Introduction to the Special Issue on Query Performance Prediction

GUGLIELMO FAGGIOLI, University of Padua, Italy
DEBASIS GANGULY, University of Glasgow, UK
SUCHANA DATTA, University College Dublin, Ireland
NICOLA FERRO, University of Padua, Italy
IADH OUNIS, University of Glasgow, United Kingdom

Query Performance Prediction (QPP) aims to estimate the effectiveness of a retrieval system for a given query without requiring relevance judgments. While traditionally studied for sparse retrieval, recent advances in neural ranking, dense retrieval, and large language models (LLMs) have prompted a shift towards QPP methods that better reflect modern IR systems. This Special Issue on *Query Performance Prediction Towards Novel Information Retrieval Paradigms* presents recent advances along two complementary directions. The first focuses on LLM- and representation-based approaches that leverage query variants, semantic interactions, and comparative assessment to improve prediction for neural retrieval models. The second explores new formulations and application settings, including inverse learning, prediction of reliability in retrieval-augmented generation, and the incorporation of multimodal and neurophysiological signals. These works highlight emerging challenges in evaluating QPP in neural and generative settings by demonstrating that QPP is evolving into a broader framework for estimating uncertainty and reliability across complex IR pipelines. We believe that this Special Issue will foster further research towards robust and generalizable QPP methods for next-generation information access systems.

ACM Reference Format:
Guglielmo Faggioli, Debasis Ganguly, Suchana Datta, Nicola Ferro, and Iadh Ounis. 2018. Introduction to the Special Issue on Query Performance Prediction. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Query Performance Prediction (QPP) is a long-standing problem in Information Retrieval (IR) that seeks to estimate how well a retrieval system is likely to perform for a given query, in the absence of explicit relevance judgments [3, 30, 36, 37]. At its core, QPP addresses a fundamental uncertainty in search: while a retrieval system can return a ranked list of documents for almost any input query, it does not intrinsically know whether the resulting ranking is likely to satisfy the user's information need. Accurate performance prediction is therefore valuable for a wide range of query-specific interventions, including selective query expansion [10], adaptive re-ranking [39, 45], clarification [1], result diversification [28], mixed-initiative conversational search actions such as clarification and query reformulation [20, 27, 30], and deciding when more expensive retrieval or generation components should be invoked [21].

Authors' Contact Information: Guglielmo Faggioli, guglielmo.faggioli@unipd.it, University of Padua, Padova, Italy; Debasis Ganguly, debasis.ganguly@glasgow.ac.uk, University of Glasgow, Glasgow, UK; Suchana Datta, suchana.datta@ucd.ie, University College Dublin, Dublin, Ireland; Nicola Ferro, ferro@dei.unipd.it, University of Padua, Padova, Italy; Iadh Ounis, University of Glasgow, Glasgow, United Kingdom, iadh.ounis@glasgow.ac.uk.

Historically, QPP research has been closely tied to classical ad-hoc retrieval settings, where queries are short textual expressions, documents are represented primarily through lexical statistics, and retrieval effectiveness is measured against test collections with human relevance judgments [15]. In this setting, a rich set of pre-retrieval and post-retrieval predictors has been developed, exploiting signals such as query specificity [19], collection statistics [18], score distributions [36, 37], clarity [6, 19], robustness to perturbation [33, 36], and the agreement among retrieved documents [11, 44]. These methods have provided important insights into why some queries are inherently easier than others and how retrieval systems can adapt their behaviour on a per-query basis.

However, contemporary IR systems differ substantially from the settings in which many traditional QPP methods were originally developed [13, 14, 35]. Modern IR pipelines increasingly rely on neural architectures, including dense retrieval and cross-encoder reranking models, which represent queries and documents in continuous vector spaces and capture rich semantic interactions beyond exact term matching [22, 26]. The recent integration of large language models (LLMs) into IR further shifts the paradigm towards systems that combine retrieval, reasoning, and generation capabilities [46]. In particular, retrieval-augmented generation (RAG) pipelines tightly couple neural retrieval with generative models, enabling knowledge-grounded answer generation but also introducing new sources of uncertainty and evaluation challenges [24, 32].

