

Report on the 1st Workshop on Conversational Search for Complex Information Needs (CoSCIN'26) at ECIR 2026

Roxana Petcu

University of Amsterdam
The Netherlands
r.m.petcu@uva.nl

Mert Yazan

Amsterdam University of Applied Sciences
The Netherlands
m.yazan@hva.nl

Mohanna Hoveyda

Radboud University
The Netherlands
mohanna.hoveyda@ru.nl

Jirui Qi

University of Groningen
The Netherlands
j.qi@rug.nl

Maarten de Rijke

University of Amsterdam
The Netherlands
m.derijke@uva.nl

Abstract

We present the insights gathered at the 1st Workshop on Conversational Search for Complex Information Needs (CoSCIN'26) held on 2 April 2026, in conjunction with the Forty-eighth European Conference on Information Retrieval (ECIR). The workshop brought together keynote presenters and industry representatives for lively discussions on recent developments in conversational search and its future directions, complemented by Q&A sessions for paper presentations. The rapid developments in conversational search enabled interesting discussions that went beyond generative dialogue and factoid question-answering tasks, towards addressing complex information needs and improving the user experience by providing personalized, adaptive answers to users based on style. Moreover, the work that was presented supported exploratory, multi-step information needs and addressed challenges such as long-form answer generation, personalization, agent orchestration, and societal considerations, which remain largely open areas of research. The participants shared their views from diverse backgrounds in information retrieval, natural language processing, and human-computer interaction, exchanged results and prospective ideas around conversational AI, and discussed its impact on users and the need for adaptivity. Through these discussions, the workshop fostered collaboration and connected researchers specializing in diverse facets of the topic, ultimately advancing the field toward creating more user-centric, intentionally designed, responsible conversational systems.

Date: 2 April 2026.

Website: <https://convsearch-complex-info-needs.github.io/>.

1 Introduction

When we talk about **conversational search**, we refer to a multi-turn dialogue setting in which the user’s information need evolves: At each turn, the system must interpret the underlying intent and generate a response that effectively addresses it. In this setting, **conversational AI** takes a step back and refers to broader methods and pipelines such as Retrieval Augmented Generation (RAG) or agentic orchestration, to address user information needs as precisely as possible. **Information need** refers to the underlying question of the user that might be expressed explicitly or not; **complex information need** refers to a user request that often can be open-ended, and often requires multiple sources of information, reasoning or planning to answer [Kim et al., 2025; Asai et al., 2020; Yao et al., 2023].

Unlike traditional search, where users must write an explicit query that mirrors their information need and navigate ranked lists from which they extract information themselves, conversational systems present all these steps through a natural, adaptive dialogue setting. It is the task of the conversational system to rephrase the user query, extract information from relevant sources, and combine them in a comprehensive answer. As users gain more experience with conversational systems, new expectations appear. Agents are increasingly expected to be experts in specific domains, such as education, healthcare, or finance. They might also be expected to keep track of user-specific information, such as dietary restrictions or movie preferences, so that the user does not have to state them each time when relevant information is needed.

To truly personalize interactions and avoid biases due to personal backgrounds, it is essential to understand how a user expresses their expectations and intentions. At the same time, reliability and robustness become essential for conversational systems, as users increasingly rely on them across a wider range of domains (including sensitive ones) while spending less time verifying the information they receive. This shows the need for conversational systems to be adaptive and address potential problems that might arise, given the increasing expectations.

The European Conference on Information Retrieval (ECIR) has been at the forefront of Information Retrieval open research, supporting advances in theoretical and applied breakthroughs. Meanwhile, recent topics such as conversational AI, retrieval augmented generation (RAG), and evaluation have positioned themselves as a fusion of IR and natural language processing (NLP). Therefore, ECIR provided an ideal platform to explore and debate advances and future strategies for approaching conversational search for complex user scenarios, personalization, and user modeling, promoting responsible and explainable systems, and bringing the community together.

