Contribution ID: **18** Type: **not specified**

Explainability of Neural Networks in STEM fields

Friday, 20 February 2026 09:00 (2 hours)

Due to their predictive accuracy, machine learning models are increasingly employed in high-stakes decision making and in scientific contexts. These predictive capabilities often come at the price of interpretability: The models are too complex to be inherently intelligible by their human users.
In an attempt to restore this interpretability, a broad range of methods have been developed, often subsumed under the umbrella of the term Explainable AI (XAI). In this session, we will give the attendees an introduction to the field. We will start by mapping the different approaches using the standard taxonomies, before we formally introduce some of the most popular explanation methods.
Over the course of the session we will focus on the core challenges that practitioners face: (1) Translating the outputs of XAI methods into concise statements about model and data, and (2) choosing the explanation technique that is most suitable for a given task. We will tackle these questions with a particular focus on the task of learning about feature target associations, thereby preparing attendees to employ XAI methods in their own work.

Presenter: KÖNIG, Gunnar