Do Open-Weight LLM Releases Affect Web Traffic to Proprietary AI Chatbots? Evidence from Llama 3.1

Jikhan Jeong[1]

Extended Abstract

In recent years, generative artificial intelligence (Gen AI) has been rapidly adopted in daily life and across a wide range of fields, including economics and business. Among Gen AI applications, proprietary AI chatbots have expanded substantially in terms of market size, user adoption, and technical capabilities since the launch of OpenAI's ChatGPT in November 2022. Within this market, ChatGPT has remained a dominant player despite the entry of several competing proprietary AI chatbots, including Microsoft Copilot, Google Gemini, Anthropic Claude, and others. These proprietary chatbots are built on pre-trained large language models (LLMs) and delivered to end users through web- or mobile-based user-interfaces (UIs). However, despite their widespread use and market importance, proprietary chatbots remain largely closed systems: users cannot directly access the architecture of the underlying pre-trained LLMs, the pre-training dataset, or the model weights.

In contrast, open-weight LLMs make their model weights publicly available, allowing users to download, run, and fine-tune the models in local computing environments. A well-known example is Meta's Llama family of open-weight models (Touvron et al. 2023). Researchers, developers, and other users have widely adopted Llama and fine-tuned these models on their own data for specific purposes. As a result, open-weighted LLMs can be adapted to user-specific tasks and provide a foundation for developing new LLM models and applications. Indeed, many derivative LLM models have been built on or adapted from Llama across various fields and tasks.

However, relatively few studies have examined how open-weight LLMs affect consumer demand for proprietary AI chatbots. This is partly because external users, including researchers, have limited access to internal user-level data. In addition, some leading AI firms, such as OpenAI and Anthropic, remain privately held as of today and are therefore not subject to the same mandatory public reporting requirements as publicly traded companies.

To address these challenges, prior studies have relied on alternative approaches. First, some AI chatbot firms and AI service providers examine user behaviors using proprietary internal data. Second, AI firms and researchers collaborate to conduct digital experiments based on internal data. Third, some researchers conduct digital, lab, or survey experiments using recruited participants. Finally, other studies use web traffic data to examine demand in AI chatbot markets and assess their impact on users, businesses, industries, and society.

However, to the best of the author's knowledge, no prior study has empirically investigated the impact of the release of open-weight LLMs on web traffic and unique visitors to proprietary AI chatbots. Addressing this research question is important because it helps improve our understanding of: (1) how global web traffic to proprietary AI chatbots changes following the

[1] Assistant professor of economics, Jaggi School of Business, Missouri University of Science and Technology, Rolla, MO 65409, USA. E-Mail: jikhan.jeong@mst.edu

release of open-weight LLMs; (2) how competitive market dynamics evolve after such release. More broadly, open-weight LLMs may or may not attract more attention to LLMs and create low-cost outside options for users, thereby increasing competitive pressure on proprietary AI chatbot providers, motivating firms' strategic responses in pricing, product differentiation, marketing, and innovation.

To fill this gap, this paper empirically examines the impact of releases in Meta's Llama 3.1 on global visits and unique visitors to proprietary AI chatbots using Similarweb data and two empirical approaches: a synthetic different-in-difference (SDID; Arkhangelsky et al. 2021) and event-study SDID (Ciccia et al 2024; Ciccia 2024).

The main empirical results show that the estimated average treatment effect on treated (ATTs) of the Llama 3.1 releases are statistically significant at the 5% level for both dependent variables: log visits and log unique visitors to proprietary AI chatbot websites.

Importantly, the release of an open-weight LLM increases visits and unique visitors to proprietary AI chatbots built on closed-weight LLMs, rather than reducing visits and unique visitors to them through substitution. One plausible explanation is that the release of an open-weight LLM may increase overall attention to LLMs, including proprietary AI chatbots, among users. Some of this increased attention may then shift toward proprietary AI chatbots because of their relative advantages, such as simpler and better user interfaces (UIs).

