



Topic

Innovating AI Evaluation: Beyond Accuracy and Precision.

Description

As the landscape of artificial intelligence continues to evolve, the need for comprehensive and nuanced evaluation methods increases as well. Traditional metrics such as accuracy and precision, while important, are insufficient for fully capturing the complexities and impacts of AI systems.

The SAIL Spring School aims to address this gap by introducing participants to a diverse array of evaluation strategies, such as user evaluations, ethical and societal impacts, evaluating outcomes that are co-constructed between user and AI, mathematical guarantees, interpretability and transparency assessments, context-specific metrics, etc.

More information and registration under: www.sail.nrw/springschool/

When & Where

When: March 26-28, 2025

Where: CITEC lecture hall, Bielefeld University, Germany

AI Evaluation for Complex Knowledge Domains



**Bettina
Finzel**

*University of
Bamberg*

Complex knowledge domains are characterized by manifold relatedness of concepts. These concepts and relations stem from (possibly structured) human knowledge. In contrast to simple knowledge, complex knowledge is based on expressive representations and usually combines multiple modalities to describe entities from the domain of interest. AI, especially machine learned models, are nowadays broadly applied to all kinds of data sets that originate from complex knowledge domains, like medicine, biology and the natural sciences.

However, the fact that these knowledge domains use various representations and modalities to describe concepts and their relations, is often neglected in AI evaluation. This tutorial introduces novel and well-established evaluation methods from the fields of trustworthy artificial intelligence, multimodal machine learning and neuro-symbolic artificial intelligence. Participants will learn about open challenges in AI evaluation for complex knowledge domains. They will further gain insights into possible solutions that may foster better understanding of an AI system's (true?) performance.

Funded by

