Personalized Autocompletion of Interactions with LLM-based Chatbots

Shani Goren*[1,2], Oren Kalinsky[1], Tomer Stav[1], Nachshon Cohen[1], Yuri Rapoport[1], Yaron Fairstein[1], Ram Yazdi[1], Alexander Libov[1], and Guy Kushilevitz**[3]

[1] Amazon Research
{stavt, nachshon, rtu, ramyazdi, alibov}@amazon.com
[2] Technion - Israel Institute of Technology
{shani.goren, orenkalinsky, yyfairstein}@gmail.com
[3] Aigency.ai
guy@aigency.ai

Abstract. Composing messages in chatbot interactions is often time-consuming, making autocompletion an appealing way to reduce user effort. Different users have different preferences and therefore different expectations from autocompletion solutions. We study how personalization can improve the autocompletion process, evaluating four schemes defined along two axes: generation vs. ranking, and prior messages vs. external features. Experiments on the WildChat and PRISM datasets with the Mistral-7B and Phi-3.5-mini models show consistent gains. Our results highlight personalization as a key factor in building effective chatbot autocomplete systems, and assist researchers and practitioners in deciding where and how to invest in improving these solutions.

Keywords: LLM · Autocomplete · Personalization

1 Introduction

Large Language Models (LLMs) have transformed natural language interfaces, turning chatbots from narrow, domain-specific tools into flexible assistants capable of addressing a wide variety of tasks [2–4]. Users therefore spend considerable effort formulating their messages in these interactions. Autocompletion provides a natural solution to reduce users' effort by suggesting completions as users type. Autocomplete methods have been shown to be effective in saving users' time and reducing cognitive load in many different use-cases [12]. While autocompletion has been explored in various domains, it has only recently been studied in the context of chatbot interactions [9]. Meanwhile, personalization has emerged as a key theme in natural language processing, aiming to adapt outputs to individual users' preferences, style or context. In this work, we bring this perspective to

* This project was done during an internship at Amazon.
** This project was done while working at Amazon.

chatbot autocompletion (task exhibited in Figure 1 (left)), demonstrating that personalization can make suggestions more relevant and further reduce typing effort. We investigate personalization as a way to tailor autocompletions to individual needs. Specifically, we evaluate four personalization strategies defined along two axes: (1) whether personalization is applied during completion generation or in post-hoc ranking of the completions, and (2) whether it relies on prior user messages or external user features, across multiple datasets and models. We show that personalization is an effective way to produce better autocomplete suggestions and improve performance, as shown in Figure 1 (right). To facilitate reproducibility and future research, we make our code and experimental framework publicly available.[4] Finally, we provide practical recommendations for improving this process. For example, personalizing completion generation is more effective than completion ranking, and leveraging user history is more effective than relying on external features.

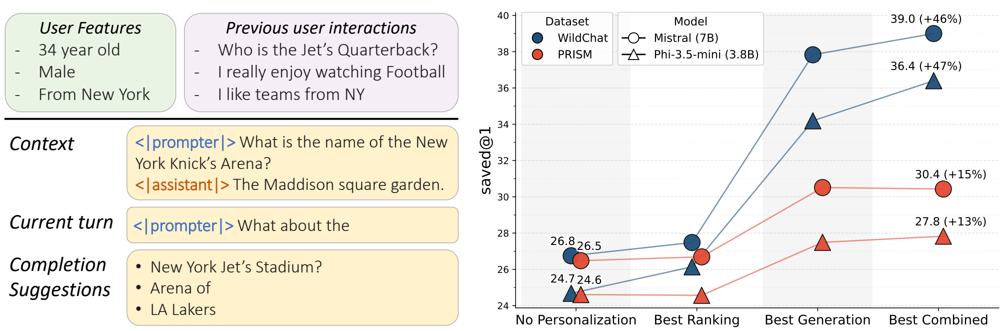

Fig. 1: **Left:** Task Example. The autocomplete solution receives the context and the current turn and returns completion suggestions. In the personalized version, it is also given the previous user interactions or user features.
Right: Best saved@1 (section 3) grouped by scheme per dataset and model.

