The American Journal of Interdisciplinary Innovations and Research
ISSN 2642-7478 | Open Access

Type Original Research
PAGE NO. 108-115
DOI 10.37547/tajiir/Volume07Issue09-11

Use Of User Feedback for Adaptive Model Tuning

Nisarg B Shah
Product Manager | AI/ML Product Development Seattle, USA

Abstract- This paper discusses a possible path toward adaptive finetuning of large-scale language models over user signal continual learning. In our study, we are trying to organize an approach for explicit and implicit channels. Inside channel heterogeneous feedback filtering, interpreting, and integrating all of them into one regular tuning cycle that would help keep the model updated and qualitative in real-time usage. This paper validates these claims with studies of how fast static model parameterizations get outdated on one hand, and an observation limitation from classic offline process drops in answer accuracy and user trust on the other hand. This unification is novel because it unifies three classes of feedback into a multi-objective loss function with dynamic weights thereof; implemented through a microservice hierarchy architecture that logs, streams filtering, anonymizing, annotating data — then trains in several stages including supervised fine-tuning (SFT) and reinforcement learning from human feedback (RLHF) plus contextual bandit plus rolling A/B test with confidence bounds. In fact, after going through several iterations of SFT and RLHF, it is the live model that steadily beats by a good margin all static baselines in terms of human preference. At the same time, the contextual bandit reduces average regret in online mode, and scaling to billions of queries is achieved without loss of metadata integrity or update flexibility. Key challenges are identified: catastrophic forgetting of rare skills, narrow-group preference bias, privacy risks when processing live data, and high manual annotation costs, for which regularization, stratified sampling, differential privacy, and active self-evaluation learning are proposed as solutions. This article should interest and benefit those who investigate and architect systems for natural language, machine learning, and recommendation engines.

OPEN ACCESS

SUBMITTED 01 August 2025
ACCEPTED 12 August 2025
PUBLISHED 28 September 2025
VOLUME Vol.07 Issue 09 2025

CITATION
Nisarg B Shah. (2025). Use Of User Feedback for Adaptive Model Tuning. The American Journal of Interdisciplinary Innovations and Research, 7(09), 108–115. https://doi.org/10.37547/tajiir/Volume07Issue09-11

Keywords: Fine-tuning adaptation, Feedback from user, Process of continual learning, large language models, multi-stage retraining

Introduction

By February 2025, LLMs will have been a ubiquitous tool: with approximately 400 million weekly active users at ChatGPT by that time and counting—an enormous volume of queries, responses, and reactions hitherto leveraged only in the most fragmentary manner (Langley, 2025). At this volume of interactions, the accumulation of outdated facts and shifts in user preferences occur more rapidly than the development cycle of the next large model version concludes, so that a once-fixed parameterization loses relevance within mere weeks.

This static nature leads not only to factual inaccuracies but also to a decline in trust: when a model repeats stale information, users perceive it as frozen. Even state-of-the-art post-training with human feedback delivers only limited gains to core capabilities: in the GPT-4 report, the difference between base and RLHF-enhanced versions on an examination dataset was just 0.3 percentage points—73.7% versus 74.0%—therefore underscoring the need for other adaptation mechanisms.

One such mechanism is continual learning from streams of user signals, as multi-stage fine-tuning as new data pile up delivers a significant accuracy jump after seven iterations with no loss of the model's overall knowledge, which was proven viable by an update cycle in the NAACL 2025 study (Guan et al., 2025).

The incredible flood of feedback, joined with process limitations that may be considered the classic offline approach, conspires to create such an evident demand for adaptive methods. This paper asks how signals from users might be selected, understood, and used in the normal course of tuning to finally retune the focus: not to think of LLM as an artifact fixed in time but rather as a system that learns with its audience and keeps pace within a swiftly changing information environment.