2 Topics of Interest

The workshop provided a collaborative platform for researchers to present the latest advancements in the rapidly evolving field of conversational search. It included original contributions focusing on a diverse range of topics that are relevant in the current space:

1. **Complex information needs:** User queries that go beyond factoid retrieval have been shown to pose significant challenges for retrieval systems [Petcu et al., 2025; van den Elsen et al., 2025]. These queries require reasoning, synthesis of multiple pieces of information [Trivedi et al., 2023], or planning to fill knowledge gaps [Jin et al., 2025]. Such queries

may involve implicit logical operations (e.g., negation, disjunction, conjunction) [Malaviya et al., 2023; Zhang et al., 2025], multi-step dependencies [Petcu et al., 2026; Jeong et al., 2024], topical complexity [Killingback and Zamani, 2025], or require information that lies outside the model’s training data [Mallen et al., 2023; Su et al., 2025]. In this regard, we welcome contributions that propose (i) *evaluation frameworks or datasets* to effectively expose the limitations of current retrieval and RAG systems, including analyses of how learned representations encode logical structures that lead to performance failure, (ii) *models or frameworks* that more effectively handle complex information needs.

2. **Data generation:** Methods for creating new benchmark datasets that capture complex conversational search needs [Askari et al., 2025]. This includes approaches for generating new datasets, augmenting existing datasets, and comparing the new resources with existing benchmarks.
3. **Orchestration:** Modern information systems increasingly rely on coordinating multiple models, tools, and retrieval components [Li et al., 2024]. Orchestration refers to adaptive and agentic approaches that dynamically decide how these resources should be combined to address a user’s request [Hoveyda et al., 2025]. This involves selecting answering strategies that enable adapting system behavior to the complexity of users’ queries. We thereby encourage contributions on (i) *orchestration mechanisms*, (ii) *optimization approaches*, and (iii) *evaluation under dynamic conditions*.
4. **Personalization and adaptivity:** Personalization encompasses many areas, including personalized text generation, response adaptation to diverse user backgrounds, and context tracking. Personalized text generation refers to generating text in an individual’s style and is especially relevant for query rewriting and adapting the system to the user’s style [Yazan et al., 2025]. In this workshop, we approach the personalization problem as an adaptivity problem, and tackle the problem of relating to users from diverse backgrounds, given the inherent biases and stereotypes of LLMs [Plaza-del Arco et al.].
5. **User experience:** Users report higher satisfaction with conversational search compared to traditional search engines, due to exercising less effort [Liang et al., 2025]. This, however, leads to new issues such as overreliance; users blindly trust conversational interfaces without noticing errors [Spatharioti et al., 2025]. Therefore, a new paradigm has emerged: users actively interact with both types of interfaces for balancing weaknesses [Menon and Shilpa, 2023]. There is a lack of studies on how people approach this new paradigm and how their expectations and needs have changed.
6. **Bias, fairness, interpretability:** Responsible systems w.r.t. ethical and societal considerations, bias mitigation, inclusivity, and transparency [Wallat et al., 2025], including works on mitigating misinformation and applications in sensitive domains.
7. **Evaluation:** Evaluation remains a key challenge for conversational search, as current metrics rely on pre-defined gold answers, failing to assess dimensions of the generated answer that go beyond mere overlap with some reference answer. Traditional relevance measures are insufficient for assessing long-form answers, faithfulness of reasoning steps, context tracking, or user-centered aspects such as safety, bias, and personalization [Salemi et al., 2025]. We invite contributions that tackle these gaps with novel metrics and evaluation frameworks.

3 Workshop Report

The 1st Workshop on Conversational Search for Complex Information Needs (CoSCIN'26) included two academic keynote speakers, one industry speaker, oral presentations of submitted papers, and dedicated talks from ECIR-accepted papers on the topic.

3.1 Keynotes

Modeling Complex Conversations

Speaker: Mohammad Aliannejadi

Mohammad Aliannejadi addressed the key challenges of modeling complex conversations in Information Retrieval, focusing on the drop in reliability that Large Language Models (LLMs) experience during multi-turn, underspecified dialogues. He emphasized the challenge a model faces in recovering when taking a wrong turn. To tackle this, the presentation introduced the MQ4CS framework [Abbasiantaeb et al., 2026], developed with Zahra Abbasiantaeb, which breaks down complex, multi-aspect queries into individual sub-queries to retrieve highly relevant, fused rankings instead of relying on standard single-query rewrites. Additionally, the talk presented the DiSCo framework [Lupart et al., 2025], together with Simon Lupart, which bypasses expensive inference-time LLM calls by distilling similarity scores with documents towards a hyperplane, allowing a model with 100 times fewer parameters to run approximately 10 times faster than its LLM teachers while achieving superior effectiveness. Finally, the presentation outlined the upcoming iKAT [Aliannejadi et al., 2024a,b] 2026 track in Melbourne, which will focus on evaluating these conversational systems using live, interactive dialogues against simulated users.

