SHORT-PAPER

Workshop on Explainability in Information Retrieval

MARIA HEUSS, University of Amsterdam, Amsterdam, Noord-Holland, Netherlands

CATHERINE CHEN, Brown University, Providence, RI, United States

AVISHEK ANAND, Delft University of Technology, Delft, Zuid-Holland, Netherlands

CARSTEN EICKHOFF, University of Tübingen, Tubingen, Baden-Wurttemberg, Germany

SUZAN VERBERNE, Leiden University, Leiden, Zuid-Holland, Netherlands

Open Access Support provided by:

Brown University

University of Amsterdam

Leiden University

University of Tübingen

Delft University of Technology

Published: 13 July 2025

Citation in BibTeX format

SIGIR '25: The 48th International ACM SIGIR Conference on Research and Development in Information Retrieval
July 13 - 18, 2025
Padua, Italy

Conference Sponsors:
SIGIR

Workshop on Explainability in Information Retrieval

Maria Heuss
m.c.heuss@uva.nl
University of Amsterdam
IRLab
Amsterdam, The Netherlands

Catherine Chen
catherine_s_chen@brown.edu
Brown University
Providence, Rhode Island, USA

Avishek Anand
Avishek.Anand@tudelft.nl
Delft Institute of Technology
Delft, The Netherlands

Carsten Eickhoff
c.eickhoff@acm.org
University of Tübingen
Tübingen, Germany

Suzan Verberne
s.verberne@liacs.leidenuniv.nl
Leiden University
Leiden Institute of Advanced
Computer Science (LIACS)
Leiden, The Netherlands

Abstract

As models grow more complex and societal demands for transparency increase with emerging regulations, explainability has become an even more important research area. However, despite its recognized relevance, explainability research in IR has seen slower progress than in related fields. This full day workshop aims to advance research in explainable information retrieval by providing a more in-depth platform to reflect on recent developments and facilitate discussions to address new and persistent challenges. Our goal is to bring together a diverse group of researchers to build a shared understanding of key tasks and challenges that will lay the foundation for the future of explainable IR research.

CCS Concepts

• **Information systems** → **Information retrieval**.

Keywords

Explainable Search, Explanation Evaluation, Explainability, Interpretability, Transparency, Fairness

ACM Reference Format:

Maria Heuss, Catherine Chen, Avishek Anand, Carsten Eickhoff, and Suzan Verberne. 2025. Workshop on Explainability in Information Retrieval . 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, New York, NY, USA, 4 pages. https://doi.org/10.1145/3726302.3730361

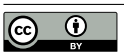
1 Motivation

Information retrieval (IR) systems play a crucial role in our daily lives, from search engines that address our information needs to recommender systems offering personalized content and product suggestions. These systems are becoming increasingly integrated into society. As they become increasingly larger and more complex (e.g., relying on sophisticated Transformer architectures [6, 10] and large amounts of training data), gaining deeper insights into their inner workings is essential to understanding their decision-making and ensuring effective human oversight.

Explainable AI research aims to explain the decision-making process of models in human-understandable terms. Explainability[1] enhances system performance and increases personalization, and is critical for ensuring safety in high-stakes domains such as finance, healthcare, and law. Additionally, it supports compliance with emerging AI regulations, such as the EU AI Act [2] and the Colorado AI Act [1], that stipulate a "right to explanation" for AI systems deployed in high-stakes situations.

Related communities, including Natural Language Processing, Computer Vision, and Recommender Systems, have seen significant progress in explainability research, supported by dedicated workshops at top conferences. Examples include, *XAI in Action* at NeurIPS 2023 [11], the *Mechanistic Interpretability Workshop* at ICML 2024 [3], and *BlackboxNLP* [4], which has been running at ACL/EMNLP for the past seven years and has had hundreds of attendees (up to 500 in 2023) and dozens of submissions. The collective efforts in these communities have contributed to faster progress in the field, with researchers acknowledging that explainability research has played a key role in accelerating advancements, as highlighted in a recent survey on the impact of interpretability and analysis research in NLP [7]. For IR, explainability has been listed as a relevant area under the SIGIR main conference call for papers since 2022, under the broader FATE Track (Fairness, Accountability, Transparency, Ethics, and Explainability). Despite IR being the backbone of AI systems in many high-stakes scenarios, we see slow progress in explainability research within IR, with a majority of work in the field focusing on recommender systems.

