MAAQR: An LLM-based Multi-Agent Framework for Adaptive Query Rewriting in Alipay Search

Qi Zheng*
qizai.zq@antgroup.com
Ant Group
Hangzhou, China

Mingjie Zhong*†
mingjie.zmj@antgroup.com
Ant Group
Hangzhou, China

Saisai Gong
gongsaisai.gss@antgroup.com
Ant Group
Hangzhou, China

Huimin Jiang
cuixing.jhm@antgroup.com
Ant Group
Hangzhou, China

Kaixin Wu
daniel.wkx@antgroup.com
Ant Group
Hangzhou, China

Hong Liu
yizhou.lh@antgroup.com
Ant Group
Hangzhou, China

Jia Xu
steve.xuj@antgroup.com
Ant Group
Hangzhou, China

Linjian Mo
linyi01@antgroup.com
Ant Group
Hangzhou, China

Abstract

Query rewriting is essential in e-commerce search, as it bridges the lexical gap between user queries and item descriptions, thereby enhancing search performance. Despite recent advancements, current rewriting approaches are still limited by an inadequate comprehension of domain-specific knowledge and a lack of mechanisms for adaptive refinement in response to new or changing query-item relationships. To overcome these limitations, we propose a large language model (LLM) based Multi-Agent Framework for Adaptive Query Rewriting (MAAQR) in Alipay Search. Initially, we perform knowledge-enhanced fine-tuning to improve the LLM's understanding of query and item semantics. Subsequently, a multi-agent collaborative rewriting architecture is employed to enhance rewrite quality and adaptability. MAAQR has been successfully deployed to serve Alipay's mini-app search since December 2024. Through offline experiments and online A/B testing, MAAQR significantly improves click-through rates (CTR) and the number of transactions for target queries, while substantially reducing the zero-results rate (ZRR).

CCS Concepts

• **Information systems → Query reformulation**; • **Computing methodologies → Multi-agent systems**.

Keywords

Query Rewriting, Large Language Models, Multi-Agent, Information Retrieval

*Both authors contributed equally to this research.
†Corresponding author.

SIGIR '25, Padua, Italy

ACM ISBN 979-8-4007-1592-1/2025/07
https://doi.org/10.1145/3726302.3731950

ACM Reference Format:
Qi Zheng, Mingjie Zhong, Saisai Gong, Huimin Jiang, Kaixin Wu, Hong Liu, Jia Xu, and Linjian Mo. 2025. MAAQR: An LLM-based Multi-Agent Framework for Adaptive Query Rewriting in Alipay Search. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25), July 13–18, 2025, Padua, Italy.* ACM, Padua, Italy, 5 pages. https://doi.org/10.1145/3726302.3731950

1 Introduction

E-commerce search engines like Alipay Search[1] help users quickly find desired products, services, or other relevant information on e-commerce platforms. An important challenge in e-commerce search is the vocabulary mismatch between user queries and item descriptions [1, 2]. To address this challenge, query rewriting or expansion is widely used to reformulate user queries, allowing them to align more closely with item descriptions [2–4].

Traditional query rewriting approaches focus on enhancing the original query by adding semantically related terms or phrases. They typically leverage pseudo-relevance feedback (PRF) [5, 6], search logs [7, 8], or ontologies/thesauri [9, 10] to extract new query terms. An alternative line of research explores the use of generative models for query reformulation. In this context, seq2seq models, often with a limited number of parameters, are trained to generate rewritten queries [11–13]. However, the limited ability of traditional methods to understand the semantics of queries and items hinders search performance improvement [14].

Recent advancements in LLMs have inspired efforts to harness their exceptional capabilities in language comprehension and generation for query rewriting, effectively bridging the vocabulary gap [14]. Existing LLM-based rewriting falls into two categories: prompting-based [1, 4, 15–20] and training-based methods [2, 3]. Prompting-based methods leverage the internal knowledge of LLMs to generate rewritten queries or other supplementary information

[1]https://www.alipay.com/

without additional training [1]. However, their efficacy may significantly decline when LLMs fail to capture the domain-specific knowledge required by the downstream task [3, 21]. Training-based methods tailor pre-trained LLMs to specific domains or tasks through supervised fine-tuning (SFT) or reinforcement learning (RL), enhancing their domain-specific performance [2, 3]. However, they often struggle to adapt to evolving relationships between queries and items, as they are not continuously updated to reflect new items, emerging patterns, and changing user needs.