Exploiting Locality for Efficient Conversational Search

Speaker: Cristina-Ioana Muntean

Cristina-Ioana Muntean, Senior Researcher at ISTI-CNR in Pisa, addressed the escalating challenge of efficiency and scalability in conversational search, arguing that computational costs and latency have become the primary bottlenecks as gains in retrieval quality slow down. To tackle this, the talk focused on exploiting topical locality, i.e., the tendency of conversational turns to cluster tightly within the embedding space, allowing systems to search locally rather than globally. Cristina introduced two primary contributions: first, a client-side embedding cache [Frieder et al., 2022] that uses geometric properties to dynamically reuse previously retrieved candidate documents without indexing; and second, the TopLoc framework [Muntean et al., 2025], which speeds up approximate nearest neighbor (ANN) searches. TopLoc improves indexing by leveraging the conversation's first utterance to restrict the search space, either via centroid caching for inverted file indexes or by establishing privileged entry points for hierarchical graphs, yielding speedups of up to 10x with negligible impact on retrieval effectiveness.

3.2 Industry talk

Towards Multi-Agent Evolving AI Scientists for End-to-End Scientific Discovery

Speaker: Yougang Lyu

Yougang Lyu talked about the limitations of existing AI scientist systems and proposed EvoScientist [Lyu et al., 2026]: a self-evolving multi-agent framework designed to continuously improve

both scientific idea generation and experiment execution through cross-task learning. The talk introduced a system decomposed into three specialized agents: a researcher agent (RA) for crafting the idea and research proposal, an engineer agent (EA) for implementing and running experiments, and an evolution manager agent (EMA) as the core driver of self-improvement. The presentation demonstrated the framework’s remarkable real-world potential: under automatic and human expert evaluation, EvoScientist systematically outperformed seven state-of-the-art baselines, substantially increasing code execution success rates, and successfully generating six papers that were all accepted to the AI Scientist Track of ICAIS 2025, with two receiving major paper awards.

3.3 Paper Presentations

Sensitivity-Aware Retrieval-Augmented Intent Clarification [Larooij, 2026]: This work explores the deployment of retrieval-augmented conversational search systems within highly sensitive domains like healthcare, government, and legal contexts. To prevent the leakage of private data when Large Language Models (LLMs) clarify complex user intents, the authors propose a multi-agent framework that acts as a mediator and gatekeeper for sensitive retrieval collections. The paper systematically addresses this security vs. utility trade-off through a three-step methodology: defining a robust threat model, designing sensitivity-aware defense mechanisms at the retrieval level, and establishing empirical evaluation metrics to balance document protection against overall system utility.

Self-Optimizing Multi-Agent Systems for Deep Research [Câmara et al., 2026]: This work addresses the limitations of current deep research systems that rely on static architectures and hand-engineered prompts, which often result in brittle and expensive optimization processes. The authors propose a self-play optimization framework that enables an orchestrator agent and parallel worker agents to explore and dynamically refine prompt combinations. By allowing the multi-agent system to iteratively plan, retrieve, and synthesize evidence across extensive document collections, this approach automates the discovery of high-performing agent behaviors, ultimately matching or outperforming expert-crafted prompt designs.

Let Me Explain — Knowledge-based RAG for Agricultural Recommendation Explanations [Di Scala and de Boer, 2026]: Combining Knowledge Graphs (KGs) and Large Language Models (LLMs), this framework addresses the challenge of providing both factual and fluent recommendations in high-stakes domains like agriculture. The authors introduce KRAGGER, a framework that matches users and items using three distinct graph-based paradigms and generates natural explanations through a hybrid rule- and LLM-based method. Evaluation results and a study with domain experts demonstrate that fine-grained attribute matching paired with RB-LLM rephrasing achieves the highest ratings for explanation simplicity, fluency, coherence, and trustworthiness.