Additionally, the recent shift toward more complex, highly parametric models (e.g. Transformer-based models) has made it more challenging to interpret and explain retrieval results, in contrast to more traditional methods (e.g., boolean search or BM25) that were more inherently explainable. Along with the emerging regulations, calling for explainability in high-stakes decision-making, there is a need for focused attention on explainability in IR systems.



SIGIR '25, Padua, Italy

ACM ISBN 979-8-4007-1592-1/2025/07
https://doi.org/10.1145/3726302.3730361

[1] *Interpretability* is a closely related term, similarly used to describe efforts to provide insight into a model's internal processes. For simplicity, we do not distinguish between the two and use *explainability* throughout this paper, encompassing any definitions of *interpretability* under its scope.

The first *Workshop on Explainability in Information Retrieval (WExIR)* at SIGIR is dedicated entirely to explainability in IR. We aim to provide more opportunities to advance explainability research while also addressing key challenges, such as evaluation and real-world requirements for explainability, that are not often the primary focus in the main conference program. This *full day* workshop aims to address these gaps by fostering discussions on these key challenges and providing a platform for researchers to discuss a vision for the future of the field.

2 Theme and Purpose of the Workshop

WExIR aims to address challenges of the field by providing a platform with specific interaction and engagement activities beyond what is possible in the dedicated main conference session. Our overall objective is to establish a shared understanding of explainability concepts, methods, and challenges specific to IR, and to outline a full research agenda to help guide future research.

Thus, we have three primary goals:

(1) **Community Building:** Bring together researchers interested in explainable IR (XIR) to foster discussions and identify key challenges in the field. This includes both established experts in the field as well as new researchers who are interested in working on XIR in the future.
(2) **Showcase Recent Progress:** Provide a platform for sharing and discussing XIR research, including work in progress and recently published studies, to increase exposure and encourage discussion on previously faced challenges in the field.
(3) **Future of XIR:** Consolidate notions and definitions, establish best practices, and identify challenges of explainability to outline a shared trajectory for future research.

For a full list of topics of interest see the Call for Papers, Section 3.2.

3 Workshop Format

This **full-day** workshop will have two keynote speakers and short presentations of recently published relevant work or work in progress. To encourage active discussions, idea exchange, and collaboration among participants, we will have multiple interactive activities throughout the course of the day. One will be the traditional poster session in conjunction with the broader SIGIR workshop poster sessions preceded by lightning talks introducing the respective posters, allowing authors to share their work, receive feedback, and engage in discussion with participants from other workshops. The two other activities targeting participant engagement will include discussion panels that aim to highlight the particular challenges that come with the practical application of explainability methods for real-world applications and the evaluation of explanations. Furthermore, we plan to work in breakout discussion groups with the goal of taking a first step towards creating a research agenda for these topics and strengthening the community in their shared understanding of those challenges. More details about these activities are provided in the remainder of this section.

3.1 Planned Activities

Keynote. As high-profile invited speakers are expected to attract additional participants to the workshop beyond the presenters, we will have a keynote speaker whose work could be of interest to a larger audience: *Oana Inel* (confirmed). Oana is a postdoctoral researcher at ETH Zurich working on human computation, HCI, and the responsible use of data, with a particular interest in the human-centric side of explanations and evaluation through user studies.

Author Lightning Talks and Posters. Instead of a traditional paper presentation session, we will invite authors with submitted and accepted work to our workshop to present a 2-3 minute *lightning talk* to give an overview of their submitted work. This will provide authors an opportunity to promote their posters and allow other participants to more effectively decide which posters they may want to visit during the shared workshop poster session.