The SDID estimators used in this paper under the simple adoption design can be disaggregated into dynamic treatment effects under an event-study framework (Ciccia 2024). Accordingly, the following subsection estimates dynamic treatment effects on treated using event-study SDID and examines whether users of proprietary AI chatbot websites may respond to the Llama 3.1 release with a delay

The empirical results from the event-study SDID show that the estimated effects of the Llama 3.1 release on visits and unique visitors to proprietary AI chatbots begin to increase from the four and five weeks after the release week, respectively. Overall, these results suggest that users of proprietary AI chatbots may respond positively to the Llama 3.1 releases with a delay.

This subsection presents additional platform-specific analyses for individual LLM-related platforms using SDID and event-study SDID, since SDID can be implemented even when there is only a single treated unit, whereas traditional DiD cannot.

One dominant player among proprietary AI chatbots is chatgpt.com. Individual users can access and use ChatGPT as a conversational AI service through chatgpt.com on desktops or mobile devices. In contrast, advanced users, such as developers, researchers, and firms, access OpenAI's developer platform through platform.openai.com. Users visit platform.openai.com primarily to obtain and use their Application Programming Interface (API) keys for specific generative AI models, rather than to interact with ChatGPT through a chatbot interface.

The empirical results show that the estimated ATTs of the Llama 3.1 releases are statistically significant for both dependent variables: log visits and log unique visitors to chatgpt.com, but statistically insignificant for platform.openai.com. In addition, the event-study SDID results show that the estimated effects of the Llama 3.1 release on visits and unique visitors to chatgpt.com and platform.openai.com begin to increase from a few weeks after the release week.

Further platform-specific analyses show that the estimated ATTs of the Llama 3.1 releases are statistically significant for both log visits and log unique visitors to perplexity.ai (an AI answer engine that retrieves relevant information), but statistically insignificant for google.com (a dominant search engine). In particular, perplexity.ai incorporated Llama 3.1 on the same date as the Llama 3.1 release date. Accordingly, users of perplexity.ai show relatively quicker responses in visits than users of other domains, while showing no delayed response in unique visitors.

Additional platform-specific analysis shows that the Llama 3.1 release increases visits and unique visitors to huggingface.co (an open-source AI model and dataset platform). In contrast, the estimated effects are statistically insignificant results for both visits and unique visitors to github.com (a software development and coding platform) and stackoverflow.com (a knowledge sharing platform for programming).

The robustness analysis section uses traditional differences-in-difference (DiD) and event-study DiD, and conducts sensitivity analyses with a longer pre-trend period. The empirical results from these robustness analyses also support the finding that visits and unique visitors to proprietary AI chatbots increase with a delay following the release.

Overall, the release of open-weight LLMs may increase visits and unique visitors to proprietary AI chatbots with a delay, rather than reducing demand for these services. One plausible conjecture is that open-weight LLM releases may increase overall attention to LLMs and attract more web traffic to proprietary AI chatbots. Accordingly, the release of an open-weight LLM increases visits and unique visitors to AI answer engines and open-source model platforms, while being statistically insignificant for search engines, coding platforms, and knowledge-sharing platform.

The approaches and findings can help market participants, including developers, researchers, AI firms, and policymakers, to better understand the impact of open-weight LLMs on LLM markets and their dynamics.

Declaration of generative AI and AI-assisted technologies

OpenAI's ChatGPT, Microsoft's Copilot, and Grammarly were used to polish and correct grammar in the manually written draft to improve the quality of the manuscript. OpenAI's ChatGPT was used in some cases to assist with coding.

Reference

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. American economic review, 111(12), 4088-4118.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). LLaMA: Open and efficient foundation language models. arXiv. https://doi.org/10.48550/arXiv.2302.13971

Ciccia, D. (2024). A short note on event-study synthetic difference-in-differences estimators. arXiv preprint arXiv:2407.09565.

Ciccia, D., Clarke, D., & Pailañir, D. (2024). SDID_EVENT: Stata module providing synthetic

difference-in-differences (SDID) event-study estimators. Statistical Software Components S459354, Boston College Department of Economics.

Clarke, D., Pailanir, D., Athey, S., & Imbens, G. (2024). On synthetic difference-in-differences and related estimation methods in Stata. The Stata Journal, 24(4), 557-598.