2 Related Work

Research at the intersection of personalization, autocompletion, and LLM-based interactions remains unexplored. Prior work has typically addressed only two of these dimensions at a time (marked in *italic*). Several studies adapt LLM outputs to individual users (thus **Personalizing LLMs**), either by retrieving relevant history [18, 19], constructing user embeddings from prior activity or style [6, 16, 17], or summarizing past interactions for prompt augmentation [18]. Personalization has also been studied in long-term dialogue [13] and recommender systems

[4] https://github.com/amazon-science/Personalized-chat-interaction-autocomplete

[7, 8, 14, 15]. In *Autocompletion, personalization* has been studied in structured domains, such as search queries, personalization is typically used to improve the ranking of predicted completions [5, 20, 21]. Such methods are not designed for the free-form, multi-turn nature of chatbot conversations. The ChaI-TeA benchmark [9] introduced the task of *Chatbot Interaction Autocomplete*, defining a framework also used in this paper, and providing baseline results. However, it treats all users uniformly, deferring the question of how personalization can enhance LLM-based autocompletion to future work.

3 Problem Definition and Setup

Modeling autocompletion of user-chatbot interactions. We use the sequential modeling proposed in [9]: at each step, an autocomplete solution (AC) receives a context consisting of all previous conversation turns, along with the prefix of the current turn. AC produces a list of candidate completions. These candidates are ranked to determine the top-k suggestions shown to the user, who may accept a suggestion or continue typing. If a completion is accepted, it's appended to the prefix. The process repeats until the end of the user turn.
Personalization Schemes. Personalization can be applied at two stages of the autocompletion solution: *generation*, where the context is augmented so that the candidates themselves reflect user-specific information, and *ranking*, where generated candidates are reordered using additional user signals. We consider two sources of personalization signals: prior interactions, capturing a user's style and topical continuity, and external features, such as demographics or stated preferences. We evaluate each scheme individually to assess its contribution and examine their combinations to explore potential compounding effects.

Datasets. We evaluate on two datasets containing conversations between users and a chatbot assistant, selected as they maintain user identification and capture multiple interactions per user. *WildChat* [22] is a comprehensive multi-turn dataset consisting of timestamped conversations collected via a chatbot service powered by GPT API. *PRISM* [11] gathers conversations conducted by diverse participants from different countries, as well as sociodemographic features and other stated preferences. Table 1 summarizes the stats of the curated datasets. Following [9] we first filter out non-English conversations. Then, for each user-turn we extract all possible prefixes and pair each with the entire conversation history up to that point as its context. The suffix of the prompt is the ground-truth completion. History Length is average characters per user.

Table 1: Dataset statistics.

	WildChat	**PRISM**
Users	871	246
Conversations	9,324	1,483
Messages	18,571	4,996
Prefixes	41,171	2,825
History Length	7,313	1,995

Models. We evaluate our methods on two open-source LLMs: Mistral-7B [10] and Phi-3.5-mini (3.8B) [1]. These models were selected as compact yet competitive representatives, balancing efficiency with performance. Candidate comple-

tions are generated by prompting the model with the context appropriate to the evaluated personalization scheme, followed by the current user prefix.
Metric. We use the saved@k metric proposed by [9] to measure the proportion of characters the autocomplete process saves for the user. Formally, saved@$k = \dfrac{len(\text{accepted\_text}) - \#\text{acceptances}}{len(\text{full\_turn}) - 1}$, where k denotes the number of suggestions shown to the user at each step. We simulate acceptances using exact string match.

4 Results

4.1 Personalized Generation

In this section, we explore the effect of adding a user's personal data to the context when generating completions. Table 2 summarizes the results for the WildChat and PRISM datasets. We experimented with prepending previous user messages to the prefix, and with using personal features like age, location etc.

