Machine Learning at the Edge of Ignorance

Krikamol Muandet

(CISPA Helmholtz Center for Information Security)

15/04/2026, 4:15 pm

Department of Statistics, Ludwigstr. 33, Seminar Room 144
and online via Zoom (Link)
(Meeting-ID: 631 1190 7291; Password: StatsCol)

Machine learning has transformed computation from executing well-defined, exact, and repeatable tasks into a paradigm of learning from data and generalising to unseen scenarios. It now serves as a foundational technology driving scientific discovery, enabling new forms of creative expression, and reshaping society. However, as machine learning systems are deployed in increasingly open-ended and high-stakes environments, their fundamental limitations become apparent. Phenomena such as distribution shift, adversarial manipulation, lack of robustness, catastrophic forgetting, hallucinations, safety risks, and misalignment all point to a common challenge: operating reliably at "the edge of ignorance"—the boundary between what a model knows and what it does not.

In this talk, I will present recent work on out-of-distribution generalisation, hypothesis testing, strategic learning, AI regulation, and causal machine learning that aims to move beyond this limitation. The central theme is the development of intelligent systems that can explicitly represent, reason about, and communicate the limits of their own knowledge, transforming epistemic boundaries from hidden vulnerabilities into principled objects of analysis and decision-making.

About the Speaker:
Krikamol Muandet is a chief scientist and tenure-track faculty at CISPA Helmholtz Center for Information Security, where he leads the Rational Intelligence (RI) Lab. From 2018 to 2022, he was a research group leader affiliated with the Empirical Inference Department at Max Planck Institute for Intelligent Systems, Tübingen, and from January 2016 to December 2017, he was a lecturer at the Department of Mathematics, Faculty of Science, Mahidol University in Thailand. Krikamol graduated summa cum laude with a PhD degree specializing in kernel methods in machine learning under supervision of Prof. Bernhard Schölkopf. His research interest lie in the area of machine learning. He is particularly interested in, for example, statistical learning theory, kernel methods, causal inference, counterfactual prediction, and algorithmic game theory.