Exploring User Simulators in Conversational Search: A Comparison between LLMs and Humans [Lu and Crestani, 2026]: Analyzing the over-reliance on fine-tuning or heavy in-context learning, this work investigates the zero-shot capabilities of LLM-based user simulators in conversational search. The authors propose a flexible simulator that models diverse user profiles and contexts across both single-response trajectories and interactive, agent-driven (ReAct) environments. Extensive evaluations on lexical diversity, syntactic complexity, and readability reveal

that zero-shot LLM-generated utterances exhibit significantly richer language, greater syntactic complexity, and higher reading difficulty than those produced by humans.

OrLog: Resolving Complex Queries with LLMs and Probabilistic Reasoning [Hoveyda et al., 2026]: Addressing the failure of current neural retrievers to enforce logical operators (conjunction, disjunction, negation) when resolving complex information needs, this paper introduces a novel neuro-symbolic framework. The authors present OrLog, which decouples predicate-level plausibility estimation from logical reasoning by using a single, decoding-free forward pass of a Large Language Model (LLM) to score atomic predicates, followed by a probabilistic reasoning engine that calculates the posterior probability of query satisfaction. Across multiple LLM backbones and logical constraint variations, OrLog significantly boosts top-rank precision—particularly on disjunctive queries—while cutting mean token usage by approximately 90% per query-entity pair compared to monolithic LLM-as-reasoner baselines.

RAC: Retrieval-Augmented Clarification for Faithful Conversational Search [Kebir et al., 2026]: This work addresses the risk of conversational search systems asking fluent but “ungrounded” clarification questions, which cannot be answered by the underlying document collection. To ensure corpus faithfulness, the authors introduce RAC, a framework that couples optimized document indexing with a fine-tuned Large Language Model (LLM) trained to generate strictly evidence-based questions. By applying contrastive preference optimization, the system learns systematically to favor questions supported by retrieved passages over ungrounded alternatives. Evaluating across four benchmarks using LLM-as-Judge assessments alongside novel natural language inference (NLI) and data-to-text metrics, the study demonstrates that RAC significantly improves question faithfulness and grounding compared to traditional baselines.

Multi-Step Semantic Reasoning in Generative Retrieval [Dong et al., 2026]: Addressing the challenges of existing Generative Retrieval (GR) models to resolve complex, numerical queries over structured documents like financial reports, this paper introduces ReasonGR. The authors propose a framework that combines a structured prompting strategy—coupling task-specific instructions with stepwise reasoning guidance—with a reasoning-focused adaptation module to optimize parameter learning. Evaluated on the FinQA dataset, ReasonGR significantly improves both retrieval accuracy and consistency, demonstrating its potential to advance generative search paradigms in reasoning-intensive domains.

LLM-Assisted Pseudo-Relevance Feedback [Otero and Parapar, 2026]: This work addresses the vulnerability of classical pseudo-relevance feedback (PRF) methods caused by noisy or tangential documents in initial retrieval stages. Rather than relying on purely generative LLM expansions that risk hallucination and terminology misalignment, the authors propose a hybrid alternative that inserts an LLM-based filtering step prior to feedback estimation. Under this framework, the LLM evaluates the relevance of documents in the initial top- k ranking, ensuring that RM3 is computed exclusively over accepted, on-topic passages. Evaluated across multiple datasets, this targeted intervention significantly improves retrieval effectiveness over both blind PRF and strong baselines while maintaining classical interpretability.

WildClaims: Conversational Information Access in the Wild(Chat) [Joko et al., 2025]: Critiquing the traditional focus on explicit, search-driven conversations, this work investigates how information access occurs implicitly in real-world user-ChatGPT interactions. The authors

introduce WildClaims, a novel dataset of 121,905 check-worthy factual claims extracted from the WildChat corpus, revealing that up to 76% of seemingly non-informational dialogues (such as creative writing) actually contain assertions requiring factual verification. By mapping these implicit information needs, the study establishes a realistic benchmark that challenges conversational systems to act as active, truth-seeking gatekeepers rather than passive responders.

4 Conclusion and Outlook

The 1st Workshop on Conversational Search for Complex Information Needs brought together a rich mix of perspectives on the future of conversational search, spanning topics like intent clarification, self-optimization, agricultural RAG systems, user simulators, reasoning, faithfulness, generative retrieval, and information access.

