Session Program

18–20 Feb 2025

Lamarr Lab Visits: 2025.1

Trustworthy AI & Human-Centered AI

Lamarr/RC Trust Dortmund, JvF25/2-232 - Meeting Room North
Lamarr Institute/TU Dortmund University Joseph-von-Fraunhofer-Straße 25 44227 Dortmund

Wednesday 19 February

10:00

Trustworthy AI & Human-Centered AI: Research Area Meeting

Session |
Location: Lamarr/RC Trust Dortmund, JvF25/3-302 - Co-Working Space, Joseph-von-Fraunhofer-Str. 25 44227 Dortmund |
Conveners: Jakob Rehof, Maram Akila, Bahavathy Kathirgamanathan, Gennady Andrienko, Natalia Andrienko

10:00–10:15 Welcome

Speaker
Jakob Rehof

10:15–10:35 The Anatomy of Evidence - An Investigation into Explainable ICD Coding

Speaker
Katharina Beckh

10:35–10:55 Directional ExplainableA AI: The Problem of the Rabbit and the Duck

Speaker
Carina Newen

10:55–11:15 Does the model think as we expect?

Speakers
Gennady Andrienko, Natalia Andrienko

11:15–11:35 Text as parameter interactive prompt optimisation for large language models

Speaker
H.S. Lin

11:35–11:45 Coffe Break

11:45–12:05 From Local to Global Explanations

Speaker
Maram Akila

12:05–12:25 Fast Linear Decomposition of ReLU Networks

Speaker
A. Baudzus

12:25–12:40 Reinforcement Learning from Self-feedback

Speaker
C. van Niekerk

12:40–12:55 Leveraging Human-Centered ML to create more Explainable ML models

Speaker
Bahavathy Kathirgamanathan

12:55–13:00 Closing

13:00