Simulation of Interactive Information Retrieval: A Guided Tour

Saber Zerhoudi
University of Passau
Passau, Germany
szerhoudi@acm.org

Adam Roegiest
Zuva
Toronto, Canada
adam@roegiest.com

Johanne R. Trippas
RMIT University
Melbourne, Australia
j.trippas@rmit.edu.au

Abstract

Interactive information retrieval (IIR) systems, including search engines and conversational systems, are increasingly central to user experiences. However, rigorously evaluating their performance, particularly as interactions become highly personalized, remains a scientific challenge. While user simulation offers a powerful methodology for reproducible evaluation, its adoption is hindered by a steep learning curve and a fragmented landscape of complex tools. This half-day tutorial provides a practical, hands-on introduction to user simulation at varying levels of complexity, from foundational statistical models to advanced, LLM-driven frameworks. Through a series of guided problems, participants will acquire practical skills in using popular libraries, learning user models from data, and applying large language models (LLMs) to simulate user behavior. The tutorial concludes with evaluating the simulators themselves, providing participants with guidance on appropriate use cases and fidelity assessment.

CCS Concepts

• **Information systems → Users and interactive retrieval**; **Language models**.

Keywords

Interactive Information Retrieval, Large Language Models, Simulation

1 Introduction

Interactive information retrieval (IIR) systems are defined by the process of a user interacting with a system to satisfy an information need [3]. Appropriately evaluating these systems remains a core, open challenge for the research community. Traditional evaluation methodologies, such as controlled user studies, observational studies, and online A/B testing, face several limitations [5]. User studies are costly, difficult to scale, and suffer from an inherent lack of exact replication and consistent reproducibility [9], making it difficult to compare multiple interactive systems. These issues are compounded when researchers seek to investigate specific, hard-to-access user populations.

To address these challenges, user simulation has emerged as a promising methodology [6]. Simulation allows researchers to model system performance with user interaction at a low cost and, crucially, in a repeatable and reproducible manner [2]. This enables the comparison of interactive systems and provides a controlled environment for analyzing system performance under different user behaviors. In addition, with the recent rise of large language models (LLMs) and generative models has introduced a new, powerful, but complex class of simulation techniques [4, 8, 10, 12]. Where researchers previously relied on statistical models to generate query formulations, LLMs can now be prompted to produce nuanced interactions, complete with linguistic variations, that reflect specific user traits and knowledge states.

This new potential also creates a major barrier to entry [7]. As simulation frameworks incorporate generative models, their operational complexity increases. This leaves a gap between the researchers who build these complex systems and the broader community, especially researchers with qualitative or user-focused backgrounds, who stand to benefit most from using them. The field lacks a practical, hands-on synthesis that connects foundational theory to modern application [1].

This tutorial is designed to fill that gap. It provides a systematic, hands-on walkthrough of simulation techniques, organized by increasing complexity. It will equip participants with the practical skills to select, implement, and, most importantly, validate simulation methods for their own research. The tutorial's primary objective is to provide participants with a solid foundation and practical skills for incorporating user simulation into their research. It combines theoretical underpinnings with hands-on exercises to build intuition on how simulation works.

Following this tutorial, participants will be able to:

- **Understand Foundational Principles:** Articulate the core concepts of user simulation and its critical limitations, enabling them to identify appropriate (and inappropriate) use cases for their work.
- **Implement Foundational Simulations:** Write code to implement simple simulation models to perform initial idea validation and "sanity checking" of interactive systems.
- **Assess Simulation Complexity:** Differentiate between foundational and advanced simulation approaches and determine when a complex, high-fidelity framework is necessary for a given research question.
- **Apply Modern Simulation Toolkits:** Gain hands-on skills with common, open-source simulation frameworks (e.g., SimIIR [1], UXSim [11]) to construct and execute complex interaction experiments.

- **Use LLMs for User Simulation:** Apply various prompting mechanisms to instruct LLMs to simulate different user actions and behaviors, reflecting specific personas or information-seeking strategies.
- **Evaluate Simulator Fidelity:** Implement simulation evaluation protocols to measure how well a simulated population matches a desired real-user population and diagnose common model failures.