The diversity of the presented work, alongside varied granularities of details, ranging from purely theoretical to applied perspectives, created an environment of curiosity and interaction. We discussed promising future research directions as high-level branches that we, as a community, want to expand in the domain of conversational search. We also questioned and debated the social implications of self-evolving agents, and how to keep them grounded in information and aligned with desirable behaviours. And finally, we discussed real applications in other sectors, such as agriculture, and how we can help them flourish with conversational AI.

We look forward to seeing this momentum carry into future work on conversational search. We look forward to seeing how the ideas and communication sparked here will evolve, and we plan to continue this initiative through a future second version of the workshop in 2027.

Acknowledgments

The organization of the workshop and the writing of the report were supported by the Dutch Research Council (NWO), under project number NWA.1389.20.183, 024.004.022, and KICH3.LTP.-20.006, and the European Union under grant agreement No. 101201510 (UNITE). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of their respective employers, funders, and/or granting authorities.

References

- Zahra Abbasiantaeb, Simon Lupart, and Mohammad Aliannejadi. Generating multi-aspect queries for conversational search. In Vera Demberg, Kentaro Inui, and Lluís Marquez, editors, *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8201–8217, Rabat, Morocco, March 2026. Association for Computational Linguistics. URL <https://aclanthology.org/2026.eacl-long.383/>.
- Mohammad Aliannejadi, Zahra Abbasiantaeb, Shubham Chatterjee, Jeffery Dalton, and Leif Azopardi. TREC iKAT 2023: The interactive knowledge assistance track overview. *arXiv preprint arXiv:2401.01330*, 2024a. URL <https://arxiv.org/abs/2401.01330>.

-
- Mohammad Aliannejadi, Zahra Abbasiantaeb, Shubham Chatterjee, Jeffrey Dalton, and Leif Azopardi. TREC iKAT 2023: A test collection for evaluating conversational and interactive knowledge assistants. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2024b. URL <https://doi.org/10.1145/3626772.3657860>.
- Akari Asai, Kazuma Hashimoto, Hannaneh Hajishirzi, Richard Socher, and Caiming Xiong. Learning to retrieve reasoning paths over Wikipedia graph for question answering. In *International Conference on Learning Representations*, 2020. URL <https://openreview.net/pdf?id=SJgVHkrYDH>.
- Arian Askari, Roxana Petcu, Chuan Meng, Mohammad Aliannejadi, Amin Abolghasemi, Evangelos Kanoulas, and Suzan Verberne. SOLID: Self-seeding and multi-intent self-instructing LLMs for generating intent-aware information-seeking dialogs. In Luis Chiruzzo, Alan Ritter, and Lu Wang, editors, *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 6390–6410, Albuquerque, New Mexico, April 2025. Association for Computational Linguistics. URL <https://aclanthology.org/2025.findings-naacl.357/>.
- Arthur Câmara, Vincent Slot, and Jakub Zavrel. Self-optimizing multi-agent systems for deep research. *arXiv preprint arXiv:2604.02988*, 2026. URL <https://arxiv.org/abs/2604.02988>.
- Daan L. Di Scala and Maaike H. T. de Boer. Let me explain - knowledge-based retrieval augmented generation for agricultural recommendation explanations. In *Advances in Information Retrieval: 48th European Conference on Information Retrieval, ECIR 2026*, 2026. URL https://doi.org/10.1007/978-3-032-21300-6_8.
- Steven Dong, Yubao Tang, and Maarten de Rijke. Multi-step semantic reasoning in generative retrieval. In *ECIR 2026: 48th European Conference on Information Retrieval*, 2026. URL <https://arxiv.org/abs/2603.12368>.
- Ophir Frieder, Ida Mele, Cristina Ioana Muntean, Franco Maria Nardini, Raffaele Perego, and Nicola Tonellotto. Caching historical embeddings in conversational search. *arXiv preprint arXiv:2211.14155*, 2022. URL <https://arxiv.org/abs/2211.14155>.
- Mohanna Hoveyda, Harrie Oosterhuis, Arjen P. de Vries, Maarten de Rijke, and Faegheh Hasibi. Adaptive orchestration of modular generative information access systems. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2025. URL <http://dx.doi.org/10.1145/3726302.3730351>.
- Mohanna Hoveyda, Jelle Piepenbrock, Arjen P de Vries, Maarten de Rijke, and Faegheh Hasibi. OrLog: Resolving complex queries with LLMs and probabilistic reasoning. In *ECIR 2026: 48th European Conference on Information Retrieval*, 2026. URL <https://arxiv.org/abs/2601.23085>.
- Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong Park. Adaptive-RAG: Learning to adapt retrieval-augmented large language models through question complexity. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for*