To address the aforementioned issues and improve overall search performance, we propose the MAAQR framework, which leverages multiple LLM agents to collaboratively rewrite and adapt user queries based on relevance feedback and insights from user behavior. To equip LLM agents with deeper domain-specific knowledge, we utilize keywords associated with items as a form of domain-specific knowledge, capturing the functionalities [22] or attribute values [23] of items available on e-commerce platforms. We conduct knowledge-enhanced fine-tuning of LLMs, incorporating keyword generation subtasks alongside knowledge-driven data synthesis.

After fine-tuning, MAAQR leverages the collaborative power of four LLM agents to achieve a comprehensive understanding of search intent and item content. In MAAQR, two rewrite agents with distinct roles generate diverse initial query reformulations, capturing multiple perspectives on the user's input. An adaptation agent continues to refine the query rewriting process by incorporating keywords from retrieved items w.r.t. initial rewrites. In this way, this agent can enhance LLMs with up-to-date information and adapt rewriting results to new or changing items [21]. To enhance the overall quality of rewriting, a decision agent utilizes insights on rewrite quality derived from user behavior to synthesize diverse rewritten information across agents and formulate the final results. We applied MAAQR to Alipay mini-app search.

Our contributions are as follows:(1) We propose an LLM-based multi-agent framework named MAAQR for adaptive query rewriting;(2) We incorporate knowledge-enhanced fine-tuning, relevance feedback, and user behavior insights to adaptively refine rewriting results;(3) We demonstrate the effectiveness of MAAQR through experiments on real-world industrial datasets.

2 Methodology

The overall systematic framework of MAAQR is illustrated in Fig. 1, which comprises two core components: knowledge-enhanced fine-tuning and multi-agent collaborative rewriting.

2.1 Knowledge-enhanced Fine-tuning

SFT techniques can adapt pre-trained LLMs for the specific task of query rewriting, thereby enhancing overall search performance [2, 3]. However, the training data used for SFT is usually scarce and expensive for collection [21]. Synthetic training data generation using LLMs has emerged as a promising approach to mitigate data scarcity [24]. In addition, mini-app keywords offer valuable insights into functionality-related search intent. This knowledge derived from keywords can be integrated into SFT to enhance rewriting performance. In light of this, we utilize knowledge-driven data synthesis and knowledge-enhanced fine-tuning based on keywords.

2.1.1 Knowledge-driven Data Synthesis. To enhance LLMs with specialized knowledge represented by the mini-apps, we leverage their keywords to guide the training data synthesis process. We use GPT-4 to generate pseudo-queries for each keyword of the mini-app, utilizing the prompt shown in Tab. 1. In this way, we collect a synthetic dataset consisting of <pseudo query, keyword> pairs.

2.1.2 Multi-instruction Supervised Fine-tuning (MISFT). We develop the following two SFT tasks using the instructions shown in Tab. 1.
Keyword-prediction Task: The input of this task is the <pseudo query, keyword> pairs from the synthetic dataset. Given a prompt x and its gold standard tokens $y = y_1, ..., y_m$, the training objective of this task is to minimize the following negative log likelihood:

$$\mathcal{L}_{\text{SFT,kp}}(\theta) = -\mathbb{E}_{(x,y)\sim\mathcal{D}_{\text{kp}}} \log \pi(y \mid x; \theta) \tag{1}$$

where $\pi(y \mid x; \boldsymbol{\theta}) = \prod_{j=1}^{m} p\left(y_j \mid x, y_{<j}; \boldsymbol{\theta}\right)$ denotes the output probability, D_{kp} denotes the keyword-prediction SFT data. $\pi(.)$ and θ denote our query rewriting model and its parameters.
Query-rewriting Task: We construct the training data using validated query reformulation patterns from online search logs. We achieve this by selecting high-quality query-rewrite pairs based on the condition that the CTR of items recalled from the rewritten query exceeds a predefined threshold. This forms a dataset of <query, rewrite query> sample pairs. The loss function of this task is similar to the keyword-prediction task as follows:

$$\mathcal{L}_{\text{SFT,qr}}(\theta) = -\mathbb{E}_{(x,y)\sim\mathcal{D}_{\text{qr}}} \log \pi(y \mid x; \theta) \tag{2}$$

where D_{qr} denotes the query-rewriting SFT data. The overall loss is composed of the two aforementioned loss functions as follows:

$$\mathcal{L}_{\text{SFT}}(\theta) = \mathcal{L}_{\text{SFT,kp}}(\theta) + \mathcal{L}_{\text{SFT,qr}}(\theta) \tag{3}$$

2.2 Multi-Agent Collaborative Rewriting

Our experiments show that LLMs struggle to generate accurate rewrites when using a single approach, due to their limited domain understanding. Both SFT and RL fail to adapt to query/item distribution shifts. To address this, we propose a Multi-Agent framework (Fig. 1, Step 2). First, we utilize a dual-core LLM to generate diverse and relevant rewritten queries. Second, we incorporate keywords knowledge to generate adaptive rewritten queries. Finally, we select the optimal rewritten query based on user behavior data.

2.2.1 Dual-Core LLM Semantic Generation.

- RSQR (Relevance-Steered Query Rewrite): **Focusing on text relevance**, it rewrites the original complex query into the common expression forms of search queries in Alipay.
- ISQR (Intent-Steered Query Rewrite): **Focusing on intent satisfaction**, it transforms the original complex query into a potential mini-app that fulfills the user's search intent. Unlike RSQR, limited to text-level rewriting, ISQR offers greater diversity while ensuring intent satisfaction.

2.2.2 Keyword-Aware Query Adaptation (KAQA). **Focusing on adaptively rewriting queries to match mini-app keyword expressions.** Keywords represent the core functionalities of mini-apps and are regularly updated by professionals. To improve recall rates, query rewrites must adaptively reflect the expressions of

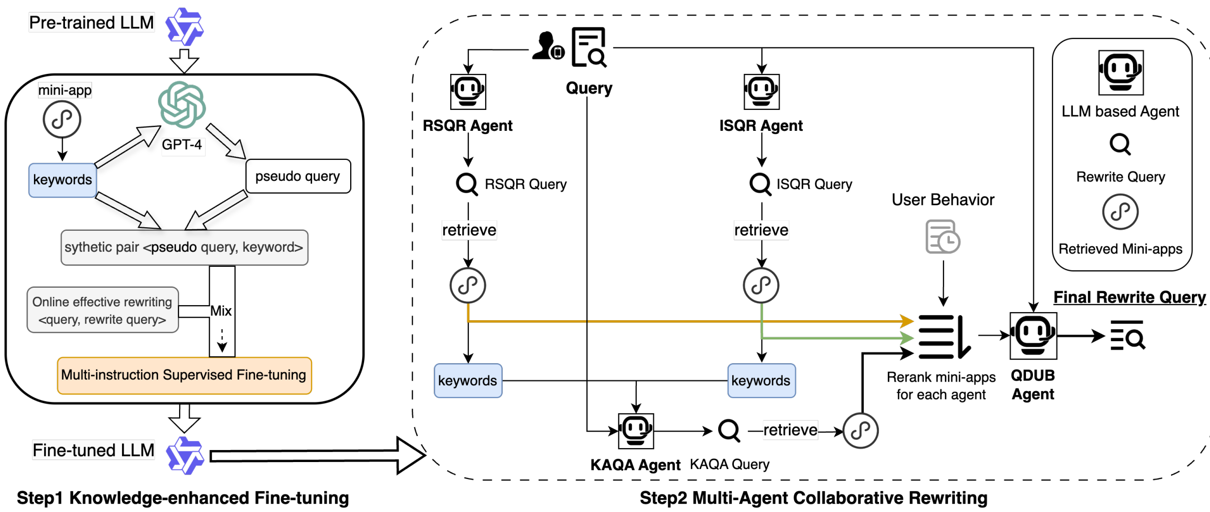

Figure 1: The Framework of MAAQR . Step1 illustrates the offline training pipeline. Step2 demonstrates the inference workflow, where RSQR and ISQR generate rewrites based on textual relevance and intent, respectively. KAQA further refines the query using feedback from RSQR and ISQR retrieval results. Finally, QDUB determines the optimal query incorporating user behavior.