Discussion Panels. We will have two discussion panels aimed at addressing key challenges in explainability for IR, to foster dialogue and idea exchange within the community. The first panel focuses on the *Needs and Challenges of Stakeholders and Practitioners in XIR*. This panel addresses one of the workshop's primary motivations: in domains where explanations of IR systems and outcomes influence consequential decisions (e.g., legal, finance, health), how can we bridge the gap between developers of explainable systems and the practitioners who use them? To promote conversation on this topic, we invite professionals from these domains to share their insights on what is needed for real-world adoption of explainable IR systems.

The second panel centers on the *Evaluation of Explanations*. In a field that centers on rigorous evaluation, it is currently unclear how explainable IR systems should be evaluated. To address this challenge, we invite speakers who work on explainable IR or specialize in evaluation to provide their insights on how the field of XIR can develop rigorous methods and frameworks for evaluating the quality and effectiveness of explainable IR systems.

Round Table Discussions. We will break up into groups of 5-10 people to discuss some of the topics of interest outlined in the Call for Papers (Section 3.2), with a particular focus on topics discussed during the discussion panels. The goal of this session will be to collect input from the community, bring attention to some of the challenges that the field faces and connect researchers who are interested in taking steps toward solving these challenges. During the discussions, the organizers will take notes that will be used to write a workshop report. Participants will also be invited after the conference to continue the discussion and contribute to future writing if interested.

Tentative Schedule of Events The tentative schedule is shown in Table 1.

3.2 Call for Papers

To foster the growth of an active community, we aim to involve both senior and junior researchers in our workshop. To curate an insightful and engaging program, we will invite submissions of **extended abstracts** up to 2 pages in length for research across various stages: already published work, ongoing work not yet submitted elsewhere, and finished work that has not yet been published. The workshop will also allow authors of in-progress work to receive constructive feedback from the community to help refine their papers. We will actively encourage people (in particular students) to submit ongoing work and preliminary results and reach out to authors

Table 1: Tentative workshop schedule

09:00-09:45	Opening Remarks & Keynote 1
09:45-10:30	Discussion Panel: *Needs and Challenges of Stakeholders/Practitioners in XIR*
10:30-11:00	Coffee Break
11:00-11:45	Round Table Discussions
11:45-12:30	Author Lightning Talks
12:30-13:30	Lunch Break
13:30-14:15	Workshop Posters
14:15-15:00	Keynote 2 (TBD)
15:00-15:30	Coffee Break
15:30-16:15	Discussion Panel: *Evaluation of Explanations*
16:15-17:00	Round Table Discussions
17:00-17:05	Closing Remarks

of previously published relevant work to encourage submissions. Below is a list of potential topics for submissions.

- **Novel Methods and Models for Explainability:** Development of new algorithms, techniques, or models to explain IR systems.
- **Internal Analysis and Model Interpretability:** Novel techniques for analyzing the inner workings of IR systems, insights learned from analysis and interpretation.
- **Evaluation:** Evaluation of explanation quality, metrics for explainable systems.
- **Human-Centered Explainability:** Work focused on identifying, investigating, or integrating users' explainability needs in IR systems.
- **Practical Applications of Explainability:** Explainability in high-stakes domains (e.g., healthcare, legal, finance, etc.), applications to personalized search/recommendation, bias mitigation and fairness.
- **Other:** Other topics not mentioned in this list but still pertaining to explainability in IR.

In addition to "traditional" topics in explainability, we will also open extended abstract submissions for *Perspectives on the Challenges of Explainability in IR*, which include but are not limited to the following topics:

- **Explainability in the Era of LLMs:** What challenges for explainability arise with the emergence of LLMs in the IR pipeline and how should we begin to solve them?
- **Defining and Evaluating Explainability:** How can we unify different notions and definitions of explainability and achieve consensus over what constitutes a "good" explanation?
- **Insights from Other Fields:** What takeaways or solutions from fields outside of IR can be used to advance XIR research?
- **Interdisciplinary XIR:** How do ethical concerns, regulatory requirements (e.g., GDPR), and domain-specific constraints (e.g., health, legal, finance, etc.) affect explainability research?

3.3 Special Requirements and Contingency Plans

AV equipment for speakers and panel discussions will be important for the workshop but as this equipment is typically already provided as part of the main conference, we do not expect this to be an issue.