-
- Computational Linguistics: Human Language Technologies*, 2024. URL <https://arxiv.org/abs/2403.14403>.
- Bowen Jin, Hansi Zeng, Zhenrui Yue, Jinsung Yoon, Serkan O Arik, Dong Wang, Hamed Zamani, and Jiawei Han. Search-R1: Training LLMs to reason and leverage search engines with reinforcement learning. In *Second Conference on Language Modeling*, 2025. URL <https://arxiv.org/abs/2503.09516>.
- Hideaki Joko, Shakiba Amirshahi, Charles L. A. Clarke, and Faegheh Hasibi. WildClaims: Information access conversations in the Wild(Chat). *arXiv preprint arXiv:2509.17442*, 2025. URL <https://arxiv.org/abs/2509.17442>.
- Ahmed Rayane Kebir, Vincent Guigue, Lynda Said Lhadj, and Laure Soulier. RAC: Retrieval-augmented clarification for faithful conversational search. *arXiv preprint arXiv:2601.11722*, 2026. URL <https://arxiv.org/abs/2601.11722>.
- Julian Killingback and Hamed Zamani. Benchmarking information retrieval models on complex retrieval tasks. *arXiv preprint arXiv:2509.07253*, 2025. URL <https://arxiv.org/abs/2509.07253>.
- Takyoung Kim, Kyungjae Lee, Young Rok Jang, Ji Yong Cho, Gangwoo Kim, Minseok Cho, and Moontae Lee. Learning to explore and select for coverage-conditioned retrieval-augmented generation. In Luis Chiruzzo, Alan Ritter, and Lu Wang, editors, *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 1460–1480, Albuquerque, New Mexico, April 2025. Association for Computational Linguistics. ISBN 979-8-89176-195-7. doi: 10.18653/v1/2025.findings-naacl.80. URL <https://aclanthology.org/2025.findings-naacl.80/>.
- Maik Larooij. Sensitivity-aware retrieval-augmented intent clarification. *arXiv preprint arXiv:2603.06025*, 2026. URL <https://arxiv.org/abs/2603.06025>.
- Xingxuan Li, Ruochen Zhao, Yew Ken Chia, Bosheng Ding, Shafiq Joty, Soujanya Poria, and Lidong Bing. Chain-of-knowledge: Grounding large language models via dynamic knowledge adapting over heterogeneous sources. In *The Twelfth International Conference on Learning Representations*, 2024. URL <https://arxiv.org/abs/2305.13269>.
- Yidong Liang, Zhijing Wu, Fan Zhang, Dandan Song, and Heyan Huang. How users interact with generative information retrieval systems: A study of user behavior and search experience. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2025. URL <https://doi.org/10.1145/3726302.3729998>.
- Lili Lu and Fabio Crestani. Exploring user simulators in conversational search: A comparison between llms and humans. In *ECIR 2026: 48th European Conference on Information Retrieval*, 2026. URL https://doi.org/10.1007/978-3-032-21300-6_3.
- Simon Lupart, Mohammad Aliannejadi, and Evangelos Kanoulas. DiSCo: LLM knowledge distillation for efficient sparse retrieval in conversational search. *arXiv preprint arXiv:2410.14609*, 2025. URL <https://arxiv.org/abs/2410.14609>.

