Attention: Your Conversational Data is What They Need

Anonymous Authors[1]

1. Introduction

The launch of ChatGPT, the first consumer-facing large language model (LLM), made the often praised AI revolution suddenly tangible for the general public. With its remarkable ability to provide tailored responses, ChatGPT took users by storm (Wu et al., 2023). Many attributed its usability to the integration of supervised fine-tuning and reinforcement learning from human feedback applied to the underlying foundation model (Ouyang et al., 2022). With around 100 million weekly active users (OpenAI, 2024a), OpenAI now possesses an extensive volume of conversational data, which could be valuable for fine-tuning.

However, recent research has shown that LLMs can not only memorize and leak data from pre-training datasets (Nasr et al., 2023) but also from fine-tuning datasets (Borkar, 2023). Some companies had already feared that proprietary firm information could be leaked and restricted the use of ChatGPT (Tilley & Kruppa, 2023). Moreover, the question of how users' conversational data can be used for additional model training is a topic of heated debate, as demonstrated by the recent backlash from Slack users (Belanger, 2024).

In this work, we discuss how EU regulations view the use of users' conversational data for improving LLMs. By conversational data, we refer to the prompts and responses generated by users during interactions. Specifically, we examine how ChatGPT navigates these legal frameworks. Following a temporary ban by the Italian Data Protection Agency, OpenAI introduced a right for users to opt out of usage for model training purposes (Chiara, 2023).

2. Background

When ChatGPT was launched in November 2022, conversations could be used by OpenAI for further model training, as was already the case with InstructGPT (Ouyang et al., 2022). Users were not asked for their consent, but OpenAI affirmed that they would remove personally identifiable information from the data they used. The first change came for data submitted through the API. In December 2022, OpenAI made it possible to opt out by sending an email.[1] In February 2023, they also offered a Google Form for submitting a "Data Opt Out Request". In March 2023, OpenAI announced that they would no longer use the API data for training, except when organizations explicitly opted in.

With conversational data in ChatGPT, which is the focus of this work, the situation is different. In March 2023, the Italian Data Protection Agency temporarily blocked ChatGPT, asserting that Italian users should have a right to opt out from their data being used (Garante, 2023). On April 25, 2023, ChatGPT introduced the option to deactivate `Chat History & Training` in the settings (OpenAI, 2023). Before this, OpenAI reverted again to using a Google Form, which is quite suprising for a company valued at over 30 billion USD at that time (Glasner, 2023). On April 28, 2023, ChatGPT was again accessible to Italian users, although the investigation is ongoing (Garante, 2024).

Since then, the opt-out has become established, and in early 2024, OpenAI has introduced temporary chats, allowing users to disable data usage for model training in a session-based window (OpenAI, 2024b). However, not everything was completely smooth. Between January and April 2024, numerous users reported a bug that could only be resolved by disabling the opt-out setting (OpenAI Community, 2024). As OpenAI did not quickly fix this issue, it led to speculation about whether this friction was intentional to coerce users into sharing their data.

3. Legal Contextualization

In the nascent days of ChatGPT, the regulation of (generative) AI was still in its infancy. Yet, in Europe, OpenAI did not operate in a legal vacuum. Using conversational data from ChatGPT users likely involves the processing of personal data, which falls within the scope of the EU's General Data Protection Regulation (GDPR). Since the GDPR has extraterritorial application, OpenAI cannot avoid compliance by claiming it is only a US-based company while also

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

[1]This section's timeline is partially derived from data on archived OpenAI websites in the Internet Archive.

serving European users.[2]

A key provision of the GDPR is that the processing of personal data requires a legal basis.[3] Interestingly, OpenAI does not rely on consent, but on legitimate interest as detailed in section 9 of their privacy policy (OpenAI, 2024c). According to recent case law of the European Court of Justice, product improvement may indeed be considered a necessary legitimate interest.[4] It is expected that European Data Protection Agencies will challenge this (EDPB, 2024). To use legitimate interest as a legal basis, it is also required that OpenAI's interests outweigh those of its users. This balancing will depend on three factors. First, the need for real user data instead of data from employees, contractors or synthetic data, and whether this complies with data minimization principles. Second, what specific safeguard OpenAI has implemented to remove personal data and prevent personal data from leakage. Third, the reasonable expectations of the affected users (Veil, 2018). Another complication could arise regarding whether legitimate interest is even applicable when users share sensitive data in conversations, such as when a user chats about health issues.[5]

In addition, the legal basis of legitimate interest requires that users be informed and given the option to opt-out.[6] From January to April 2024, users on various online platforms, including the OpenAI Community Forum, reported a `Conversation key not found` error after they opted out and used ChatGPT (OpenAI Community, 2024). The bug could be resolved by turning the chat history back on. OpenAI's delayed response to this issue led to speculation that the friction was intentional to coerce users into sharing their data. Creating friction in users' ability to exercise their rights is not a new phenomenon online. This practice is often referred to as *dark patterns* and is well studied with cookies (Luguri & Strahilevitz, 2021). It would also not be the first instance of such tactics being used to collect data for training purposes (Potoroaca, 2024). The legal qualification of this issue will be determined on a case-by-case basis. However, in a regulatory investigation, OpenAI would need to provide truthful information about whether there were technical reasons for the delayed response.[7]

