic ChatGPT	Direct	Organic search	Paid search
All entry-matched pairs (305 pairs)				
Advertiser × Post (DiD)	0.079 (0.142)	-0.027 (0.139)	0.022 (0.112)	-0.095 (0.152)
p-value	.578	.845	.848	.534
Implied volume effect	8.2%	-2.7%	2.2%	-9.0%
95% CI (implied %)	[-18%, 43%]	[-26%, 28%]	[-18%, 27%]	[-32%, 23%]
MDE (80% power)	+49%/−33%	+47%/−32%	+37%/−27%	+53%/−35%
Observations	20,920	24,334	24,290	21,344
Fixed Effects: Site×Device, Pair×Week×Device	Yes	Yes	Yes	Yes

NOTES: Advertisers with observed paid ChatGPT referrals are matched one-to-one, without replacement, to websites with no paid ChatGPT referrals, at the advertiser's first paid-referral week, on organic ChatGPT volume during the twelve weeks before (weeks −12 to −1) and on its change between the two six-week halves, exactly on the dominant visitor country and within a caliper of 1 standardized distance (features standardized within each entry week's pool; advertisers without a control inside the caliper stay unmatched). PPML stacks event weeks −12 to +12 around entry on zero-filled weekly US sessions with website×device and pair×calendar-week×device fixed effects. First referral is an outcome-defined proxy, not campaign start: zero-click campaigns have no clock, and later weeks need not contain an active or clicked ad. Effects equal $\exp(\widehat{\beta}) - 1$; MDEs use 80% power and 5% two-sided size. Estimates are descriptive; SEs are website-clustered, in parentheses, with the exact two-sided p-value of each coefficient on the line below.

Table N.2: Paid ChatGPT Performance Across Advertiser Websites Against the Largest Disclosed Account

	Conversion rate	Revenue per visit
Largest disclosed account (Opascope 2026): about $60,000 over 15 days, $1.49 per dollar of spend at $1.72 per click	2.35%	about $2.56
Median website	0.00%	$0.00
90th percentile across websites (each column separately)	0.14%	$0.25
Largest paid account (52,232 sessions; 66% of transactions)	0.76%	$1.99
Websites with any paid ChatGPT transaction	18 of 102	
Websites meeting both observed thresholds, any volume	1 of 102 (1.0%)	
with at least 100 paid ChatGPT sessions	0 of 102	

NOTES: The 102 websites of the financial sample with any paid ChatGPT sessions; conversion and revenue per visit are ratios of each website's paid ChatGPT totals over the window across all visitor countries. A website matches the account when it meets both of the account's values; the last row counts the matches with at least 100 paid ChatGPT sessions. The one website below that volume that matches does so on 9 paid ChatGPT sessions. The largest disclosed account reported about $60,000 of spend over 15 days, a 2.35% conversion rate and $1.49 of revenue per dollar of spend at $1.72 per click, about $2.56 per visit (Opascope 2026). Transactions per session need not equal the share of distinct visitors purchasing.

Web Appendix O What We Estimate and What the Data Can Show

Condition A. The analysis behind Condition A is causal: it estimates what the introduction of ads did to organic ChatGPT referrals, using the staggered rollout as a natural experiment. The treatment is the introduction of ChatGPT advertising in a country in a given week. Throughout, a market is a visitor country, and countries assigned the same launch week form a cohort (Web Appendix B). The estimates measure the average effect of that introduction on the organic ChatGPT referrals the panel websites receive from visitors in that country, whether or not the referring user saw an ad, an intention-to-treat effect. The effect is measured by comparing the change in launch countries with the change in the untreated European Union, and organic ChatGPT referrals with organic search referrals to the same website; the assumptions behind these comparisons are stated in Web Appendix D. We observe the rollout by country and know which subscription tiers carry ads; we never observe whether an individual user was shown one.

The two dependent variables of Condition A capture the observable margins of organic referral performance. Referral volume is the number of organic ChatGPT sessions a website receives from visitors in a country. Conversion rate is transactions per organic ChatGPT session, and describes purchasing among the visits that arrive, whose composition can itself change. Together they show whether fewer visits reach websites and whether those that arrive still buy. Neither measures user trust or behavior inside ChatGPT directly.

Condition B. The channel comparison behind Condition B is descriptive: adjusted for website-by-device and device-by-week effects, it compares conversion, order value, revenue per session, and engagement between sessions arriving through paid ChatGPT and through paid search or direct visits. It shows the observed outcomes of sessions in these channels, not what an advertiser would have earned by buying one channel instead of another. The decision to advertise is the advertiser's own, not random, and a website becomes visible as an advertiser only once it records a paid-tagged

ChatGPT visit, so entry is dated at that first observed visit, which can lag the campaign start.

The matched comparisons at entry, which the Condition B section cites for value that a paid click may bring through other channels, ask whether an advertiser's traffic gains are confined to the channel it bought and whether they appear at entry. They can show patterns consistent with additional traffic; they cannot separate such traffic from selection, and they are not an experiment on advertiser returns.

Financing arithmetic (the Discussion of the main text). The parity price and the financing calculations, revenue per free user against the cost of free access and the ad load that would close the gap, are arithmetic on the Condition B contrasts and on public price, usage, and cost figures. They are illustrative calculations under stated assumptions, not estimates from the rollout and not predictions (Web Appendix P).

Web Appendix P Financing Arithmetic

This section sets out the arithmetic behind the Discussion of the paper with its sources (Table P.1). Beyond the parity price it uses no estimate from this study.