Table 2: Personalized generation results on WildChat and PRISM. Best and second-best for each configuration (dataset, model, k) are bold and underlined.

Dataset	Personalized Generation Scheme	#prev.	Mistral (7B)			Phi-3.5-mini (3.8B)		
			saved@1	saved@5	saved@100	saved@1	saved@5	saved@100
WildChat	No Personalization	-	26.75	39.85	41.43	24.68	37.31	39.15
	Previous Messages	1	33.93	48.04	49.82	30.73	44.17	46.29
	Previous Messages	3	36.43	51.05	53.11	33.39	47.29	49.52
	Previous Messages	5	37.26	52.19	54.14	<u>33.91</u>	<u>48.10</u>	<u>50.23</u>
	Previous Messages	10	**37.83**	**52.75**	**54.77**	**34.19**	**48.21**	**50.44**
	Previous Messages	20	<u>37.61</u>	<u>52.45</u>	<u>54.49</u>	32.42	45.99	47.96
	Previous Messages	All	37.01	51.59	53.56	29.81	41.96	43.92
PRISM	No Personalization	-	26.48	38.66	39.55	24.61	35.54	37.07
	Previous Messages	1	26.76	39.54	40.46	25.15	35.96	37.49
	Previous Messages	3	28.25	41.78	43.33	26.25	37.66	39.37
	Previous Messages	5	29.78	41.83	43.04	25.48	37.36	38.60
	Previous Messages	10	29.95	43.92	45.50	26.80	38.69	40.08
	Previous Messages	20	29.04	43.17	45.61	27.49	38.74	40.73
	Previous Messages	All	**30.51**	44.10	45.44	27.05	38.06	39.49
	Personal Features	-	29.06	41.40	42.46	25.95	38.21	39.79
	Combined	1	28.52	42.30	44.07	25.51	38.26	39.73
	Combined	3	29.26	42.52	44.12	25.63	37.01	38.65
	Combined	5	28.68	41.93	43.70	27.20	38.40	40.21
	Combined	10	29.52	43.81	45.18	27.46	39.23	40.93
	Combined	20	29.77	43.92	<u>45.81</u>	<u>27.58</u>	<u>39.51</u>	<u>40.97</u>
	Combined	All	30.43	**45.22**	**46.73**	**27.83**	**39.56**	**41.42**

Previous Messages. We use a varying number of the most recent messages. For the WildChat dataset, we observe substantial gains using this scheme. Most gains came from the first few messages: for example, saved@1 increased from 26.75% to 33.93% for Mistral, and from 24.68% to 30.73% for Phi when adding just a single prior message. Performance peaked at 10 messages before declining,

suggesting that overly long contexts may dilute the signal.
In contrast, the PRISM dataset exhibited a more gradual trend. Gains from recent messages were modest, with minimal improvement from the first message alone (e.g.: saved@1 for Mistral improved from 26.48% to 26.76%.) However, performance continued to improve as more context was added, peaking when using 20 messages for Phi and all available messages for saved@1 and saved@5 for Mistral, suggesting that the larger Mistral is more capable of handling longer contexts. Differences between the datasets can be attributed to the longer messages in WildChat compared to PRISM (Table 1), which likely provide a richer signal for personalization. Moreover, in the curation process of the PRISM dataset users were encouraged to ask LLMs about different topics, limiting the relevance of previous messages. In summary, incorporating prior messages improves personalization, with the extent of the benefit varying across datasets and models.

Personal Features. Using PRISM's survey data, we utilized the available user features by constructing a *personal intro* for each user - a short natural language summary of the user's features, such as: age, country, education, etc. The personal features are inserted into a fixed template. Following is a shortened illustration of such a summary (actual intros contained more features):

> I am a 45-54 year old male from the United States of America.
> I have a graduate degree. My marital status is married. I have 2 kids.
> I value learning and self-improvement.