Table 1: Prompt engineering for SFT . This table shows the prompts designed for data synthesis and SFT.

Phase	Task	Prompt
Data Synthesis	Keyword2query	Consider a mini-app called {*mini-app title*}, which has the following keywords that describe its functionalities: {*keywords*}. Your task is to create a reasonable query for the current keyword. keyword: {keyword 1} Query: {query 1}, keyword: {keyword 2} Query: {query 2} keyword: {*keyword*} Query:
SFT	Keyword-prediction	Please provide the exact keyword for the most relevant mini-app related to the given query Query: {*query*} System: {*keyword*}
	Query-rewriting	You are a query rewrite system designed to optimize queries ... Please rewrite the following query Query: {*query*} System: {*rewrite*}

keywords. Inspired by InteR [16], we enhance the adaptability of LLM-generated rewrites through retrieval feedback. Specifically, rewrites from RSQR and ISQR are used to retrieve mini-apps and extract relevant keywords from them. These keywords are then combined with the original query and fed back into the LLM. This process enables the model to generate more adaptive rewrites that better capture keyword-based expressions.

2.2.3 Query decisions driven by user behavior (QDUB). **Focusing on making query rewrite decisions using user behavior to maximize business value**. To enhance the value of rewritten queries in business scenarios (e.g., CTR and GMV), we first leverage user behavior data to rank the recall sets from the three aforementioned agents, filtering out low-CTR items. Then, functioning as a simulated user proxy, the QDUB Agent evaluates these refined results by conducting pairwise comparisons between the original query and each rewritten variant's mini-app ecosystem—analyzing key metrics such as CTR, GMV, and user engagement. This process ultimately selects the most commercially viable rewritten option, ensuring alignment with both user intent and business objectives.

Through these three steps, we have developed an efficient and adaptive query rewriting framework that enhances the diversity and relevance of query rewriting and more effectively adapts to different data distributions and their shifts [25].

3 Experiments

3.1 Dataset and Experiment Setup

3.1.1 Dataset. We conducted experiments using a real search dataset sampled from long-tail, complex, and inefficient online queries. It comprises two distinct sources: 1. Q&A: approximately 25,000 long complex queries framed as questions; 2. Few-recall: approximately 30,000 low-CTR queries that produce few or no search results.

3.1.2 Baseline Models. For offline evaluation, we compare the proposed algorithm with two baseline models:

(1) As our main baseline, we use an industrial discriminative-rewriting model (DRM) that integrates various retrieval sources and embodies years of engineering optimizations.

Table 2: Comparison of different LLM query rewrite methods on offline datasets. The best results are in bold, and the second-best results are underlined.

Method	Few-recall		Q&A	
	Num_{recall}	P_{good}	Num_{recall}	P_{good}
DRM (Baseline)	–	–	–	–
Q2D (GPT-4)	+16.21	0.2703	+8.20	0.5769
Qwen_FewShot	+4.32	0.6665	+0.84	0.6611
Qwen_SFT	+16.67	0.5612	+15.22	0.6994
Qwen_MISFT	<u>+17.08</u>	0.6326	<u>+16.77</u>	<u>0.7369</u>
MAAQR w/o feedback	+13.96	**0.7137**	+16.46	0.7245
MAAQR	**+32.03**	<u>0.6868</u>	**+23.92**	**0.7489**

(2) Query2Doc (Q2D) [4], a prevalent LLM-based query rewriting approach, leverages GPT-4 to generate hypothetical answers for enhanced document matching.