In parallel, IR systems are increasingly deployed in conversational search settings, where multi-turn interactions, query reformulation, and clarification play a central role, as well as in multimodal retrieval environments, which combine textual, visual, and other modalities for effective information access [4, 31]. These developments suggest that retrieval effectiveness is no longer determined solely by lexical matching or even ranking quality, but by how well systems capture semantics, context, and interaction dynamics. Moreover, the notion of "performance" itself is expanding. In addition to rank-based effectiveness measures such as NDCG, MAP, or reciprocal rank, LLM-based answer generation systems may need to predict factuality, answer reliability [41], and hallucination risk [2] in generated responses. Other aspects such as user effort in relevance assessments [16, 43], fairness [23], robustness, and the overall quality of generated answers are also becoming important targets of evaluation [41].

This Special Issue on *Query Performance Prediction Towards Novel Information Retrieval Paradigms* brings together recent work that builds on and extends an already active line of research on QPP for modern IR systems. In recent years, QPP approaches have moved beyond lexical retrieval settings, aiming to predict the performance of neural and dense retrieval models that rely on semantic representations and learned query–document interactions [3, 8, 13, 34]. More broadly, there has been a growing interest in designing predictors that are better aligned with the behaviour of neural ranking architectures and that exploit representation-level signals [13, 38] rather than traditional term-based statistics.

Against this backdrop, the papers accepted in this Special Issue further advance this trajectory. Collectively, they demonstrate that QPP is no longer confined to predicting the effectiveness of lexical document retrieval alone, but is increasingly being viewed as a broader framework for estimating uncertainty, reliability, and expected utility across complex IR and retrieval-augmented generation pipelines.

At a general level, the papers in this Special Issue can be grouped into two main categories. The first category consists of works that develop new LLM- and neural-representation-based approaches for QPP, with particular emphasis on query variants, semantic interactions, and comparative assessment of retrieved results. These papers investigate how auxiliary query evidence, large language models, contrastive representation learning, and pairwise comparison strategies can be better used to estimate the effectiveness of modern retrieval systems.

The second category consists of papers that broaden the conceptual and application scope of QPP. These works include a novel inverse-learning formulation of QPP, the use of QPP-like signals for assessing reliability in medical retrieval-augmented question answering, and the incorporation of neurophysiological and behavioural signals for predicting performance in reading and listening tasks. Together, these papers show that QPP is evolving from a technique for estimating ranked-list effectiveness into a broader framework for reasoning about uncertainty, reliability, and expected performance in diverse information access settings.

1.1 LLM-based QPP: Query Variants, Semantic Interaction, and Comparative Assessment

The first group of papers investigates how QPP can benefit from the representational and generative capabilities of large language models, as well as from richer forms of interaction between queries and retrieved results [3, 9, 12]. These papers are motivated by the observation that traditional QPP methods, although often effective for lexical retrieval models, may be less well aligned with neural rankers whose behaviour depends on semantic matching, contextual representations, and learned query-document interactions.

The paper *RAQG-QPP: Query Performance Prediction with Retrieved Query Variants and Retrieval Augmented Query Generation* studies the role of query variants in improving QPP for neural retrieval systems [42]. Prior work has shown that variants of the original query can provide useful evidence about the stability, ambiguity, and specificity of an information need. However, variants produced through simple term expansion may be incoherent, off-topic, or insufficiently grounded. To address this limitation, the authors propose retrieving query variants from logs of past queries and then using LLMs to generate additional variants conditioned on these retrieved examples. This combines the realism of historical user queries with the flexibility of generation. Experiments on TREC Deep Learning 2019 and 2020 show that this retrieval-augmented query generation strategy improves QPP accuracy, particularly for neural ranking models such as MonoT5. The work illustrates how LLMs can be used not merely as direct predictors, but also as generators of auxiliary evidence for estimating retrieval effectiveness.

A different strategy is illustrated in *Adaptive Query Performance Prediction for Retrieval Augmented Generation: Bridging Retrieval Quality and Generation Relevance* [39]. More specifically, the approach proposed in the manuscript employs a learning-based QPP that employs multiple signals to predict the performance in a Retrieval Augmented Generation scenario. The signals investigated in [39] are designed to explore different facets of the query and the retrieved documents. In particular, they consider similarity-based attributes to capture similarity aspects between documents and passages, ranking indicators to extract signals of sudden changes in the retrieval scores of the retrieved documents, semantic consistency measures to capture the geometric spread of the passage embeddings, lexical signals, and traditional QPP scores. *Angle-QPP: Improving Query Performance Prediction through Large Language Models and Angle Interaction in Complex Vector Space* [25] focuses on a neural QPP framework that combines LLM-based representations, contrastive learning, and an angle-based interaction mechanism in complex embedding space. The motivation is that existing binary-encoder and cross-encoder approaches may not fully capture the subtle semantic differences between queries and retrieved documents that are important for QPP. Angle-QPP addresses this by first using a contrastive warm-up phase to improve representation quality, and then applying an angular interaction mechanism to model fine-grained query-document relationships. Across TREC DL 2019–2022, the method substantially improves over previous binary-encoder QPP models, with larger LLM backbones yielding a stronger performance. This paper therefore contributes both a new architecture for semantic interaction and empirical evidence that scaling language-model representations can benefit QPP.