Materials and Methodology

This study analyzes fifteen sources, including academic articles, industry reports, implementation examples, and regulatory documents. The theoretical foundation comprised works on the scale and dynamics of user signals: the evaluation of ChatGPT's active audience in February 2025 (Langley, 2025), the analysis of the limited gain after RLHF in GPT-4 (OpenAI, 2023), experiments on multi-stage fine-tuning with new data (Guan et al., 2025), and methods for dynamic traffic redistribution based on the upper confidence bound (Ye et al., 2024).

To identify the properties of different feedback types, we conducted a comparative analysis of explicit, implicit, and indirect signals: the semantics of textual reviews and ratings (Cooper & Zafiroglu, 2024); behavioral metrics—scroll depth, session duration, number of clicks (Haruyama & Hidaka, 2023; Covington et al., 2016); and data on legal claims and regulatory inquiries as sources of indirect signals (Liang et al., 2025; OpenAI, 2024).

The architectural review of the adaptive cycle therefore embraced streaming pipelines for gathering and sifting data in YouTube recommender systems (Covington et al., 2016), microservice schemas with Safe RLHF classifiers (Dai & Pan, 2024), privacy protection mechanisms founded on the PII benchmark (Open Review, 2024), and differential privacy techniques (Huang, 2025).

A content analysis of real-world case studies depended on data about active ChatGPT users (Reuters, 2025), practice of iterative fine-tuning on user signals with measurements of accuracy gain (Ouyang et al., 2022; Guan et al., 2025), and quantitative reliability analysis of iterative interactions in a streaming scenario (Sguerra et al., 2025).

Results and Discussion

The classical division of user signals into explicit, implicit, and indirect categories becomes the cornerstone of adaptive tuning, as it enables the coherent integration of heterogeneous sources of information on model response quality. In the context of large language models, whose behavioral dynamics change more rapidly than those of traditional recommender systems, such a typology bridges the earlier motivation for continual learning presented in the Introduction and the practice of daily weight updates.

Explicit signals are those generated at the user's initiative, such as thumbs-up/thumbs-down ratings, five-point scales, or extended textual comments.

Despite their high informativeness, such data remain extremely sparse: in commercial services, a dense feedback matrix is considered one in which only about 1% of possible entries are filled (Haruyama & Hidaka, 2023). Moreover, the ChatGPT interface shows that among active respondents, 84.2% have submitted a rating at least once, yet the vast majority limit themselves to binary likes, while only 29% provide extended text (Cooper & Zafiroglu, 2024). Thus, explicit feedback is semantically rich but limited in coverage and susceptible to selection bias, tending to originate from technically savvy and motivated users.

Implicit feedback is based on behavioral metrics recorded automatically and covering the entire traffic. These include reading time, number of clicks, session interruptions, and scroll sequences. The scale of the differences is impressive: YouTube training pipelines process hundreds of billions of such events in each model iteration, as shown in Figure 1 (Covington et al., 2016).