For the round table discussion session, it would be nice to have tables to create actual "round table" discussions. However, this is not a must-have and if tables are not available, we can easily rearrange chairs to create spaces that still foster engaging discussions.

In the event of a small number of submissions to our workshop, we will invite authors of the accepted papers from the main SIGIR conference for lightning talks to provide them with more exposure and enrich the discussion and reallocate the extra time from the Author Lightning Talks to the Discussion Panel Audience Q&A and/or the Round Table Discussion sessions.

To minimize the risk of last-minute cancellations, we will prioritize inviting speakers and panelists who are already likely to attend the conference due to their work in IR or related areas (i.e., RecSys). Given that SIGIR attracts a diverse group of attendees from academia and industry, we believe that this approach will still provide a wide range of perspectives and insights for our workshop. Each panel will feature 3–4 speakers, which may include keynote or invited paper presenters. This will allow us to maintain a robust lineup even if a speaker cancels at the last minute.

3.4 Envisioned Workshop Output

Following the workshop, we plan to submit a SIGIR Forum report on the workshop itself, as well as insights gained and possible next steps for future work. Ideally, a first set of requirements and desirables for explainability in IR can be established. In the long term, we would like to use insights from this workshop to write a perspective paper, for example for IRRJ[2] and aim to establish a lasting workshop series by identifying a dedicated core group of researchers in this inaugural edition to help drive future efforts.

4 Workshop Organizers

All organizers are expected to attend onsite.

Maria Heuss is a final-year PhD Candidate at the IRLab, University of Amsterdam, currently supervised by Maarten de Rijke and Avishek Anand. Her research focuses on questions around the fairness and interpretability of IR and NLP systems, with particular interest in rigorous formalizations and evaluation. She has published at leading conferences on information retrieval such as SIGIR and CIKM.

Catherine Chen is a fourth-year PhD Candidate at Brown University, currently supervised by Carsten Eickhoff. Her research centers on explainable information retrieval, particularly developing methods for explainable search. She is also broadly interested in the explainability of large language models and the applications of ML/NLP/IR to the biomedical domain and has published papers in leading IR and NLP venues, such as SIGIR and EMNLP.

Avishek Anand is an associate professor at the Delft University of Technology. His research focuses on explainable IR and other

[2]https://irrj.org/

topics around Question Answering and information retrieval. He has published, organized, and given tutorials in leading IR venues.

Carsten Eickhoff is a professor at the University of Tübingen and holds a visiting appointment at Brown University. His research focuses on transparency and explainability of language models. His lab pursues these topics both in the context of information retrieval as well as natural language processing, and frequently explores applications in biomedicine.

Suzan Verberne is a professor of Natural Language Processing at Leiden University. Her recent work centers around interactive information access for low-resource contexts. She has a strong interest in the interplay between search engines and LLMs and the role of the user in both. Suzan is highly active in the NLP and IR communities, holding chairing positions in the large world-wide conferences.

5 PC Members and Selection Process

The workshop is open for participation for everyone interested. Selection of extended abstracts (2 pages ACM format) will be mainly based on relevance to the workshop. We invite submissions from different stages of research, ranging from perspective papers to work in progress to previously published work, and welcome submissions from students. Reviewing will be done by the organizers themselves as well as a number of PC members.

6 Target Audience

As previously mentioned, the target audience for our workshop is both junior and senior researchers who currently work on XIR or are interested in working on XIR in the future. Due to the highly interdisciplinary nature of explainability, we will encourage participation not only from researchers in academia, but from stakeholders and practitioners who have a vested interest in deploying explainable systems for real-world applications. Thus, SIGIR is an appropriate venue to attract such an audience, as the conference attracts a diverse group of academics, as well as a wide variety of industry and government specialists who may rely on explanations of IR systems to inform high-stakes decisions and outcomes.

7 Related Workshops

The *International Workshop on Explainable Recommendation and Search (EARS)*[3] was held at SIGIR from 2018-2020 [12] with the main goal of presenting the latest research and connecting researchers in the community. Topics covered by the series include models for explainable search and recommendation, information sources for explanations, user behavior analysis, new explanation types, applications, and evaluation methods, and the series resulted in a SIGIR forum report [13].