Intros were prepended to the input prefix. Adding a personal intro alone improved performance, indicating that explicit user features provide a meaningful personalization signal. For Mistral, the improvement was comparable to including 2-3 recent messages, while for Phi it matched the effect of 5-10 messages.
Combined. We also *combine* the personal intro with previous messages (by placing the intro before the messages) and observe a compounding effect: performance improves further, suggesting that personal intros and message history provide complementary forms of personalization. While intros summarize static user attributes, they lack the stylistic and contextual cues that can be found in prior messages.

4.2 Personalized Ranking

In this section, we explore improvement in the ranking of completions offered to the user. This seems beneficial as we observe a substantial performance gap between smaller realistic ks such as 1 or 3 and larger values such as 5 or more.

Previous Messages. We experimented with re-ranking the top 5 completions (as determined by log-likelihood) using various similarity-based methods: cosine similarity between embeddings, edit distance, and Jaccard distance. For each method, we tested two aggregation strategies: maximum similarity (i.e., similarity to the most similar prior message) and mean similarity (mean across all previous messages). To compute these scores, we concatenated the input prefix with its candidate completion and compared the resulting message to the

Table 3: Personal ranking comparison (`saved@1`) on WildChat and PRISM: Log-likelihood (LogL) vs. Jaccard-based solution (Jaccard + LogL) similarity.

Dataset	Personalized Generation Scheme	#prev.	Mistral (7B)			Phi-3.5-mini (3.8B)		
			LogL	Jaccard + LogL	Δ	LogL	Jaccard + LogL	Δ
WildChat	No Personalization	-	26.75	**27.48**	+0.73	24.68	**26.14**	+1.46
	Previous Messages	All	37.01	**39.00**	+1.99	29.81	**32.81**	+3.00
PRISM	No Personalization	-	26.48	**26.69**	+0.21	**24.61**	24.57	-0.04
	Previous Messages	All	30.51	**30.70**	+0.18	**27.05**	26.93	-0.12

user's previous messages. We also explored combining similarity scores via a weighted sum. Out of all of these variations, using a combination of Jaccard and log-likelihood aggregated using maximum similarity yielded the best results, and is reported in Table 3. This is likely due to the fact that Jaccard similarity captures repeated words from previous turns, providing a strong signal for ranking.

For the WildChat dataset, re-ranking with Jaccard similarity improved results by 0.73% for Mistral and 1.46% for Phi. Combining this solution with personalized generation increased gains, reaching up to 3%. The largest improvements were observed when using all previous messages, suggesting that this approach may help mitigate the adverse effects of excessive context, by steering the ranking back toward the most relevant parts of the user's history. Results on the PRISM dataset were mixed, with gains being modest. For Mistral, personalizing using the Jaccard-based solution modestly improved results (+0.18%). Combining with personalized generation slightly reduced the gap (+0.11%), while Phi actually showed minimal drops in performance. This echoes the trend we observed: WildChat's longer topic-focused turns provide strong personalization signals, while PRISM's inherent topic shifts limit their usefulness.

Personal Features. To complete the four schemes, we also re-ranked completions using external user features from PRISM. We started by using a strong model for this task to gauge the ability of the most sophisticated models available today to utilize these features - we chose Claude 3.7 Sonnet to re-rank the top five log-likelihood completions according to user features. We also experimented with the DSPy[5] framework to optimize prompts for the task to improve even further. Despite using such a strong solution, which is impractical for online use-cases, this approach underperformed compared to log-likelihood ranking. On Mistral with PRISM, saved@1 dropped from 26.96% with log-likelihood to 24.43% with feature-based ranking, while saved@2 decreased from 32.61% to 31.02% and saved@3 from 35.64% to 34.12%. This is an important negative result of our work - reranking with external user features does not seem to bring value. We hypothesize that this is due to the fact that user style, which is important for reranking, is not well represented in these signals.