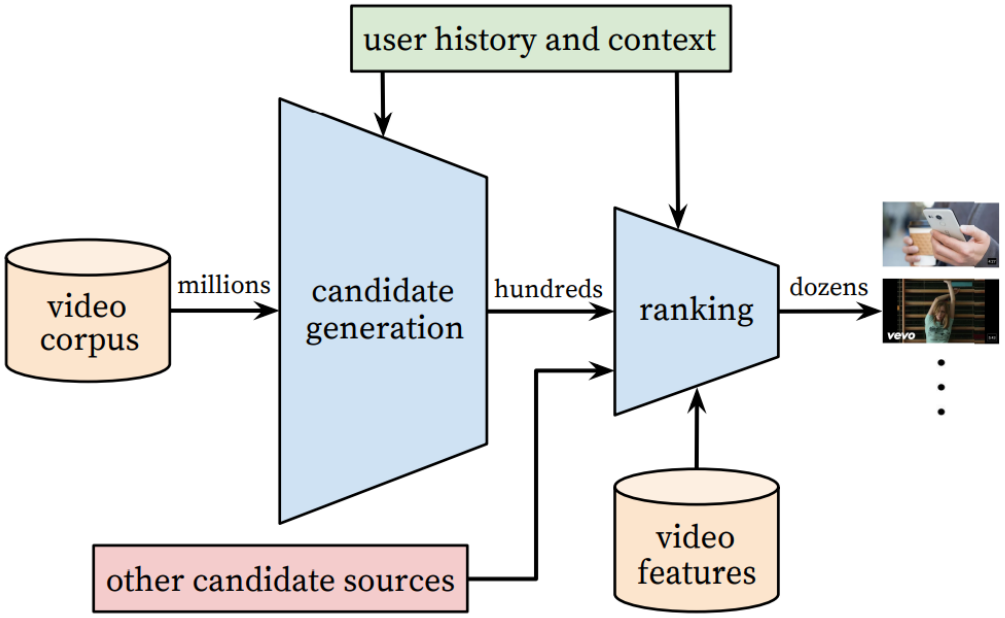

Fig. 1. Recommendation system architecture demonstrating the funnel where candidate videos are retrieved and ranked before presenting only a few to the user (Covington et al., 2016)

In music streaming, repeat listens form statistically significant patterns that must be assigned reliability coefficients due to high noise levels (Sguerra et al., 2025). Signal redundancy helps to smooth measurement errors, but requires Bayesian confidence estimates and regularization; otherwise, the model rapidly overfits to the short tail of anomalous actions.

Indirect feedback occurs when the user does not rate the response directly but turns to a third party: files a complaint, reports a violation, or initiates a legal claim. Thus, the corpus of financial claims in the United States for 2022–2024 comprises 687,241 records, and already 18% of the text was generated with LLM participation, opening the possibility of training models on real examples of conflict cases (Liang et al., 2025). From a regulatory standpoint, in the first half of 2024, the OpenAI platform received 29 requests from authorities for non-content data and 8 for content, disclosing information on 60 accounts in total, as shown in Fig. 2 (OpenAI, 2024).

