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 insu cient 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

Evaluation (Metrics) for Generative AI

Steffen Eger
University of Technology Nuremberg

Christoph Leiter
University of Technology Nuremberg

In this tutorial, we will discuss evaluation metrics for generative AI, especially for machine translation, summarization, and text-to-image generation. Our discussion will first give a historical overview over metrics from BLEU and ROUGE in the early 2000s to BERTScore and MoverScore in the 2020s to the latest generation of prompt-based evaluation schemes with Large Language Models such as Gemba-MQM and VQAScore. We will especially focus on explainability, efficiency, and robustness of recent high-quality metrics. We will also discuss human evaluation and particular use cases such as human and automatic evaluation of automatic literary translation.