[5] We used the MIPROv2 algorithm: https://github.com/stanfordnlp/dspy

5 Conclusions

We extensively studied personalization for chatbot autocompletion, evaluating four schemes across generation and ranking using prior messages and external features. Our experiments show that personalization consistently improves performance, in some cases reaching improvements of over 45%. This highlights personalization as an important component in building effective autocomplete systems that produce effective completions. We find that personalizing the generation step of the autocompletion solution is more important than doing so for the ranking step. Also, that leveraging conversational history brings the lion's share of the gains, while external features have weaker impact, but offer complementary signals to the history approach. These findings serve as a spotlight for researchers and practitioners trying to personalize autocomplete solutions.

References

1. Abdin, M., Aneja, J., Awadalla, H., Awadallah, A., Awan, A.A., Bach, N., Bahree, A., Bakhtiari, A., Bao, J., Behl, H., Benhaim, A., Bilenko, M., Bjorck, J., Bubeck, S., Cai, M., Cai, Q., Chaudhary, V., Chen, D., Chen, D., Chen, W., Chen, Y.C., Chen, Y.L., Cheng, H., Chopra, P., Dai, X., Dixon, M., Eldan, R., Fragoso, V., Gao, J., Gao, M., Gao, M., Garg, A., Giorno, A.D., Goswami, A., Gunasekar, S., Haider, E., Hao, J., Hewett, R.J., Hu, W., Huynh, J., Iter, D., Jacobs, S.A., Javaheripi, M., Jin, X., Karampatziakis, N., Kauffmann, P., Khademi, M., Kim, D., Kim, Y.J., Kurilenko, L., Lee, J.R., Lee, Y.T., Li, Y., Li, Y., Liang, C., Liden, L., Lin, X., Lin, Z., Liu, C., Liu, L., Liu, M., Liu, W., Liu, X., Luo, C., Madan, P., Mahmoudzadeh, A., Majercak, D., Mazzola, M., Mendes, C.C.T., Mitra, A., Modi, H., Nguyen, A., Norick, B., Patra, B., Perez-Becker, D., Portet, T., Pryzant, R., Qin, H., Radmilac, M., Ren, L., de Rosa, G., Rosset, C., Roy, S., Ruwase, O., Saarikivi, O., Saied, A., Salim, A., Santacroce, M., Shah, S., Shang, N., Sharma, H., Shen, Y., Shukla, S., Song, X., Tanaka, M., Tupini, A., Vaddamanu, P., Wang, C., Wang, G., Wang, L., Wang, S., Wang, X., Wang, Y., Ward, R., Wen, W., Witte, P., Wu, H., Wu, X., Wyatt, M., Xiao, B., Xu, C., Xu, J., Xu, W., Xue, J., Yadav, S., Yang, F., Yang, J., Yang, Y., Yang, Z., Yu, D., Yuan, L., Zhang, C., Zhang, C., Zhang, J., Zhang, L.L., Zhang, Y., Zhang, Y., Zhang, Y., Zhou, X.: Phi-3 technical report: A highly capable language model locally on your phone (2024), https://arxiv.org/abs/2404.14219
2. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al.: Gpt-4 technical report. arXiv preprint arXiv:2303.08774 (2023)
3. Anil, R., Borgeaud, S., Wu, Y., Alayrac, J., Yu, J., Soricut, R., Schalkwyk, J., Dai, A.M., Hauth, A., Millican, K., Silver, D., Petrov, S., Johnson, M., Antonoglou, I., Schrittwieser, J., Glaese, A., Chen, J., Pitler, E., Lillicrap, T.P., Lazaridou, A., Firat, O., Molloy, J., Isard, M., Barham, P.R., Hennigan, T., Lee, B., Viola, F., Reynolds, M., Xu, Y., Doherty, R., Collins, E., Meyer, C., Rutherford, E., Moreira, E., Ayoub, K., Goel, M., Tucker, G., Piqueras, E., Krikun, M., Barr, I., Savinov, N., Danihelka, I., Roelofs, B., White, A., Andreassen, A., von Glehn, T., Yagati, L., Kazemi, M., Gonzalez, L., Khalman, M., Sygnowski, J., et al.: Gemini: A family of highly capable multimodal models. CoRR **abs/2312.11805** (2023). https://doi.org/10.48550/ARXIV.2312.11805, https://doi.org/10.48550/arXiv.2312.11805
4. Anthropic: The claude 3 model family: Opus, sonnet, haiku (2023), https://assets.anthropic.com/m/61e7d27f8c8f5919/original/Claude-3-Model-Card.pdf, accessed: 2025-05-13
5. Cai, F., Liang, S., de Rijke, M.: Time-sensitive personalized query auto-completion. In: Li, J., Wang, X.S., Garofalakis, M.N., Soboroff, I., Suel, T., Wang, M. (eds.) Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM 2014, Shanghai, China, November 3-7, 2014. pp. 1599–1608. ACM (2014). https://doi.org/10.1145/2661829.2661921, https://doi.org/10.1145/2661829.2661921
6. Doddapaneni, S., Sayana, K., Jash, A., Sodhi, S.S., Kuzmin, D.: User embedding model for personalized language prompting. CoRR **abs/2401.04858** (2024). https://doi.org/10.48550/ARXIV.2401.04858, https://doi.org/10.48550/arXiv.2401.04858