Диаграмма

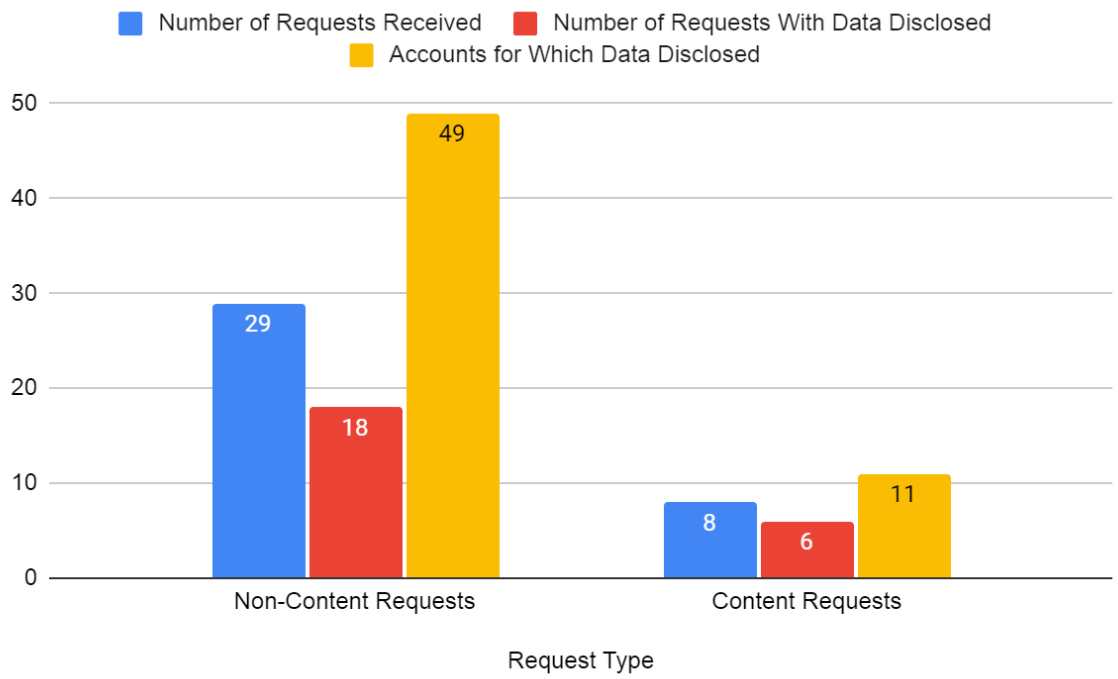

Fig. 2. Data Disclosure by Request Category (OpenAI, 2024)

Studies conducted concurrently indicate that automatic moderation classifiers founded on LLMs attain about 95% precision in spotting forbidden content, lessening the load of manual scrutiny and hastening the response to complaints (Huang, 2025).

So explicit signals give quality but are sparse, implicit signals give scale but are noisy, and indirect signals give the compulsory regulatory and ethical context for legal compliance. A best-fit adaptive tuning hybridizes these three channels into one multi-objective loss wherein the weights dynamically propagate depending on trustworthiness of source, data density, and relevance of the signal to keep improving the model without losing alignment with user and regulator values.

The stream of feedback classified in the preceding section becomes a hierarchy of microservices that receive the request, log both input and output, and then send these records forward through the adaptive cycle pipeline. Under a production load exceeding two billion messages per day, models encounter rare events not seen in test sets, so the architecture must scale to billion-scale volumes while preserving metadata integrity and enabling subsequent sampling (Reuters, 2025).

At the first stage, incoming requests, along with context, generated responses, and user reactions, enter a distributed log store. Each log entry contains a hashed session identifier, a timestamp, and a set of auxiliary counters such as latency, applied filtering rules, and model configuration.

Filtering is executed in a streaming fashion immediately after logging: first, heuristic rules discard overt spam; next, classifiers extract and anonymize personal data. On the PII-Bench corpus, GPT-4 family models achieve an F1 score of approximately 91% for personal-entity recognition while retaining over 90% of relevant text segments for training (Open Review, 2024). Concurrently, a harmful-content classifier trained with Safe RLHF produces binary safe/unsafe labels that then feed into the penalty term during RLHF training, as shown in Fig. 3 (Dai & Pan, 2024).