The manuscript *Beyond Pointwise: Exploring Comparative Approaches for Predicting Query Performance Using LLMs* examines how LLMs should be used when assessing the quality of retrieved results [17]. Existing LLM-based QPP methods often rely on pointwise judgments, where an LLM assesses retrieved items independently. This paper instead investigates comparative approaches, including pairwise comparisons between retrieved elements or between selected reference items. The study evaluates Llama-3.1-8B-Instruct and Qwen3-8B under both fine-tuned and non-fine-tuned settings, comparing the predicted performance with human-judged measures such as NDCG@10 and RR@10. The results show that pairwise approaches often outperform pointwise baselines, especially in settings without fine-tuning, although their behaviour varies across retrieval models such as ANCE and BM25. The paper also addresses computational efficiency by reducing the number of LLM inferences through comparisons against selected subsets of reference items.

These manuscripts highlight how QPP for modern retrieval systems increasingly depends on richer forms of evidence than those used in traditional lexical predictors. Query variants can expose ambiguity and robustness; using signals from different sources could improve QPP effectiveness; LLM-based representations can capture semantic relationships between queries and documents; and comparative assessment can reveal relative quality differences among the retrieved results. This research line therefore points toward a more interaction-aware and semantically grounded view of QPP.

1.2 Novel Formulations and Applications of QPP

The second research line investigated by the papers in the special issue broadens the scope of QPP beyond the conventional prediction of ranked-list effectiveness. These works ask how QPP can be reformulated, where QPP-like signals can be applied, and what types of evidence may be useful when the task is no longer standard ad-hoc retrieval. They show that QPP is becoming relevant for predicting search effectiveness and understanding retrieval behaviour, assessing answer reliability, and modelling performance in human-centred information processing tasks.

The paper *Reversing the Retrieval Engine: Query Performance Prediction as an Inverse Learning Task* proposes a novel formulation of QPP as an inverse learning problem [29]. Rather than learning a direct mapping from retrieval signals to effectiveness scores, the paper starts from the hypothesis that observed effectiveness values arise from latent query and retrieval characteristics. The proposed inverse regression framework maps performance labels, such as NDCG@10, back to estimated retrieval features including clarity, score variance, and entropy. In doing so, it seeks to reconstruct the latent retrieval properties that explain the observed performance. This formulation changes the role of QPP from direct score estimation to explanatory reconstruction. Experiments on MS MARCO and TREC Robust04 show that the inverse model substantially improves rank-correlation performance over a conventional forward baseline, suggesting that reversing the retrieval process can offer both predictive and diagnostic benefits.

The paper *Augmenting Small Language Model for Better Medical Question Answering through Source Authentication* extends QPP from ranked-list prediction to the reliability of answers produced by a medical retrieval-augmented generation pipeline [7]. Medical question answering is a high-stakes setting in which factual reliability and computational efficiency are both essential. The authors introduce a resource-constrained framework that integrates a fine-tuned 1.5B-parameter Small Language Model with retrieval-augmented generation and a calibrated fact-checking module. A central contribution of the paper is the use of internal system metrics, particularly a Factual Consistency Score, as QPP-like indicators of answer reliability. This work broadens the meaning of QPP in modern IR pipelines: in RAG systems, performance depends on whether relevant documents are retrieved and the downstream generation faithfully uses the retrieved evidence.

The paper *On the Use of Electroencephalography in Query Performance Prediction* investigates whether neurophysiological signals can improve QPP beyond text-only features [40]. The study uses datasets in which textual queries are associated with EEG and eye-tracking signals collected during reading and listening tasks. The authors propose an ensemble architecture with modality-specific models for text, EEG, and eye-tracking data, followed by a meta-learner that produces final predictions. The results show that generalized models trained across subjects can improve over text-only baselines under certain conditions. In particular, trimodal models combining EEG, eye-tracking, and text perform well for reading tasks, while EEG-plus-text models improve prediction in listening settings. Query-level analysis further suggests that neurophysiological signals are especially useful for fragmentary or semantically ambiguous queries.