Since then, explainability research has seen more engagement in the Recommender Systems field, with the workshop on *Measuring the Quality of Explanations in Recommender Systems (QUARE)* at SIGIR'22 [8] and RecSys'23 [5]. The first QUARE workshop included 28 papers (original submissions, demos, and position papers) and was also followed up by a SIGIR forum report [9].

Our workshop builds on the foundation of the topics covered by EARS, investigating the progress in recent years while introducing new focuses: (1) tackling open challenges such as consistent definitions and evaluation of explanations, (2) integrating insights from related fields such as NLP and general ML, and (3) focusing on different IR use cases and their needs for explanations. While progress in explainability for recommender systems has gained traction, explainability in IR systems has not received the same focused attention. We hope this workshop will help the community identify a shared trajectory for future exploration and advance research in this area.

[3] https://ears2019.github.io/

References

[1] 2024. Colorado Senate Bill 24-205: Concerning the Regulation of Artificial Intelligence. Colorado General Assembly. https://leg.colorado.gov/bills/sb24-205 Will go into effect on February 1, 2026.

[2] 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

[3] 2024. Workshop on Mechanistic Interpretability. Workshop at the International Conference on Machine Learning (ICML). https://icml.cc/virtual/2024/workshop/29953

[4] Yonatan Belinkov, Najoung Kim, Jaap Jumelet, Hosein Mohebbi, Aaron Mueller, and Hanjie Chen. 2024. Proceedings of the 7th BlackboxNLP Workshop: Analyzing and Interpreting Neural Networks for NLP. In *Proceedings of the 7th BlackboxNLP Workshop: Analyzing and Interpreting Neural Networks for NLP*.

[5] Oana Inel, Nicolas Mattis, Milda Norkute, Alessandro Piscopo, Timothée Schmude, Sanne Vrijenhoek, and Krisztian Balog. 2023. QUARE: 2nd Workshop on Measuring the Quality of Explanations in Recommender Systems. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 1241–1243.

[6] Jimmy Lin, Rodrigo Nogueira, and Andrew Yates. 2022. *Pretrained transformers for text ranking: Bert and beyond*. Springer Nature.

[7] Marius Mosbach, Vagrant Gautam, Tomás Vergara-Browne, Dietrich Klakow, and Mor Geva. 2024. From Insights to Actions: The Impact of Interpretability and Analysis Research on NLP. *arXiv preprint arXiv:2406.12618* (2024).

[8] Alessandro Piscopo, Oana Inel, Sanne Vrijenhoek, Martijn Millecamp, and Krisztian Balog. 2022. QUARE: 1st Workshop on Measuring the Quality of Explanations in Recommender Systems. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 3478–3481.

[9] Alessandro Piscopo, Oana Inel, Sanne Vrijenhoek, Martijn Millecamp, and Krisztian Balog. 2023. Report on the 1st Workshop on Measuring the Quality of Explanations in Recommender Systems (QUARE 2022) at SIGIR 2022. In *ACM SIGIR Forum*, Vol. 56. ACM New York, NY, USA, 1–16.

[10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, California, USA) *(NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 6000–6010.

[11] FChhavi Yadav, Michal Moshkovitz, Nave Frost, Suraj Srinivas, Bingqing Chen, Valentyn Boreiko, Himabindu Lakkaraju, J. Zico Kolter, Dotan Di Castro, and Kamalika Chaudhuri. 2023. XAI in Action: Past, Present, and Future Applications. Workshop at Advances in Neural Information Processing Systems (NeurIPS). https://neurips.cc/virtual/2023/workshop/66529

[12] Yongfeng Zhang, Xu Chen, Yi Zhang, Min Zhang, and Chirag Shah. 2020. EARS 2020: The 3rd International Workshop on ExplainAble Recommendation and Search. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2472–2474.

[13] Yongfeng Zhang, Yi Zhang, and Min Zhang. 2019. Report on EARS'18: 1st International Workshop on ExplainAble Recommendation and Search. In *ACM SIGIR Forum*, Vol. 52. ACM New York, NY, USA, 125–131.