7. Gao, Y., Sheng, T., Xiang, Y., Xiong, Y., Wang, H., Zhang, J.: Chat-rec: Towards interactive and explainable llms-augmented recommender system. CoRR **abs/2303.14524** (2023). https://doi.org/10.48550/ARXIV.2303.14524, https://doi.org/10.48550/arXiv.2303.14524
8. Geng, S., Liu, S., Fu, Z., Ge, Y., Zhang, Y.: Recommendation as language processing (RLP): A unified pretrain, personalized prompt & predict paradigm (P5). In: Golbeck, J., Harper, F.M., Murdock, V., Ekstrand, M.D., Shapira, B., Basilico, J., Lundgaard, K.T., Oldridge, E. (eds.) RecSys '22: Sixteenth ACM Conference on Recommender Systems, Seattle, WA, USA, September 18 - 23, 2022. pp. 299–315. ACM (2022). https://doi.org/10.1145/3523227.3546767, https://doi.org/10.1145/3523227.3546767
9. Goren, S., Kalinsky, O., Stav, T., Rapoport, Y., Fairstein, Y., Yazdi, R., Cohen, N., Libov, A., Kushilevitz, G.: Chai-tea: A benchmark for evaluating autocompletion of interactions with llm-based chatbots. CoRR **abs/2412.18377** (2024). https://doi.org/10.48550/ARXIV.2412.18377, https://doi.org/10.48550/arXiv.2412.18377
10. Jiang, A.Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D.S., de las Casas, D., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L.R., Lachaux, M.A., Stock, P., Scao, T.L., Lavril, T., Wang, T., Lacroix, T., Sayed, W.E.: Mistral 7b (2023), https://arxiv.org/abs/2310.06825
11. Kirk, H.R., Whitefield, A., Röttger, P., Bean, A.M., Margatina, K., Gómez, R.M., Ciro, J., Bartolo, M., Williams, A., He, H., Vidgen, B., Hale, S.: The PRISM alignment dataset: What participatory, representative and individualised human feedback reveals about the subjective and multicultural alignment of large language models. In: Globersons, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J.M., Zhang, C. (eds.) Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024 (2024)
12. Lehmann, F., Buschek, D.: Examining autocompletion as a basic concept for interaction with generative ai. i-com **19**(3), 251–264 (2021)
13. Li, H., Yang, C., Zhang, A., Deng, Y., Wang, X., Chua, T.: Hello again! llm-powered personalized agent for long-term dialogue. CoRR **abs/2406.05925** (2024). https://doi.org/10.48550/ARXIV.2406.05925, https://doi.org/10.48550/arXiv.2406.05925
14. Liu, J., Liu, C., Lv, R., Zhou, K., Zhang, Y.: Is chat-gpt a good recommender? A preliminary study. CoRR **abs/2304.10149** (2023). https://doi.org/10.48550/ARXIV.2304.10149, https://doi.org/10.48550/arXiv.2304.10149
15. Lyu, H., Jiang, S., Zeng, H., Xia, Y., Wang, Q., Zhang, S., Chen, R., Leung, C., Tang, J., Luo, J.: Llm-rec: Personalized recommendation via prompting large language models. In: Duh, K., Gómez-Adorno, H., Bethard, S. (eds.) Findings of the Association for Computational Linguistics: NAACL 2024, Mexico City, Mexico, June 16-21, 2024. pp. 583–612. Association for Computational Linguistics (2024). https://doi.org/10.18653/V1/2024.FINDINGS-NAACL.39, https://doi.org/10.18653/v1/2024.findings-naacl.39
16. Neelakanteswara, A., Chaudhari, S., Zamani, H.: RAGs to style: Personalizing LLMs with style embeddings. In: Deshpande, A., Hwang, E., Murahari, V., Park, J.S., Yang, D., Sabharwal, A., Narasimhan, K., Kalyan, A. (eds.) Proceedings of the 1st Workshop on Personalization of Generative AI Systems (PERSONALIZE