In *Uncovering the Limitations of Query Performance Prediction: Failures, Insights, and Implications for Selective Query Processing* [5], the authors analyse the generalization capabilities of QPP methods and their robustness. The authors consider both traditional lexical-based QPP approaches, such as NQC and WIG, approaches based on Learning-to-Rank (LETOR) features, and dense predictors. Furthermore, their analysis focuses on different collection categories and considers classes of target IR systems, including lexical and semantic sparse models, bi-encoders, and late interaction approaches. Their findings show high variability in accuracy, with the collection being the most prominent factor in causing changes in QPP performance. Furthermore, their analysis shows the limits of applying QPP in downstream tasks, such as QPP-driven selective query processing. Their work calls for revisiting QPP, especially in light of novel IR paradigms, as well as redesigning QPP evaluation approaches targeting downstream applications.

Together, these papers expand the boundaries of QPP in different complementary ways. The inverse learning paper [29] reconceptualizes QPP as a means of reconstructing latent retrieval properties; the medical QA paper [7] applies QPP-like reliability estimation to retrieval-augmented answer generation; and the EEG paper introduces human-centred signals for predicting performance in reading and listening tasks. Finally, the evaluation-focused paper [5] illustrates the need for revising current QPP research, both in terms of approaches, with the objective of improving their robustness across IR settings, but also in terms of QPP evaluation, which should focus on downstream tasks.

2 Discussion and Outlook

The contributions in this Special Issue point to two broad trends in the evolution of QPP. The first is methodological: QPP methods are becoming increasingly semantic, generative, and interaction-aware. The use of retrieved and generated query variants, LLM-based semantic representations, complex-space interaction mechanisms, and pairwise comparative assessment reflects a shift away from purely lexical or score-distribution-based predictors toward models that better match the behaviour of neural retrieval systems and the semantics of queries and corpora.

The second trend concerns scope: QPP is moving beyond the prediction of ranked-list effectiveness in standard ad-hoc retrieval. The contributions in this issue show that QPP can be formulated as an inverse reconstruction problem, used to estimate answer reliability in retrieval-augmented medical question answering, and enriched with neurophysiological and behavioural signals in reading and listening tasks. These developments suggest that QPP is becoming a general framework for predicting the success, reliability, or difficulty of information access processes.

Finally, across both strands of work, this Special Issue highlights the continuing importance of *evaluation*. As QPP expands beyond classical ad-hoc retrieval, the field must now develop evaluation methodologies that are appropriate for neural rankers, LLM-based predictors, RAG pipelines, multimodal settings, and answer generation. This includes new datasets, metrics, and a careful analysis of computational cost, generalization across collections, interpretability, and robustness. At the same time, the contributions highlight how, due to the advent of new IR evaluation paradigms, it is necessary to revisit how we evaluate QPP, focusing on the actual downstream task the QPP model will address.

In summary, the contributions in this Special Issue demonstrate that QPP remains a central and evolving problem in Information Retrieval (IR). Far from being limited to predicting the effectiveness of traditional search queries, QPP now provides a framework for reasoning about uncertainty, reliability, and adaptability in modern IR systems. The papers collected here show how inverse learning, LLMs, query variants, comparative assessment, factual consistency estimation, and neurophysiological evidence can all contribute to this broader vision. We hope that this Special Issue will stimulate further research into QPP methods that are accurate, interpretable, efficient, and applicable across the next generation of retrieval and retrieval-augmented systems.

Acknowledgments

The guest editors thank the ACM Transactions on Information Systems (TOIS) journal for the opportunity to host this Special Issue. We particularly thank Prof. Min Zhang, the Editor-in-Chief of the journal, and the ACM TOIS editorial and production teams for their support throughout the review and publication process. We also thank all authors for their submissions, and the reviewers for their thoughtful and constructive feedback, which helped shape the final set of accepted papers.