3.1.3 Evaluation Metrics. We use metrics Num_{recall} and P_{good} for offline evaluation [26]. Num_{recall} represents the average recall count achieved by integrating the new rewrite query, measuring the overall improvement in the retrieved documents compared to the original query during the offline simulation. P_{good} evaluates recall expansion quality by measuring the fraction of relevant items (as determined by blind expert labeling) in the additional results retrieved compared to the online DRM output.

3.1.4 Implementation Details. We chose Qwen2.5-14B as our backbone model and employed LoRA technology for fine-tuning, which covered all linear layers. We trained the model for 2 epochs with an initial learning rate of 1.0e-4 and a batch size of 8.

3.2 Offline Evaluation

In this section, we provide a detailed description of the offline evaluation and the results of MAAQR on Q&A and Few-recall datasets randomly collected from the Alipay mini-app search logs. We first generated rewritten queries using MAAQR for each data source, then used offline retrieval in our production system to measure recall quantity (Num_{recall}) and relevance quality (P_{good}) via expert blind labeling. We conducted an ablation study. As shown in Tab. 2, the results of the experiment reveal several key points.

(1) MAAQR performs best among both datasets, showing the greatest improvement compared to the DRM baseline rewrite and Q2D (GPT-4) while maintaining good relevance.
(2) Qwen2.5-14B (MISFT) outperforms Qwen2.5-14B (SFT) across all datasets and metrics, demonstrating the effectiveness of SFT samples in the Keyword2query.
(3) MAAQR outperforms MAAQR w/o feedback, highlighting that utilizing retrieval feedback on mini-app attributes and user behavior not only enhances the semantic relevance of queries but also improves consistency in text retrieval.

3.3 Online A/B Test

We conducted rigorous A/B experiments in Alipay Search. MAAQR generated initial rewrites $Q_{\text{rw\_0}}$ from complex Q&A/few-recall queries, with daily updates where $Q_{\text{rw}} = Q_{\text{rw\_t-1}} \cup Q_{\text{rw\_t}}$ aggregates historical and new rewrites. These rewritten queries are incorporated as an additional retrieval source, merging with the multiple online retrieval pathways, and the combined retrieval set is subsequently processed by downstream relevance and ranking modules.

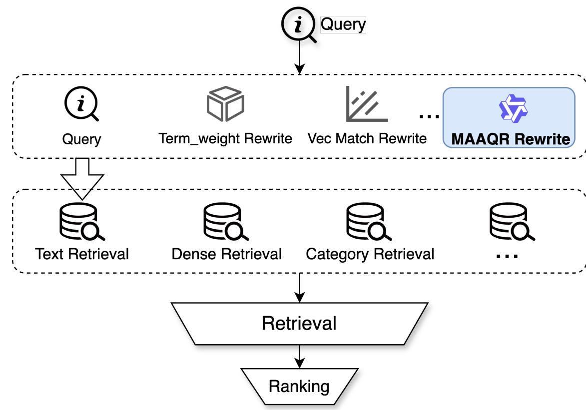

Figure 2: MAAQR online serving. The MAAQR rewriting has been applied to text retrieval.

To evaluate the effectiveness of MAAQR, we conducted a 2-week A/B test through Alipay's experimentation platform, where millions of users were randomly grouped to ensure consistent conditions. The evaluation focused on key search metrics, including the zero-results rate (ZRR), CTR, pay amount (GMV), and transaction volume (#Trans). As shown in Tab. 3, compared to online baseline DRM, MAAQR significantly reduced the ZRR (-14.46%, p<0.01) while increasing CTR (+6.25%, p<0.01) , GMV(+18.02%) and #Trans (+2.79%, p<0.01). The search system's latency only increased by 0.16 ms through Key-Value cache. The results demonstrate that MAAQR effectively improved multiple key metrics, validating its ability to enhance the rewriting of challenging queries.

Table 3: Online A/B test of MAAQR on target queries in Alipay search.. * and ** indicate that the improvement is statistically significant with p < 0.05 and p < 0.01.