-
- Yougang Lyu, Xi Zhang, Xinhao Yi, Yuyue Zhao, Shuyu Guo, Wenxiang Hu, Jan Piotrowski, Jakub Kaliski, Jacopo Urbani, Zaiqiao Meng, Lun Zhou, and Xiaohui Yan. EvoScientist: Towards multi-agent evolving AI scientists for end-to-end scientific discovery. *arXiv preprint arXiv:2603.08127*, 2026. URL <https://arxiv.org/abs/2603.08127>.
- Chaitanya Malaviya, Peter Shaw, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. QUEST: A retrieval dataset of entity-seeking queries with implicit set operations. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 2023. URL <https://arxiv.org/abs/2305.11694>.
- Alex Mallen, Akari Asai, Victor Zhong, Rajarshi Das, Daniel Khashabi, and Hannaneh Hajishirzi. When not to trust language models: Investigating effectiveness of parametric and non-parametric memories. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 2023. URL <https://arxiv.org/abs/2212.10511>.
- Devadas Menon and K. Shilpa. “Chatting with ChatGPT”: Analyzing the factors influencing users’ intention to use the OpenAI’s ChatGPT using the UTAUT model. *Heliyon*, 2023.
- Cristina Ioana Muntean, Franco Maria Nardini, Raffaele Perego, Guido Rocchietti, and Cosimo Rulli. Efficient conversational search via topical locality in dense retrieval. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR ’25, 2025. URL <http://dx.doi.org/10.1145/3726302.3730186>.
- David Otero and Javier Parapar. LLM-assisted pseudo-relevance feedback. *arXiv preprint arXiv:2601.11238*, 2026. URL <https://arxiv.org/abs/2601.11238>.
- Roxana Petcu, Samarth Bhargav, Maarten de Rijke, and Evangelos Kanoulas. A comprehensive taxonomy of negation for NLP and neural retrievers. In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng, editors, *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 15511–15533, Suzhou, China, November 2025. Association for Computational Linguistics. URL <https://aclanthology.org/2025.findings-emnlp.839/>.
- Roxana Petcu, Evangelos Kanoulas, and Maarten de Rijke. SubSearch: Intermediate rewards for unsupervised guided reasoning in complex retrieval, 2026. URL <https://arxiv.org/abs/2604.07415>.
- Flor Miriam Plaza-del Arco, Amanda Cercas Curry, Susanna Paoli, Alba Cercas Curry, and Dirk Hovy. Divine LLaMAs: Bias, stereotypes, stigmatization, and emotion representation of religion in large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*. URL <https://arxiv.org/abs/2407.06908>.
- Alireza Salemi, Julian Killingback, and Hamed Zamani. ExPerT: Effective and explainable evaluation of personalized long-form text generation. In *Findings of the Association for Computational Linguistics: ACL 2025*, 2025. URL <https://arxiv.org/abs/2501.14956>.

-
- Sofia Eleni Spatharioti, David Rothschild, Daniel G Goldstein, and Jake M Hofman. Effects of LLM-based search on decision making: Speed, accuracy, and overreliance. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 2025. URL <https://doi.org/10.1145/3706598.3714082>.
- Weihang Su, Qingyao Ai, Jingtao Zhan, Qian Dong, and Yiqun Liu. Dynamic and parametric retrieval-augmented generation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, 2025. URL <https://arxiv.org/abs/2506.06704>.
- Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. Interleaving retrieval with chain-of-thought reasoning for knowledge-intensive multi-step questions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 2023. URL <https://arxiv.org/abs/2212.10509>.
- Coen van den Elsen, Francien Barkhof, Thijmen Nijdam, Simon Lupart, and Mohammad Alian-nejadi. Reproducing NevIR: Negation in neural information retrieval. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '25, 2025. URL <http://dx.doi.org/10.1145/3726302.3730294>.
- Jonas Wallat, Maria Heuss, Maarten de Rijke, and Avishek Anand. Correctness is not faithfulness in retrieval augmented generation attributions. ICTIR '25, 2025. URL <https://arxiv.org/abs/2412.18004>.
- Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=WE_vluYUL-X.
- Mert Yazan, Suzan Verberne, and Frederik Situmeang. Improving RAG for personalization with author features and contrastive examples. In *Advances in Information Retrieval: 47th European Conference on Information Retrieval, ECIR 2025, Lucca, Italy, April 6–10, 2025, Proceedings, Part III*, 2025. URL <https://arxiv.org/abs/2504.08745>.
- Wenhao Zhang, Mengqi Zhang, Shiguang Wu, Jiahuan Pei, Zhaochun Ren, Maarten de Rijke, Zhumin Chen, and Pengjie Ren. ExcluIR: Exclusionary neural information retrieval. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025. URL <https://arxiv.org/abs/2404.17288>.