References

[1] Mohammad Aliannejadi, Hamed Zamani, Fabio Crestani, and W. Bruce Croft. 2019. Asking Clarifying Questions in Open-Domain Information-Seeking Conversations. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019*, Benjamin Piwowarski, Max Chevalier, Éric Gaussier, Yoelle Maarek, Jian-Yun Nie, and Falk Scholer (Eds.). ACM, 475–484. doi:10.1145/3331184.3331265

[2] Dang Anh-Hoang, Vu Tran, and Le-Minh Nguyen. 2025. Survey and analysis of hallucinations in large language models: attribution to prompting strategies or model behavior. *Frontiers Artif. Intell.* 8 (2025). doi:10.3389/FRAI.2025.1622292

[3] Negar Arabzadeh, Maryam Khodabakhsh, and Ebrahim Bagheri. 2021. BERT-QPP: Contextualized Pre-trained transformers for Query Performance Prediction. In *CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, November 1 - 5, 2021*, Gianluca Demartini, Guido Zuccon, J. Shane Culpepper, Zi Huang, and Hanghang Tong (Eds.). ACM, 2857–2861. doi:10.1145/3459637.3482063

[4] Gang Cheng, Yaxi Wu, Desheng Cao, et al. 2024. A Survey of Multimodal Retrieval: From Traditional Analytics to Deep Learning. *Proc. SPIE* (2024).

[5] Adrian-Gabriel Chifu, Sébastien Déjean, Moncef Garouani, Josiane Mothe, Diégo Ortiz, and Md Zia Ullah. 2025. Uncovering the Limitations of Query Performance Prediction: Failures, Insights, and Implications for Selective Query Processing. *ACM Trans. Inf. Syst.* (Nov. 2025). Just Accepted. doi:10.1145/3774427

[6] Stephen Cronen-Townsend, Yun Zhou, and W. Bruce Croft. 2002. Predicting query performance. In *SIGIR 2002: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 11-15, 2002, Tampere, Finland*, Kalervo Järvelin, Micheline Beaulieu, Ricardo A. Baeza-Yates, and Sung-Hyon Myaeng (Eds.). ACM, 299–306. doi:10.1145/564376.564429

[7] Sourav Das, Sanjay Chatterji, and Imon Mukherjee. 2026. Augmenting Small Language Model for Better Medical Question Answering through Source Authentication. *ACM Trans. Inf. Syst.* (Feb. 2026). Just Accepted. doi:10.1145/3797887

[8] Suchana Datta, Guglielmo Faggioli, Nicola Ferro, Debasis Ganguly, Cristina Ioana Muntean, Raffaele Perego, and Nicola Tonellotto. 2026. Projection-Displacement-Based Query Performance Prediction for Embedded Space of Dense Retrievers. *ACM Trans. Inf. Syst.* 44, 1 (2026), 7:1–7:30. doi:10.1145/3765617

[9] Suchana Datta, Debasis Ganguly, Derek Greene, and Mandar Mitra. 2022. Deep-QPP: A Pairwise Interaction-based Deep Learning Model for Supervised Query Performance Prediction. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining* (Virtual Event, AZ, USA) *(WSDM '22)*. Association for Computing Machinery, New York, NY, USA, 201–209. https://doi.org/10.1145/3488560.3498491

[10] Suchana Datta, Debasis Ganguly, Sean MacAvaney, and Derek Greene. 2024. A Deep Learning Approach for Selective Relevance Feedback. In *Advances in Information Retrieval - 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24-28, 2024, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 14609)*, Nazli Goharian, Nicola Tonellotto, Yulan He, Aldo Lipani, Graham McDonald, Craig Macdonald, and Iadh Ounis (Eds.). Springer, 189–204. doi:10.1007/978-3-031-56060-6_13

[11] Suchana Datta, Debasis Ganguly, Mandar Mitra, and Derek Greene. 2023. A Relative Information Gain-based Query Performance Prediction Framework with Generated Query Variants. *ACM Trans. Inf. Syst.* 41, 2 (2023), 38:1–38:31. doi:10.1145/3545112

[12] Suchana Datta, Sean MacAvaney, Debasis Ganguly, and Derek Greene. 2022. A 'Pointwise-Query, Listwise-Document' based Query Performance Prediction Approach. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*, Enrique Amigó, Pablo Castells, Julio Gonzalo, Ben Carterette, J. Shane Culpepper, and Gabriella Kazai (Eds.). ACM, 2148–2153. doi:10.1145/3477495.3531821

