Markus Pantsar, Frederik Stjernfelt and
Käte Hamburger Kolleg Aachen:
Cultures of Resaerch (c:o/re)
RWTH Aachen University

c:o/re Workshop: Explainable AI – Explanations in AI (02./03. February 2022)

Location: RWTH, HumTec, Theaterplatz 16, R303

One important challenge in machine learning is the "black box" problem, in which an artificial intelligence reaches a result without any humans being able to explain why. This problem is typically present in deep artificial neural networks, in which the hidden layers are impenetrable. To tackle this problem, researchers have introduced the notion of *explainable AI* (XAI), artificial intelligence the results of which can be understood by humans. The XAI position is usually characterised in terms of three properties: transparency, interpretability, and explainability. While the first two have standard definitions, explainability is not understood in a uniform manner. What does explainability mean? What kind of AI is explainable? Can there be properly explainable machine learning systems? In this workshop, we discuss a variety of approaches to these topics in connection to fundamental questions in artificial intelligence. What are explanations in AI? What do AI systems explain and how? How does AI explanation relate to the topics of human understanding and intelligence?

Join workshop via Zoom:
https://rwth.zoom.us/j/95389080288?pwd=QjF5NjV2d005QlFkWXBmRzBDMmVNdz09

Meeting ID: **953 8908 0288**, Passcode: **742180**

Program:

Wed., 02.02. R303 & online

13:00–13:40 Daniel Wenz (RWTH Aachen)

"Understanding Explanatory Structures in Computer-Assisted Proofs – Explanation and Understanding together Again?"

13:40–14:20 Markus Pantsar (University of Helsinki, c:o/re Aachen)

"Theorem Proving in Deep Artificial Neural Networks"

14:20 COFFEE BREAK

14:50–15:30 Frederik Stjernfelt (Aarhus University, c:o/re Aachen)

"Peirce as a Philosopher of AI"

15:30 BREAK

17:00–19:00 c:o/re Lecture Series "Cultures of Research – Digitalization of Research"

Jobst Landgrebe (Cognotekt, Cologne)

"Certifiable 'Artificial Intelligence' for Reliable Systems"

Join lecture via Zoom:
https://rwth.zoom.us/j/95907232437

Thur., 03.01. R303 & online

10:30–11:10 Gabriele Gramelsberger (c:o/re Aachen)

"Machine Learning – An Emerging Experimental Culture in Science"

11:10–12:00 Lincoln V. Schreiber, Gabriel de O. Ramos (Universidade do Rio dos Sinos, UNISINOS), Ana L. C. Bazzan (Universidade Federal do Rio Grande do Sul)

"Explainable AI for Traffic Signal Control"

Ana L. C. Bazzan (c:o/re Aachen)

"XRL: Not as Critical As XAI?"

12:00   LUNCH BREAK

13:30–14:10 Jobst Landgrebe (Cognotekt, Cologne)

"Absolute Limits of Mathematical Modeling in AI"

14:10–14:50 Andreas Kaminski (RWTH Aachen)

"Explainable AI – Transparent AI?"

14:50–15:30 Joffrey Becker (Collège de France, c:o/re Aachen)

Making Sense of Intelligent Systems. From Conception Practices to Interaction Studies

END