T2: Cognitive Effects and Biases in Large Language Models

Markus Schedl, Ralph Hertwig
Antonela Tommasel, Shahed Masoudian
https://cpjku.github.io/cobis_llm_eacl2026/
Cutting-edge
Pavillon DE RABAT (Level 1)
Saturday, March 28, 2026 - from 14:00 to 17:30

Cognitive effects such as anchoring, positional effect, or confirmation bias are core aspects of human decision making and reasoning. As LLMs increasingly act as communicative partners, reasoning tools, and evaluators, understanding how these cognitive effects influence their behavior and vice versa has become essential. While recent studies have adapted psychological experiments to detect cognitive biases in LLMs, they often use a particular kind of experimental setup from psychology that carries implications for human performance. In addition, current NLP studies often confuse cognitive effects with biases, diverging from their psychological foundations and overlooking potentially functional aspects of these phenomena.

In this tutorial, jointly organized by NLP researchers and a cognitive psychologist and decision scientist, we aim to build shared conceptual and methodological ground between the two disciplines. We begin by outlining how cognitive effects and biases are defined, validated, and sometimes debated within psychology, highlighting differences and contradictions in experimental designs. We then bridge these insights to NLP through an overview of key studies examining cognitive biases in LLMs, mapping their methodological parallels and divergences. The tutorial also includes a hands-on component where participants explore the challenges of detecting a single cognitive bias (e.g., positional bias) in multilingual LLMs, illustrating the nuances and pitfalls of such evaluations. We conclude by discussing emerging research directions and open questions at the intersection of cognitive science and large language models.

Markus Schedl, Johannes Kepler University Linz (JKU), Germany
email: markus.schedl@jku.at
website: http://www.mschedl.eu

Bio. Markus Schedl is a Full Professor at the Johannes Kepler University Linz (JKU), affiliated with the Institute of Computational Perception, leading the Multimedia Mining and Search Group and the Human-centered AI Group at the Linz Institute of Technology (LIT) AI Lab. His research interests include information retrieval, recommender systems, natural language processing, and trustworthy AI, with a keen interest in psychological aspects of AI-powered systems and their users. He (co-)authored more than 300 refereed conference papers, journal articles, books, and book chapters, and has given numerous tutorials at top information retrieval and recommender systems venues.

Ralph Hertwig, Center of Adaptive Rationality, Max Planck Institute for Human Development, Germany
email: `hertwig@mpib-berlin.mpg.de`
website: `https://www.mpib-berlin.mpg.de/staff/ralph-hertwig`
Bio. Ralph Hertwig is the Director of the Center of Adaptive Rationality (ARC) at the Max Planck Institute for Human Development in Berlin. He is also an Honorary Professor at the Free University Berlin and the Humboldt University Berlin. His research focuses on how people make decisions and how good decisions can be fostered. He is a member of the German Academy of Sciences Leopoldina, the Berlin-Brandenburg Academy of Sciences and Humanities, and the acatech, the National Academy of Science and Engineering. He has received the Leibniz Prize, Germany's most prestigious research award.

Antonela Tommasel, Johannes Kepler University Linz, Germany / UNCPBA, Argentina
email: `antonela.tommasel@jku.at`
website: `https://tommantonela.github.io`
Bio. Antonela Tommasel is an Assistant Professor at the Johannes Kepler University Linz (JKU) and at the National University of the Center of the Province of Buenos Aires (UNCPBA, Argentina), as well as a researcher at the National Scientific and Technical Research Council (CONICET, Argentina). Her research lies at the intersection of artificial intelligence and social computing, with a focus on recommender systems. Her work explores fairness and bias in LLM-based recommendations, agentic LLM systems for supporting complex decision-making in software engineering, and the dynamics of user engagement and misinformation in online platforms.

Shahed Masoudian, Johannes Kepler University Linz (JKU), Germany
email: `shahed.masoudian@jku.at`
website: `https://hcai.at/persons/masoudian`

Bio. Shahed Masoudian is a University Assistant and PhD student at the Johannes Kepler University Linz (JKU). A core part of his work focuses on the controllability of language models, treating user control over model behavior as a practical and promising solution to addressing biases within language models. He develops novel methods that allow bias mitigation to be dynamic, scalable, and adaptable to different contexts, without requiring extensive retraining. He also supervises students on various topics in information retrieval, recommender systems, controllable text generation, and fairness of language models.