[13] Guglielmo Faggioli, Thibault Formal, Simon Lupart, Stefano Marchesin, Stéphane Clinchant, Nicola Ferro, and Benjamin Piwowarski. 2023. Towards Query Performance Prediction for Neural Information Retrieval: Challenges and Opportunities. In *Proceedings of the 2023 ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR 2023, Taipei, Taiwan, 23 July 2023*, Masaharu Yoshioka, Julia Kiseleva, and Mohammad Aliannejadi (Eds.). ACM, 51–63. doi:10.1145/3578337.3605142

[14] Guglielmo Faggioli, Thibault Formal, Stefano Marchesin, Stéphane Clinchant, Nicola Ferro, and Benjamin Piwowarski. 2023. Query Performance Prediction for Neural IR: Are We There Yet?. In *Advances in Information Retrieval - 45th European Conference on Information Retrieval, ECIR 2023,*

Dublin, Ireland, April 2-6, 2023, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 13980), Jaap Kamps, Lorraine Goeuriot, Fabio Crestani, Maria Maistro, Hideo Joho, Brian Davis, Cathal Gurrin, Udo Kruschwitz, and Annalina Caputo (Eds.). Springer, 232–248. doi:10.1007/978-3-031-28244-7_15

[15] Debasis Ganguly, Suchana Datta, Mandar Mitra, and Derek Greene. 2022. An Analysis of Variations in the Effectiveness of Query Performance Prediction. In *Advances in Information Retrieval - 44th European Conference on IR Research, ECIR 2022, Stavanger, Norway, April 10-14, 2022, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 13185)*, Matthias Hagen, Suzan Verberne, Craig Macdonald, Christin Seifert, Krisztian Balog, Kjetil Nørvåg, and Vinay Setty (Eds.). Springer, 215–229. doi:10.1007/978-3-030-99736-6_15

[16] Debasis Ganguly and Emine Yilmaz. 2023. Query-specific Variable Depth Pooling via Query Performance Prediction. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Taipei, Taiwan) *(SIGIR '23)*. Association for Computing Machinery, New York, NY, USA, 2303–2307. doi:10.1145/3539618.3592046

[17] Cezary Golecki, Adrian Racki, Jakub Król, Pawel Zawistowski, and Aleksander Wawer. 2026. Beyond Pointwise: Exploring Comparative Approaches for Predicting Query Performance Using LLMs. *ACM Trans. Inf. Syst.* (Feb. 2026). Just Accepted. doi:10.1145/3793556

[18] Claudia Hauff, Djoerd Hiemstra, and Franciska de Jong. 2008. A survey of pre-retrieval query performance predictors. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008, Napa Valley, California, USA, October 26-30, 2008*, James G. Shanahan, Sihem Amer-Yahia, Ioana Manolescu, Yi Zhang, David A. Evans, Aleksander Kolcz, Key-Sun Choi, and Abdur Chowdhury (Eds.). ACM, 1419–1420. doi:10.1145/1458082.1458311

[19] Ben He and Iadh Ounis. 2004. Inferring Query Performance Using Pre-retrieval Predictors. In *String Processing and Information Retrieval, 11th International Conference, SPIRE 2004, Padova, Italy, October 5-8, 2004, Proceedings (Lecture Notes in Computer Science, Vol. 3246)*, Alberto Apostolico and Massimo Melucci (Eds.). Springer, 43–54. doi:10.1007/978-3-540-30213-1_5

[20] Yunah Jang, Kang-il Lee, Hyunkyung Bae, Hwanhee Lee, and Kyomin Jung. 2024. IterCQR: Iterative Conversational Query Reformulation with Retrieval Guidance. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), NAACL 2024, Mexico City, Mexico, June 16-21, 2024*, Kevin Duh, Helena Gómez-Adorno, and Steven Bethard (Eds.). Association for Computational Linguistics, 8121–8138. doi:10.18653/V1/2024.NAACL-LONG.449

[21] Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong Park. 2024. 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 (Volume 1: Long Papers), NAACL 2024, Mexico City, Mexico, June 16-21, 2024*, Kevin Duh, Helena Gómez-Adorno, and Steven Bethard (Eds.). Association for Computational Linguistics, 7036–7050. doi:10.18653/V1/2024.NAACL-LONG.389

[22] Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. 2020. Dense Passage Retrieval for Open-Domain Question Answering. (2020), 6769–6781. doi:10.18653/V1/2020.EMNLP-MAIN.550