2024). pp. 119–123. Association for Computational Linguistics, St. Julians, Malta (Mar 2024), https://aclanthology.org/2024.personalize-1.11/
17. Ning, L., Liu, L., Wu, J., Wu, N., Berlowitz, D., Prakash, S., Green, B., O'Banion, S., Xie, J.: User-llm: Efficient LLM contextualization with user embeddings. CoRR **abs/2402.13598** (2024). https://doi.org/10.48550/ARXIV.2402.13598, https://doi.org/10.48550/arXiv.2402.13598
18. Richardson, C., Zhang, Y., Gillespie, K., Kar, S., Singh, A., Raeesy, Z., Khan, O.Z., Sethy, A.: Integrating summarization and retrieval for enhanced personalization via large language models. CoRR **abs/2310.20081** (2023). https://doi.org/10.48550/ARXIV.2310.20081, https://doi.org/10.48550/arXiv.2310.20081
19. Salemi, A., Mysore, S., Bendersky, M., Zamani, H.: Lamp: When large language models meet personalization. In: Ku, L., Martins, A., Srikumar, V. (eds.) Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024. pp. 7370–7392. Association for Computational Linguistics (2024). https://doi.org/10.18653/V1/2024.ACL-LONG.399, https://doi.org/10.18653/v1/2024.acl-long.399
20. Shokouhi, M.: Learning to personalize query auto-completion. In: Jones, G.J.F., Sheridan, P., Kelly, D., de Rijke, M., Sakai, T. (eds.) The 36th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR '13, Dublin, Ireland - July 28 - August 01, 2013. pp. 103–112. ACM (2013). https://doi.org/10.1145/2484028.2484076, https://doi.org/10.1145/2484028.2484076
21. Yin, D., Tan, J., Zhang, Z., Deng, H., Huang, S., Chen, J.: Learning to generate personalized query auto-completions via a multi-view multi-task attentive approach. In: Gupta, R., Liu, Y., Tang, J., Prakash, B.A. (eds.) KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020. pp. 2998–3007. ACM (2020). https://doi.org/10.1145/3394486.3403350, https://doi.org/10.1145/3394486.3403350
22. Zhao, W., Ren, X., Hessel, J., Cardie, C., Choi, Y., Deng, Y.: Wildchat: 1m chatgpt interaction logs in the wild. In: The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024. OpenReview.net (2024), https://openreview.net/forum?id=Bl8u7ZRlbM