2 Syllabus

This half-day tutorial is organized into three interconnected modules. The content is structured to guide participants from foundational concepts to advanced applications, with a mix of presentations and guided, hands-on coding problems.

- **Part 1: Foundational Simulation (60 min).** This module provides a foundational understanding of how to conduct user simulation. We begin by positioning simulation as a practical tool for "sanity checking" system behavior, placing particular focus on its use for initial validation rather than in-depth fidelity.
  - *Lecture:* The "Why" of simulation; simple models of interaction (e.g., click models, Markov chains); the concept of simulation for "sanity checking" ideas.
  - *Hands-on Problem:* Participants will implement a basic simulator from scratch in a notebook to test a simple interactive search feature.
- **Part 2: Advanced Simulation Frameworks (75 min).** This module provides a guided walkthrough of using learned simulation parameters and established frameworks. We will discuss how to learn user models from data (e.g., real user logs) and incorporate them into an experiment. A significant portion is dedicated to the modern use of LLMs.
  - *Lecture:* Architectures of modern toolkits; incorporating user data; using LLMs as simulators.
  - *Hands-on Problem:*
  (1) **Framework-based Simulation:** Participants will use the SimIIR framework [1] to configure and run a high-fidelity simulation experiment, customizing it with different user types.
  (2) **LLM-based Simulation:** Participants will work with UXSim [11] to prompt an LLM to generate query reformulations and assess document relevance based on a given user persona.
- **Part 3: Evaluating Simulations (45 min).** This final module addresses the question of how to evaluate the evaluators. Participants will learn how to measure simulation fidelity and understand the limitations of their models.
  - *Lecture:* The importance of validation; metrics for comparing simulated behavior to real user logs; qualitative vs. quantitative assessment. In addition, we will discuss potential biases inherent in user simulation, including the risk of reinforcing demographic stereotypes through generative model-based personas, the lack of diversity in synthetic populations, and the transparency requirements for reporting simulated results.

  - *Hands-on Problem:* Participants will be given a sample of "real" user data and the output from their Part 2 simulator, and will implement simple metrics to measure the alignment between them.
  - *Wrap-Up and Q&A:* Summary of techniques and discussion of which level of simulation is appropriate for different research goals.

3 Target Audience and Prerequisites

This introductory tutorial targets members of the community interested in incorporating simulation into their IIR research. It is designed for graduate students, academic researchers, and industry practitioners at all career stages. We explicitly welcome and encourage participation from researchers with qualitative and user-focused backgrounds who wish to learn how to add quantitative, reproducible evaluation methods to their research.

Prerequisites: No special experience is required. A basic familiarity with programming concepts, Python, and Jupyter notebooks will be advantageous. All code and notebooks will be provided. Assistance will be provided for all hands-on components.

4 Presenters

Dr Saber Zerhoudi is a Postdoctoral Researcher at the University of Passau, Germany. His research is centered on interactive information retrieval, with a particular focus on modeling and simulating user search behaviors. His recent work involves developing user-centric agents for retrieval-augmented generation (RAG) frameworks and investigating novel methodologies for next-generation search. He is an active member of the IR community, having co-organized tutorials and workshops on user simulation and open web search at SIGIR'25 and ECIR'24-25.

Dr Adam Roegiest is the VP of Research and Technology at Zuva, a Toronto-based legal AI startup. Adam's research has focused on the application of information retrieval and machine learning technologies to legal retrieval tasks. More recently, he has extended his research into how these technologies interact with human-computer interaction and accessibility. Adam previously organised both iterations of the TREC Total Recall track, one iteration of the TREC Real-Time Summarization track, workshops at CHIIR 2024 and ECIR 2025 focusing on the future of IR, and a tutorial at CHIIR 2025. Adam is also a steering committee member for CHIIR.