[23] Ömer Kirnap, Fernando Diaz, Asia Biega, Michael D. Ekstrand, Ben Carterette, and Emine Yilmaz. 2021. Estimation of Fair Ranking Metrics with Incomplete Judgments. In *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, Jure Leskovec, Marko Grobelnik, Marc Najork, Jie Tang, and Leila Zia (Eds.). ACM / IW3C2, 1065–1075. doi:10.1145/3442381.3450080

[24] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (Eds.). https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

[25] Xianming Li and Jing Li. 2026. Angle-QPP: Improving Query Performance Prediction through Large Language Models and Angle Interaction in Complex Vector Space. *ACM Trans. Inf. Syst.* (Jan. 2026). Just Accepted. doi:10.1145/3793857

[26] Jimmy Lin, Rodrigo Nogueira, and Andrew Yates. 2021. Pretrained Transformers for Text Ranking: BERT and Beyond. (2021). doi:10.2200/S01123ED1V01Y202108HLT053

[27] Yosi Mass, Doron Cohen, Asaf Yehudai, and David Konopnicki. 2022. Conversational Search with Mixed-Initiative - Asking Good Clarification Questions backed-up by Passage Retrieval. In *Proceedings of the Second DialDoc Workshop on Document-grounded Dialogue and Conversational Question Answering*. Association for Computational Linguistics, Dublin, Ireland, 65–71. doi:10.18653/v1/2022.dialdoc-1.7

[28] David Maxwell, Leif Azzopardi, and Yashar Moshfeghi. 2019. The impact of result diversification on search behaviour and performance. *Inf. Retr. J.* 22, 5 (2019), 422–446. doi:10.1007/S10791-019-09353-0

[29] Rabeb Mbarek and Hawete Hattab. 2026. Reversing the Retrieval Engine: Query Performance Prediction as an Inverse Learning Task. *ACM Trans. Inf. Syst.* (Feb. 2026). Just Accepted. doi:10.1145/3790102

[30] Chuan Meng, Negar Arabzadeh, Mohammad Aliannejadi, and Maarten de Rijke. 2023. Query Performance Prediction: From Ad-hoc to Conversational Search. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Taipei, Taiwan) *(SIGIR'23)*. Association for Computing Machinery, New York, NY, USA, 2583–2593. https://doi.org/10.1145/3539618.3591919

[31] Chuan Meng, Negar Arabzadeh, Mohammad Aliannejadi, and Maarten de Rijke. 2023. Query Performance Prediction: From Ad-hoc to Conversational Search. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023*, Hsin-Hsi Chen, Wei-Jou (Edward) Duh, Hen-Hsen Huang, Makoto P. Kato, Josiane Mothe, and Barbara Poblete (Eds.). ACM, 2583–2593. doi:10.1145/3539618.3591919

[32] Andrew Parry, Debasis Ganguly, and Manish Chandra. 2024. "In-Context Learning" or: How I learned to stop worrying and love "Applied Information Retrieval". In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Washington DC,

USA) *(SIGIR '24)*. Association for Computing Machinery, New York, NY, USA, 14–25. https://doi.org/10.1145/3626772.3657842

[33] Haggai Roitman. 2017. An Enhanced Approach to Query Performance Prediction Using Reference Lists. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, August 7-11, 2017*, Noriko Kando, Tetsuya Sakai, Hideo Joho, Hang Li, Arjen P. de Vries, and Ryen W. White (Eds.). ACM, 869–872. doi:10.1145/3077136.3080665

[34] Abbas Saleminezhad, Negar Arabzadeh, Radin Hamidi Rad, Soosan Beheshti, and Ebrahim Bagheri. 2025. Robust query performance prediction for dense retrievers via adaptive disturbance generation. *Mach. Learn.* 114, 3 (2025), 65. doi:10.1007/S10994-024-06659-Z

[35] Payel Santra, Partha Basuchowdhuri, and Debasis Ganguly. 2026. Breaking Flat: A Generalised Query Performance Prediction Evaluation Framework. In *Advances in Information Retrieval - 48th European Conference on Information Retrieval, ECIR 2026, Delft, The Netherlands, March 29 - April 2, 2026, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 16484)*, Ricardo Campos, Adam Jatowt, Yanyan Lan, Mohammad Aliannejadi, Christine Bauer, Sean MacAvaney, Avishek Anand, Zhaochun Ren, Suzan Verberne, Nan Bai, and Masoud Mansoury (Eds.). Springer, 155–171. doi:10.1007/978-3-032-21300-6_10

