SAIL
SUSTAINABLE LIFE-CYCLE OF INTELLIGENT SOCIO-TECHNICAL SYSTEMS

Spring School 2025

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

Security of Machine Learning Systems

Thorsten Eisenhofer
Technical University Berlin

Machine learning models have been identified to be vulnerable against a variety of attacks. Yet, much of this work has treated models in isolation, failing to account for the intricacies of real-world deployments where models function as part of a broader machine learning system. This gap in understanding is especially important as the next frontier is already unfolding, with many companies starting to incorporating large language models into their systems. In this talk, I explore my research on the security of machine learning systems. By viewing the learning algorithm as a single element within a broader system, I will highlight the expanded attack surfaces of practical deployments and introduce new perspectives on how to secure them effectively.

Funded by

Ministry of Culture and Science of the State of North Rhine-Westphalia

UNIVERSITÄT BIELEFELD
UNIVERSITÄT PADERBORN
HSBI Hochschule Bielefeld University of Applied Sciences and Arts
TH OWL

sail.nrw