Method	CTR	#Trans	GMV	ZRR
DRM (Baseline)	–	–	–	–
Qwen_MISFT	+3.75% **	+0.39%	+3.68%	-7.47%
MAAQR	+6.25% **	+2.79% *	+18.02% **	-14.46% **

4 Conclusion

In this paper, we propose a novel query rewrite system: LLM-based Multi-Agent Adaptive Query Rewriting Framework (MAAQR), which performs knowledge-enhanced fine-tuning to improve the LLM's understanding of query and item semantics and employs a multi-agent collaborative rewriting architecture to enhance rewrite quality and adaptability. Through offline and online A/B testing, we validate the effectiveness of MAAQR in helping users discover previously unrecognized mini-apps, achieving a statistically significant 6.25% increase in CTR. In the future, we plan to distill the rewriting model to enhance our real-time rewriting capabilities for long-tail queries in an online environment.

References

[1] Iain Mackie, Shubham Chatterjee, and Jeffrey Dalton. Generative relevance feedback with large language models. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023*, pages 2026–2031. ACM, 2023.

[2] Aijun Dai, Zhenyu Zhu, Haiqing Hu, Guoyu Tang, Lin Liu, and Sulong Xu. Enhancing e-commerce query rewriting: A large language model approach with domain-specific pre-training and reinforcement learning. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, CIKM 2024, Boise, ID, USA, October 21-25, 2024*, pages 4439–4445. ACM, 2024.

[3] Wenjun Peng, Guiyang Li, Yue Jiang, Zilong Wang, Dan Ou, Xiaoyi Zeng, Derong Xu, Tong Xu, and Enhong Chen. Large language model based long-tail query rewriting in taobao search. In *Companion Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, Singapore, May 13-17, 2024*, pages 20–28. ACM, 2024.

[4] Liang Wang, Nan Yang, and Furu Wei. Query2doc: Query expansion with large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, EMNLP 2023, Singapore, December 6-10, 2023*, pages 9414–9423, 2023.

[5] Tao Tao and ChengXiang Zhai. Regularized estimation of mixture models for robust pseudo-relevance feedback. In *SIGIR 2006: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, Washington, USA, August 6-11, 2006*, pages 162–169, 2006.

[6] Guihong Cao, Jian-Yun Nie, Jianfeng Gao, and Stephen Robertson. Selecting good expansion terms for pseudo-relevance feedback. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2008, Singapore, July 20-24, 2008*, pages 243–250. ACM, 2008.

[7] Hang Cui, Ji-Rong Wen, Jian-Yun Nie, and Wei-Ying Ma. Probabilistic query expansion using query logs. In *Proceedings of the Eleventh International World Wide Web Conference, WWW 2002, May 7-11, 2002, Honolulu, Hawaii, USA*, pages 325–332. ACM, 2002.

[8] Sen Li, Fuyu Lv, Taiwei Jin, Guiyang Li, Yukun Zheng, Tao Zhuang, Qingwen Liu, Xiaoyi Zeng, James T. Kwok, and Qianli Ma. Query rewriting in taobao search. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022*, pages 3262–3271, 2022.

[9] Jagdev Bhogal, Andrew MacFarlane, and Peter W. H. Smith. A review of ontology based query expansion. *Information processing management*, 43(4):866–886, 2007.

[10] Aritra Mandal, Ishita K. Khan, and Prathyusha Senthil Kumar. Query rewriting using automatic synonym extraction for e-commerce search. In *Proceedings of the SIGIR 2019 Workshop on eCommerce, co-located with the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, eCom@SIGIR 2019, Paris, France, July 25, 2019*, 2019.

[11] Yiming Qiu, Kang Zhang, Han Zhang, Songlin Wang, Sulong Xu, Yun Xiao, Bo Long, and Wen-Yun Yang. Query rewriting via cycle-consistent translation for e-commerce search. In *37th IEEE International Conference on Data Engineering, ICDE 2021, Chania, Greece, April 19-22, 2021*, pages 2435–2446, 2021.

[12] Sanjay Agrawal, Srujana Merugu, and Vivek Sembium. Enhancing e-commerce product search through reinforcement learning-powered query reformulation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023*, pages 4488–4494, 2023.