[36] Anna Shtok, Oren Kurland, and David Carmel. 2010. Using statistical decision theory and relevance models for query-performance prediction. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Geneva, Switzerland) *(SIGIR '10)*. Association for Computing Machinery, New York, NY, USA, 259–266. doi:10.1145/1835449.1835494

[37] Anna Shtok, Oren Kurland, David Carmel, Fiana Raiber, and Gad Markovits. 2012. Predicting Query Performance by Query-Drift Estimation. *ACM Trans. Inf. Syst.* 30, 2, Article 11 (May 2012), 35 pages. doi:10.1145/2180868.2180873

[38] Ashutosh Singh, Debasis Ganguly, Suchana Datta, and Craig MacDonald. 2023. Unsupervised Query Performance Prediction for Neural Models with Pairwise Rank Preferences. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023*, Hsin-Hsi Chen, Wei-Jou (Edward) Duh, Hen-Hsen Huang, Makoto P. Kato, Josiane Mothe, and Barbara Poblete (Eds.). ACM, 2486–2490. doi:10.1145/3539618.3592082

[39] Kunal Sinha, Aparajita; Chakma. 2026. Adaptive Query Performance Prediction for Retrieval Augmented Generation: Bridging Retrieval Quality and Generation Relevance. *ACM Trans. Inf. Syst.* (Feb. 2026). Just Accepted.

[40] Iakovos Tenedios, Niall McGuire, and Yashar Moshfeghi. 2026. On The Use Of Electroencephalography In Query Performance Prediction. *ACM Trans. Inf. Syst.* (2026). Just Accepted. doi:10.1145/3816244

[41] Fangzheng Tian, Debasis Ganguly, and Craig Macdonald. 2026. Predicting Retrieval Utility and Answer Quality in Retrieval-Augmented Generation. In *Advances in Information Retrieval - 48th European Conference on Information Retrieval, ECIR 2026, Delft, The Netherlands, March 29 - April 2, 2026, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 16483)*, Ricardo Campos, Adam Jatowt, Yanyan Lan, Mohammad Aliannejadi, Christine Bauer, Sean MacAvaney, Avishek Anand, Zhaochun Ren, Suzan Verberne, Nan Bai, and Masoud Mansoury (Eds.). Springer, 368–385. doi:10.1007/978-3-032-21289-4_24

[42] Fangzheng Tian, Debasis Ganguly, and Craig MacDonald. 2026. RAQG-QPP: Query Performance Prediction with Retrieved Query Variants and Retrieval Augmented Query Generation. *ACM Trans. Inf. Syst.* (May 2026). Just Accepted. doi:10.1145/3815198

[43] Manisha Verma, Emine Yilmaz, and Nick Craswell. 2016. On Obtaining Effort Based Judgements for Information Retrieval. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, San Francisco, CA, USA, February 22-25, 2016*, Paul N. Bennett, Vanja Josifovski, Jennifer Neville, and Filip Radlinski (Eds.). ACM, 277–286. doi:10.1145/2835776.2835840

[44] Oleg Zendel, Anna Shtok, Fiana Raiber, Oren Kurland, and J. Shane Culpepper. 2019. Information Needs, Queries, and Query Performance Prediction. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019*, Benjamin Piwowarski, Max Chevalier, Éric Gaussier, Yoelle Maarek, Jian-Yun Nie, and Falk Scholer (Eds.). ACM, 395–404. doi:10.1145/3331184.3331253

[45] Longhui Zhang, Yanzhao Zhang, Dingkun Long, Pengjun Xie, Meishan Zhang, and Min Zhang. 2024. A Two-Stage Adaptation of Large Language Models for Text Ranking. In *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and virtual meeting, August 11-16, 2024 (Findings of ACL, Vol. ACL 2024)*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, 11880–11891. doi:10.18653/V1/2024.FINDINGS-ACL.706

[46] Yutao Zhu, Huaying Yuan, Shuting Wang, Jiongnan Liu, Wenhan Liu, Chenlong Deng, Haonan Chen, Zheng Liu, Zhicheng Dou, and Ji-Rong Wen. 2026. Large Language Models for Information Retrieval: A Survey. *ACM Trans. Inf. Syst.* 44, 1 (2026), 12:1–12:54. doi:10.1145/3748304