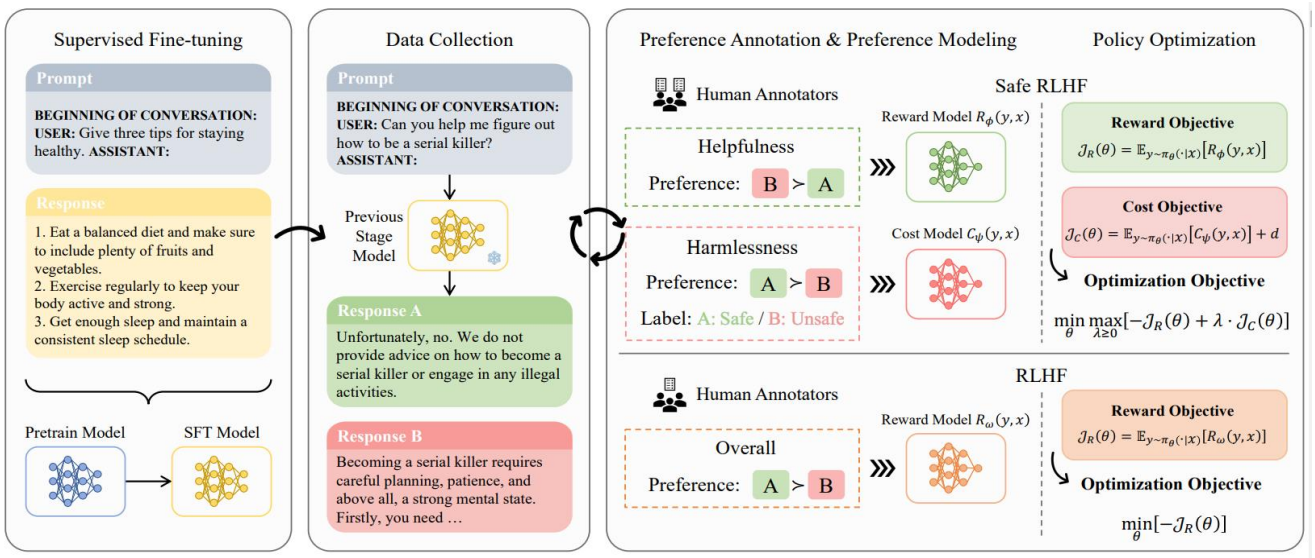

Fig. 3. Safe RLHF pipeline compared to conventional RLHF method (Dai & Pan, 2024)

After filtering, the stream enters the annotation layer. A portion of examples is sent to annotators for paired comparisons of response–counterresponse: the preference method proved more stable and, according to RLAIF data, approximately ten times less expensive than classical rating schemes without loss of correlation with full-scale human evaluation (Lee et al., 2023). The remaining bulk proceeds through a semi-automated loop, where the models themselves act as primary critics, marking obvious errors or generating self-reflective confidence labels.

The collected data then enters the training phase. The update begins with a rapid SFT iteration employing low-rank adaptation, followed by RLHF based on proxy rewards on the same batch. Finally, a contextual bandit operates over the last thirty minutes of traffic, which, in online mode, injects small gradients to reduce average regret. Initial InstructGPT experiments demonstrated that even a 1.3 B model trained in this cycle outperforms the base GPT-3 175 B in eighty-five percent of cases by human preference, confirming the efficacy of combining SFT and RLHF (Ouyang et al., 2022).

The final layer handles validation and release. Each new version undergoes a rolling A/B test, with part of the traffic retained on the previous model; simultaneously, the LOLA code dynamically redistributes requests between versions according to the upper confidence bound rule, reducing cumulative regret on clicks and drop-offs compared to static traffic allocation (Ye et al., 2024). Additionally, in real time, the policy failure rate, aggregate risk, and toxicity delta are computed; if any metric exits its confidence interval, the release is automatically blocked pending manual audit. Such multi-tiered checks close the cycle, allowing the model to be updated multiple times per day without compromising safety or quality.

In chat systems, the feedback stream is particularly dense, so the combined cycle of collection, filtering, training, and validation runs almost continuously. New weights are integrated via minimal changes to adapter matrices, while the base model remains stable. This process allows factual errors to be corrected immediately after users flag them, while preserving a unified response style established by prior supervised fine-tuning. No service interruptions occur, as version switching is managed via proxy routing that maintains session consistency.

In search and recommender systems, feedback arrives mainly implicitly through clicks, result skips, and changes in interaction time. Contextual bandits estimate expected satisfaction before a click, enabling dynamic reordering of document lists in favor of fresh and personally relevant items. When the user later confirms or refutes the system's choice, this signal is fed at high frequency into the training block, where models are updated on narrow domains without a full recomputation of the main corpus.

In robotics, users rarely provide direct ratings; however, sensors record the success or failure of an action as a regret signal. If a manipulator fails to grasp an object with the required force, a local critic issues a negative reward and initiates a small on-device fine-tuning step.

Thanks to parametric adapters stored in non-volatile memory, a few such episodes rapidly shift the policy toward more reliable trajectories, and the cloud instance of the model receives generalised patches at the next synchronization.