[13] Yifei Yuan, Chen Shi, Runze Wang, Liyi Chen, Renjun Hu, Zengming Zhang, Feijun Jiang, and Wai Lam. CO3: low-resource contrastive co-training for generative conversational query rewrite. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC/COLING 2024, 20-25 May, 2024, Torino, Italy*, pages 3394–3406. ELRA and ICCL, 2024.

[14] ChengXiang Zhai. Large language models and future of information retrieval: Opportunities and challenges. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024*, pages 481–490. ACM, 2024.

[15] Luyu Gao, Xueguang Ma, Jimmy Lin, and Jamie Callan. Precise zero-shot dense retrieval without relevance labels. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 1762–1777, 2023.

[16] Jiazhan Feng, Chongyang Tao, Xiubo Geng, Tao Shen, Can Xu, Guodong Long, Dongyan Zhao, and Daxin Jiang. Synergistic interplay between search and large language models for information retrieval. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*. Association for Computational Linguistics, 2024.

[17] Rolf Jagerman, Honglei Zhuang, Zhen Qin, Xuanhui Wang, and Michael Bendersky. Query expansion by prompting large language models. *CoRR*, abs/2305.03653, 2023.

[18] Marwah Alaofi, Luke Gallagher, Mark Sanderson, Falk Scholer, and Paul Thomas. Can generative llms create query variants for test collections? an exploratory study. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023*, pages 1869–1873. ACM, 2023.

[19] Iain Mackie, Ivan Sekulic, Shubham Chatterjee, Jeffrey Dalton, and Fabio Crestani. GRM: generative relevance modeling using relevance-aware sample estimation for document retrieval. *CoRR*, abs/2306.09938, 2023.

[20] Tao Shen, Guodong Long, Xiubo Geng, Chongyang Tao, Yibin Lei, Tianyi Zhou, Michael Blumenstein, and Daxin Jiang. Retrieval-augmented retrieval: Large language models are strong zero-shot retriever. In *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and virtual meeting, August 11-16, 2024*, pages 15933–15946, 2024.

[21] Yutao Zhu, Huaying Yuan, Shuting Wang, Jiongnan Liu, Wenhan Liu, Chenlong Deng, Zhicheng Dou, and Ji-Rong Wen. Large language models for information retrieval: A survey. *CoRR*, abs/2308.07107, 2023.

[22] Chen Wei, Yixin Ji, Zeyuan Chen, Jia Xu, and Zhongyi Liu. LLMGR: large language model-based generative retrieval in alipay search. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024*, pages 2847–2851. ACM, 2024.

[23] Keiji Shinzato, Naoki Yoshinaga, Yandi Xia, and Wei-Te Chen. Simple and effective knowledge-driven query expansion for qa-based product attribute extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 227–234, 2022.

[24] Zhuoyan Li, Hangxiao Zhu, Zhuoran Lu, and Ming Yin. Synthetic data generation with large language models for text classification: Potential and limitations. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, EMNLP 2023, Singapore, December 6-10, 2023*, pages 10443–10461, 2023.

[25] Orion Weller, Kyle Lo, David Wadden, Dawn J. Lawrie, Benjamin Van Durme, Arman Cohan, and Luca Soldaini. When do generative query and document expansions fail? A comprehensive study across methods, retrievers, and datasets. In *Findings of the Association for Computational Linguistics: EACL 2024, St. Julian's, Malta, March 17-22, 2024*, pages 1987–2003. Association for Computational Linguistics, 2024.

[26] Sen Li, Fuyu Lv, Taiwei Jin, Guiyang Li, Yukun Zheng, Tao Zhuang, Qingwen Liu, Xiaoyi Zeng, James T. Kwok, and Qianli Ma. Query rewriting in taobao search. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022*, pages 3262–3271. ACM, 2022.

Company Portrait

Alipay is China's leading digital interconnectivity and third-party payment open platform, with over 3 million merchants and institutions providing convenient services such as bill payments, transfers, online food ordering, online wealth management, and live shopping to 1 billion consumers.

Presenter Biography

Presenter: **Qi Zheng** is an algorithm engineer at Alipay, focusing on researching and building search systems that help users quickly find the items they want. During his work on Alipay, he applied machine learning to retrieval and ranking in search.