Dr Johanne Trippas is a Vice-Chancellor's Senior Research Fellow at RMIT University, specializing in intelligent systems, focusing on digital assistants and conversational information seeking. Their research aims to enhance information accessibility through conversational systems, interactive information retrieval, and human-computer interaction. Additionally, Johanne is part of the NIST TREC program committee and is an ACM CHIIR steering committee member. They serve as vice-chair of the SIGIR Artifact Evaluation Committee, workshop chair for ACM CHIIR'25, and program chair for ACM CHIIR'26. Johanne has organized the ACM Conversational User Interfaces (CUI'24) conference, workshops (CHIIR'20-22, '24, ECIR'24–26), a TREC Track (CAsT'22), and tutorials (CHIIR'21, '25, SIGIR'22, WebConf'23, and ECIR'24).

References

[1] Leif Azzopardi, Timo Breuer, Björn Engelmann, Christin Kreutz, Sean MacAvaney, David Maxwell, Andrew Parry, Adam Roegiest, Xi Wang, and Saber Zerhoudi. 2024. SimIIR 3: A Framework for the Simulation of Interactive and Conversational Information Retrieval. In *Proceedings of the 2024 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region, SIGIR-AP 2024*. ACM, 197–202. doi:10.1145/3673791.3698427

[2] Krisztian Balog, Nolwenn Bernard, Saber Zerhoudi, and ChengXiang Zhai. 2025. Theory and Toolkits for User Simulation in the Era of Generative AI: User Modeling, Synthetic Data Generation, and System Evaluation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2025*. 4138–4141. doi:10.1145/3726302.3731697

[3] Peter Ingwersen and Kalervo Järvelin. 2005. *The Turn - Integration of Information Seeking and Retrieval in Context*. The Kluwer International Series on Information Retrieval, Vol. 18. Kluwer. doi:10.1007/1-4020-3851-8

[4] Vahid Sadiri Javadi, Fryderyk Róg, Aska Aksa, Johanne R. Trippas, Svitlana Vakulenko, and Lucie Flek. 2026. CHARISMA: Character-Based Interaction Simulation with Multi-LLM Agents Toward Computational Social Psychology. In *Proceedings of the ACM Conference on Information Interaction and Retrieval (CHIIR'26)*. 1–5.

[5] Diane Kelly. 2009. Methods for Evaluating Interactive Information Retrieval Systems with Users. *Found. Trends Inf. Retr.* 3, 1-2 (2009), 1–224. doi:10.1561/1500000012

[6] David Maxwell and Leif Azzopardi. 2016. Simulating Interactive Information Retrieval: SimIIR: A Framework for the Simulation of Interaction. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016*. 1141–1144. doi:10.1145/2911451.2911469

[7] Johanne Trippas, Oleg Zendel, and Adam Roegiest. 2025. Applying Large Language Models to Interactive Information Retrieval: A Practical Exploration. In *Proceedings of the 2025 Conference on Human Information Interaction and Retrieval*. ACM.

[8] Johanne R. Trippas and J. Shane Culpepper. 2025. Report from the 4th Strategic Workshop on Information Retrieval in Lorne (SWIRL 2025). *SIGIR Forum* 59, 1 (2025), 1–68.

[9] Peilin Yang and Hui Fang. 2016. A Reproducibility Study of Information Retrieval Models *(ICTIR '16)*. Association for Computing Machinery, New York, NY, USA, 77–86. doi:10.1145/2970398.2970415

[10] Saber Zerhoudi and Michael Granitzer. 2024. Generative Agents Navigating Digital Libraries. In *Sustainability and Empowerment in the Context of Digital Libraries - 26th International Conference on Asia-Pacific Digital Libraries, ICADL 2024, Bandar Sunway, Malaysia, December 4-6, 2024, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 15493)*. Springer, 171–188. doi:10.1007/978-981-96-0865-2_14

[11] Saber Zerhoudi and Michael Granitzer. 2025. UXSim: Towards a Hybrid User Search Simulation. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM '25)*. ACM. doi:10.1145/3746252.3761640

[12] Erhan Zhang, Xingzhu Wang, Peiyuan Gong, Yankai Lin, and Jiaxin Mao. 2024. USimAgent: Large Language Models for Simulating Search Users. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024*. ACM, 2687–2692. doi:10.1145/3626772.3657963