Per free user. OpenAI reported $1 billion in projected annualized advertising revenue less than 200 days after launch, in the same announcement that puts ChatGPT at more than 1 billion weekly active users (OpenAI 2026g); about 95% of them do not pay (O'Brien 2026), so the free base is about 950 million users, and the $1 billion over it is $1.05 per free user per year, rounded to $1 in the paper, averaged over all markets including those without ads. The inference cost of a free user is about $5 per year (Muppidi and Palazzolo 2026). The 2030 projection of $102 billion in advertising revenue (Muppidi 2026) over about 2.52 billion free users, the 2.75 billion weekly users OpenAI's April 2026 forecast assumes for 2030 (Fried 2026) net of the 8.5% subscriber share of its November 2025 forecast (Muppidi 2025), is $40.48 per free user per year, rounded to $40.

Ads per commercial prompt. A free user sends about seven prompts a day, the CFO's figure for free users specifically, with the paid tiers at two, three, and eleven times that (Friar 2026), a fifth of them with commercial intent (Adams 2026), 511 commercial prompts a year. Revenue per free user per year is the product of commercial prompts, ads per commercial prompt, the click-through rate, and the value per click. Within each scenario the paper holds the last two fixed, at the third-party click-through estimate of 0.50% (Mordecai 2026) and the parity price of $0.262, and lets ad load vary; holding the click-through rate constant as ad load rises is a conservative assumption, and the results are average ad units per commercial prompt, not a placement on every prompt. One ad on every commercial prompt then earns $0.67 per free user per year; the $5 cost takes 7.5 ads per commercial prompt, more than seven in the paper, and the $40 takes 59.8, about 60 (Table P.1, Panel B, on the rounded targets). Panel B varies the click-through rate over the range that single advertisers and third parties report, 0.35% to 1.30% (Horwitz 2026, Fletcher 2026, Deda 2026b); at the most favorable rate the $5 still takes 2.9 ads per commercial prompt and the $40 takes 23.0.

Today's ad load. Today's $1 at the reported account price of $1.72 per click (Opascope 2026) is 0.58 clicks per free user per year, which at a 0.50% click-through rate is 0.23 ad units per commercial prompt; at the parity price the same $1 would take 1.50 (Panel B), the difference being the price per click. The paper's statement that about one in four commercial prompts carries an ad rests on the observed sample: ads on 25.9% of 50,006 commercial US prompts in July 2026, one ad slot per answer (Deda 2026a). The arithmetic figure is close to it but combines global revenue, one account's price, and a third-party click-through rate, and it equals prompt coverage only if ad-bearing answers carry one unit each.

Table P.1: The Financing Arithmetic of the Discussion: Inputs, Sources, and Ads per Commercial Prompt

Panel A: inputs and derived quantities

	Value	Source
Prompts per free user per day	7	Friar (2026)
Share of prompts with commercial intent	20%	Adams (2026)
Commercial prompts per free user per year ($7 \times 365 \times 0.20$)	511	derived
Value of a paid ChatGPT click at parity	\$0.262	Web Appendix L
Click-through rate of ChatGPT ads, third-party estimate, June 2026	0.50%	Mordecai (2026)
Advertising revenue per free user today[a]	\$1.05, rounded to \$1	OpenAI (2026g); O'Brien (2026)
Inference cost of a free user per year[b]	\$5	Muppidi and Palazzolo (2026)
Revenue per free user in the 2030 projection[c]	\$40.48, rounded to \$40	Muppidi (2026)
Reported account price per click	\$1.72	Opascope (2026)
Revenue per free user per year at one ad per commercial prompt ($511 \times 0.005 \times 0.262$)	\$0.67	derived
Ad units per commercial prompt[d]	0.23	derived
Share of commercial prompts carrying an ad, observed (50,006 US prompts, July 2026)	25.9%	Deda (2026a)

Table P.1: The Financing Arithmetic of the Discussion: Inputs, Sources, and Ads per Commercial Prompt (continued)

Panel B: ad units per commercial prompt needed at the parity price and the rounded targets, by click-through rate

Click-through rate	for today's $1	for $5	for $40
0.35%	2.14	10.7	85.5
0.50%	1.50	7.5	59.8
1.30%	0.58	2.9	23.0

[a] $1 billion in projected annualized advertising revenue divided by about 950 million free users: more than 1 billion weekly active users, about 95% of them non-paying. This averages projected advertising revenue over all free users, including users in markets without ads. At the 90% free share cited by the CFO in September 2026 (CNBC 2026a), the corresponding figure is $1.11; the rounding to $1 is unchanged.

[b] The $5 figure is an estimate of annual inference cost, not the full cost of serving a free user.

[c] $102 billion in projected 2030 advertising revenue divided by about 2.52 billion free users. The free-user count is constructed from a projection of 2.75 billion weekly users in 2030 (Muppidi and Palazzolo 2026, Fried 2026) net of the 8.5% subscriber share in the earlier forecast (Muppidi 2025).

[d] As implied by today's $1 at $1.72 per click ($(1/1.72)/(511 \times 0.005)$).

NOTES: Revenue per free user per year = commercial prompts per year × ads per commercial prompt × click-through rate × value per click. Within each scenario, the click-through rate and the value per click are held fixed and ads per commercial prompt, an average over commercial prompts, is the free variable; holding the click-through rate constant as ad load rises is an assumption. Ads per commercial prompt counts ad units per prompt; OpenAI's documentation allows one or more units per response and one or more items per unit (OpenAI 2026a). Panel B click-through rates are one advertiser's own account, 0.35% (Horwitz 2026); Similarweb's June 2026 estimate, 0.50%, the rate used in the paper (Mordecai 2026); and SE Ranking's own campaigns, 1.30% (Deda 2026b). Adthena's pilot sample, 0.92% (Fletcher 2026), and Similarweb's May 2026 figure, 0.68% (Mordecai 2026), lie inside that range; no platform-wide rate is published. The parity price is $0.262 = $0.72 × 0.363 (Table K.1).