HELMHOLTZ
IMAGING

Contribution ID: **77**   Type: **Talk**

Interpretable Vision and Language Models

Tuesday 14 May 2024 11:15 (45 minutes)

Clearly explaining a rationale for a visual classification decision to an end-user can be as important as the decision itself. For the communication to be effective, the decision maker needs to recognize the class-discriminative properties of the object that are present in the image but also it needs to understand the intend of the communication partner. In this talk, I will present my past and current work on Explainable Machine Learning focusing on large vision and language models where we show (1) how to learn compositional representations of images that go beyond recognition towards understanding, (2) how to generate visual features using natural language descriptions when no visual data is available to train deep models, and (3) how our models focus on discriminating properties of the visible object, jointly predict a class label, explain why/not the predicted label is chosen for the image.

Primary author: AKATA, Zeynep (Helmholtz Munich)

Session Classification: Keynote by Zeynep Akata: Interpretable Vision and Language Models