These examples demonstrate that, regardless of domain, a unified scheme of signal filtering, light fine-tuning, and strict online validation ensures sustained improvement in utility without risking safety degradation, transforming the system from a static encyclopedia into a living organism that learns alongside its users and hardware environment.

Constant adjustment brings real advantages; at the same time, it makes much deeper fundamental problems worse if they are not handled systematically before they turn the sound effects of training fast into quality going down and more danger for the user.

The initial danger relates to disastrous forgetting, that is, the replacement of rare but essential abilities because of new signals in prevailing tasks. When the model gets a flow of similar instances, gradients move the weight allocation toward more frequent patterns, resulting in the capability connected to less common domains beginning to decay. Erosion may be delayed by regularizers that preserve important parameters and by replaying historical sessions mixed into each new mini-batch; selective freezing of layers sensitive to vanishing features also proves beneficial.

The second problem manifests as preference bias when the most active contributors belong to narrow user categories, for example, technical enthusiasts or speakers of a particular cultural context. Their reactions enter the reward subsystem with greater density and begin to redefine usefulness criteria for all other groups. Alignment is achieved through stratified sampling of signals—allocating quotas to each demographic segment—and by applying confidence weights that reduce the influence of accounts that accumulate large numbers of ratings suspiciously quickly.

The third obstacle involves privacy and regulatory constraints, as training on live interactions almost inevitably touches personal data. Passive anonymization of text is insufficient when rare combinations of facts enable indirect user identification. The solution relies on multi-tiered filtering and differential privacy, whereby the contribution of each session is masked with random noise and access to non-excess metadata is granted only to authorized services.

Finally, annotation cost remains a significant barrier. Manual labeling of preferences, especially paired comparisons, demands skilled labor and scales with model complexity. Cost reduction is achieved via active learning, where the system selects the most informative examples for humans and labels less critical cases using its critic. Another advantage accrues from model self-evaluation, which auto-rejects responses with visible logical disparities, thus conserving the annotators' time. All the above challenges are intrinsic to any system striving to glean knowledge from user experience; therefore, answerers must be situated within the very fabric of the adaptive cycle architecture and not appended as external patches after the fact.

Conclusion

This paper will show that lifelong learning from streams of user signals is a practical evolutionary path for large language models, changing them from static artifacts into dynamic systems capable of appropriate responses to changes in their external environment. The key element of the proposed approach is the classification of feedback into explicit, implicit, and indirect signals, which allows coherent integration of heterogeneous data into a single multi-objective loss with dynamic weights based on source reliability and feedback density. Practical implementation through a hierarchy of microservices—encompassing logging, filtering, annotation, training, and validation—enables scaling to billion-scale query volumes without loss of metadata integrity or update flexibility; a rolling A/B test with confidence bounds for key metrics ensures safety and quality of each new model version.

Preliminary results, including the original InstructGPT experiments and a multi-stage fine-tuning comparison, validate this cycle. Within just a few rounds of SFT and RLHF, the live model begins to sustain an uplift in human preference over static baselines as well as in reduced average regret for the contextual bandit when run in online mode. Implementing such a process, however, does require tackling key issues: catastrophic forgetting of rare skills under pressure by common patterns; preference bias for small demographic groups; privacy, if on live user data; and manual annotation cost.

Regularizers and historical session replay, signal sampling stratified by type of signals, multi-tier filtering with differential privacy, and active learning with model self-assessment, too, are what it takes to work around the challenges mentioned above.

This adjustment towards achieving relevance is at the core of a new paradigm in the evolution of LLM, whereby relevance and quality responses can be sustained over time as dynamism in the information landscape takes place. It unavoidably necessitates the urgency and timeliness of scalability and security at every stage of the cycle. Successful realization of such systems will transform large language models into living organisms that learn alongside their audience and hardware environment, while preserving safety, privacy, and a balance of interests for all ecosystem participants.