The new provisions for foundation models under the EU's AI Act are expected to become applicable from July 2025. In Chapter 5, the EU has introduced new rules for General Purpose AI models (GPAIs) that will apply alongside the GDPR rules described before.[8] It is likely that OpenAI's flagship model will be qualified as a *Systemic risk* GPAI because the floating-point operations for training exceeded 10^{25} (Epoch AI, 2024) or due to other criteria such as the high number of users.[9] Despite this, the AI Act does not require opting in for the use of data for model improvement. However, it imposes documentation duties on GPAIs vis-à-vis the Office for AI and providers of AI systems regarding the training of their models (Friedl & Gasiola, 2024).[10]

4. Discussion

As LLMs are getting widely used and mulitmodal (Stanford University, 2024), more conversational data is being generated. Although the pre-training and copyright law receive much scholarly attention, the issue of using conversational data to improve models is becoming increasingly important. In the case of ChatGPT, there is currently no legal certainty regarding the use of conversational data for further model training. The ongoing investigations of the Data Protection Agencies are expected to provide clearer guidance. As of now, we can state that silently collecting and using data is likely to violate EU law. If a company is unwilling to ask for consent, it should ensure that there is an opt-out option that is communicated to users and works without friction.

In the law and tech literature the so-called pacing problem describes how the law often lags behind technological advancements (Marchant, 2011). However, this is only one part of the story, especially in the European context. As the example of ChatGPT shows, there is also a *racing* problem. ChatGPT was rapidly made available worldwide, adhering once again to the Silicon Valley philosophy of "move fast and break things". With the first race now over and many new providers entering the market, users will hopefully have options where their data is not silently used by default.

[2] Art. 3(2) GDPR. For improved readability, references to laws will be found in footnotes.

[3] Art. 6(1) GDPR.

[4] Meta Platforms Inc. and Others v. Bundeskartellamt, Case C-252/21, ECLI:EU:C:2023:537, para. 122 (2023) ("it cannot be ruled out from the outset that the controller's interest in improving the product or service with a view to making it more efficient and thus more attractive can constitute a legitimate interest capable of justifying the processing of personal data and that such processing may be necessary in order to pursue that interest").

[5] Art. 9(1) GDPR.

[6] Art. 21 GDPR.

[7] Art. 31 GDPR.

[8] Recital 10 of the AI Act.

[9] Art. 51(2) and 51(1)(b) AI Act and its Annex XIII.

[10] Art. 53 and 55 AI Act.

References

Belanger, A. Slack defends default opt-in for ai training on chats amid user outrage, 2024. URL https://arstechnica.com/tech-policy/2024/05/slack-defends-default-opt-in-for-ai-training-on-chats-amid-user-outrage/.

Borkar, J. What can we learn from data leakage and unlearning for law? *arXiv preprint arXiv:2307.10476*, 2023.

Chiara, P. G. Italian DPA v. OpenAI's ChatGPT: The reasons behind the investigations and the temporary limitation to processing. *Journal of Law and Technology*, 2023.

EDPB. Report of the Work Undertaken by the ChatGPT Taskforce, 2024.

Epoch AI. Epoch AI Database, 2024. URL https://epochai.org/data/epochdb/visualization?startDlEra=1950-12-28.

Friedl, P. and Gasiola, G. G. Examining the EU's Artificial Intelligence Act, 2024. URL https://verfassungsblog.de/examining-the-eus-artificial-intelligence-act/.

Garante. Temporary limitation of processing of italian users' data against OpenAI, 2023. URL https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870847#english.

Garante. ChatGPT: Italian DPA notifies breaches of privacy law to OpenAI, 2024. URL https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9978020#english.

Glasner, J. Startup valuation fluctuations, 2023. URL https://news.crunchbase.com/venture/startup-valuation-fluctuations-ai-openai-msft-eoy-2023/.

Luguri, J. and Strahilevitz, L. J. Shining a light on dark patterns. *Journal of Legal Analysis*, 13(1):43–109, 2021.

Marchant, G. E. *The Growing Gap Between Emerging Technologies and the Law*, pp. 19–33. 2011.

Nasr, M., Carlini, N., Hayase, J., Jagielski, M., Cooper, A. F., Ippolito, D., Choquette-Choo, C. A., Wallace, E., Tramèr, F., and Lee, K. Scalable extraction of training data from (production) language models. *arXiv preprint arXiv:2311.17035*, 2023.

OpenAI. New ways to manage your data in ChatGPT. 2023. URL https://openai.com/index/new-ways-to-manage-your-data-in-chatgpt/.

OpenAI. Start using ChatGPT instantly, 2024a. URL https://openai.com/index/start-using-chatgpt-instantly/.

OpenAI. Memory and new controls for ChatGPT, 2024b. URL https://openai.com/index/memory-and-new-controls-for-chatgpt/.

OpenAI. Privacy policy, 2024c. URL https://openai.com/policies/privacy-policy/.

OpenAI Community. Chat history off - conversation key not found error, 2024. URL https://community.openai.com/t/chat-history-off-conversation-key-not-found-error/594342.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., and Lowe, R. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744, 2022.

Potoroaca, A. Facebook will soon use posted content to train its ais, opting out requires a form, 2024. URL https://www.techspot.com/news/103179-facebook-soon-use-posted-content-train-ais-opting.html.

Stanford University. The AI Index Report, 2024. URL https://aiindex.stanford.edu/report/.

Tilley, A. and Kruppa, M. Apple restricts use of ChatGPT, joining other companies wary of leaks, 2023. URL https://www.wsj.com/articles/apple-restricts-use-of-chatgpt-joining-other-companies-wary-of-leaks-d44d7d34?

Veil, W. Einwilligung oder berechtigtes Interesse?: Datenverarbeitung zwischen Skylla und Charybdis. 2018.

Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L., and Tang, Y. A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, 10(5):1122–1136, 2023.