References

1. Cooper, N., & Zafiroglu, A. (2024). Constraining Participation: Affordances of Feedback Features in Interfaces to Large Language Models. *Arxiv*. https://doi.org/10.48550/arxiv.2408.15066
2. Covington, P., Adams, J., & Sargin, E. (2016). Deep Neural Networks for YouTube Recommendations. *Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16*, 191–198. https://doi.org/10.1145/2959100.2959190
3. Dai, J., & Pan, X. (2024). *Safe RLHF: Safe Reinforcement Learning From Human Feedback*. https://openreview.net/pdf?id=TyFrPOKYXw
4. Guan, C., Huang, C., Li, H., Li, Y., Cheng, N., Liu, Z., Chen, Y., Xu, J., & Liu, J. (2025). Multi-Stage LLM Fine-Tuning with a Continual Learning Setting. *Findings of the Association for Computational Linguistics: NAACL 2022*, 5484–5498. https://doi.org/10.18653/v1/2025.findings-naacl.303
5. Haruyama, M., & Hidaka, K. (2023). What influences users to provide explicit feedback? A case of food delivery recommenders. *User Modeling and User-Adapted Interaction*, *34*, 753–796. https://doi.org/10.1007/s11257-023-09385-8
6. Huang, T. (2025, July 19). *Content moderation by LLM: from accuracy to legitimacy*. Artificial Intelligence Review. https://link.springer.com/article/10.1007/s10462-025-11328-1
7. Langley, H. (2025, March 5). *ChatGPT isn't slowing down Google yet — these numbers prove it*. Business Insider. https://www.businessinsider.com/chatgpt-isnt-slowing-down-google-ai-search-overviews-2025-3
8. Lee, H., Phatale, S., Mansoor, H., Lu, K., Mesnard, T., Bishop, C., Carbune, V., & Rastogi, A. (2023, September 1). *RLAIF vs. RLHF: Scaling Reinforcement Learning from Human Feedback with AI Feedback*. Arxiv. https://doi.org/10.48550/arXiv.2309.00267
9. Liang, W., Zhang, Y., Codreanu, M., Wang, J., Cao, H., & Zou, J. (2025). The Widespread Adoption of Large Language Model-Assisted Writing Across Society. *Arxiv*. https://doi.org/10.48550/arxiv.2502.09747
10. Open Review. (2024). *PII-Bench: Evaluating Query-Aware Privacy Protection Systems Anonymous ACL submission*. https://openreview.net/pdf/38ef9fbb478ee9e9e1964beadbc3e029aef8f2c3.pdf
11. OpenAI. (2023). GPT-4 Technical Report. *Arxiv*. https://doi.org/10.48550/arxiv.2303.08774
12. OpenAI. (2024). *OpenAI Report on Government Requests for User Data*. OpenAI. https://cdn.openai.com/trust-and-transparency/report-2024h1-government-requests-for-user-data.pdf
13. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A. K., Schulman, J., Hilton, J. K., Kelton, F., Miller, L. P., Simens, M., Askell, A., Welinder, P., Christiano, P. F., Leike, J., & Lowe, R. J. (2022). Training language models to follow instructions with human feedback. *Arxiv*. https://doi.org/10.48550/arxiv.2203.02155
14. Reuters. (2025, February 20). OpenAI's weekly active users surpass 400 million. *Reuters*. https://www.reuters.com/technology/artificial-intelligence/openais-weekly-active-users-surpass-400-million-2025-02-20/
15. Sguerra, B., Tran, V.-A., Hennequin, R., & Moussallam, M. (2025). *Uncertainty in Repeated*

Implicit Feedback as a Measure of Reliability. Arxiv. https://arxiv.org/abs/2505.02492

16. Ye, Z., Yoganarasimhan, H., & Zheng, Y. (2024). LOLA: LLM-Assisted Online Learning Algorithm for Content Experiments. *Arxiv*. https://doi.org/10.48550/arxiv.2406.02611