



ETHICS AND NORMATIVITY OF EXPLAINABLE AI

Issues in XAI #7

Paderborn University

Paderborn, Germany

May 15-17, 2024

Organised by:

Tobias Matzner, Suzana Alpsancar, Martina Philippi, Wessel Reijers.

The Workshop is part of the series Issues XAI, which continues and expands! We have now appointed a steering committee (SC) consisting of some of its initiators: Juan M. Durán (Delft), Lena Kästner (Bayreuth), Rune Nyruup (Aarhus), and Eva Schmidt (Dortmund). If you are interested in hosting one of the upcoming events, please just get in touch with one of the SC members.



Paderborn Workshop

This workshop will be held at **Paderborn University** on **May 15, 16, and 17**. It will bring together world-leading academics and experts in the field of ethics of explainable AI.

The urgency of this workshop is emphasised by the proliferation of AI systems and their use in decision making processes, from credit score assessments, via risk assessment models, to judicial rulings. This development has only been aggravated with the introduction of machine learning to the public, in particular in the guise of generative AI. This means that people in myriad of contexts and everyday situations will be facing AI-generated decisions. These decisions may have far-reaching consequences, for instance if they lead to a loan rejection, a fraud investigation, or a rejection of parole.

This development has gone hand in hand with a normative call for making AI decision-making *explainable*, explicable, and interpretable. That is, in certain situations it may be desirable that a decision made by an AI *black box* system can be explained, that an answer can be given to the question of why the decision was made.

Yet, Explaining AI is not per se a good thing. In the worst case, explanations can be used to manipulate users or to create acceptance for a technology that is ethically or legally unacceptable. Therefore, research on explaining and explainability needs ethical reflection.

The workshop kickstarts this conversation, focusing on three different themes that are spread out over three workshop days. On the **first day** (May 15), we critically assess the need for explainable AI. Is XAI a good and useful aspect of AI systems at all? And what are its limitations?

The **second day** (May 16) will explore the epistemic and normative dimensions of explainable AI. How can we understand the relation between XAI and epistemic notions of inference and deliberation? How can we consider normative requirements of XAI in the light of a right to explanation and the AI responsibility gap?

The **third day** (May 17) makes the turn to use cases of XAI, embedding the epistemic and normative debates in discussions around XAI in medical and military contexts, as well as in the design of smart cities.

Programme

DAY 1: May 15 – *Why Explainable AI?*

13:30 Welcome and Registration

14:00 Tobias Matzner, Suzana Alpsancar;
Introduction to the workshop

14:30 Katharina Rohlfing: Welcome words

15:00 Dorothea Winter: *Responsibility and Artificial Intelligence*

15:30 Joris Graff: *Why Explainable AI Does not Provide Reason Explanations*

16:00 Scott Robbins: *The (Ethical) Uselessness of XAI*

16:30 Break

17:00 Keynote 1: **Andrea Aler Tubella**

18:00 End of programme

19:00 Dinner first day

DAY 2: May 16 – *Epistemic and Normative Aspects of XAI*

09:30 Welcome coffee

10:00 Zachary Daus: *On the Relation Between Inference and Explanation in AI*

10:30 Martina Philippi & Wessel Reijers: *Ethics of Explainable AI*

11:00 Barend de Rooij: *The Right to Explanation under Conditions of Epistemic Injustice*

11:30 Break

12:00 Daniel Tramp: *Whom to Trust? Simplicity or Transparency?*

12:30 Timo Handwerk: *Towards a Holistic Concept of Explainable AI*

13:00 Keynote 2: **Elettra Bietti**

14:00 Lunch

15:30 Thomas Grote & Norbert Paulo: *A Minimalist Account of the Right to Explanation*

16:00 Filippos Stamatou: *The Responsibility Gap in Explainable AI*

16:30 Wrap up and discussion

17:00 End of programme

19:00 Dinner second day

DAY 3: May 17 – *Use cases of XAI*

09:30 Welcome coffee

10:00 Keynote 3: **Reinhold Häb-Umbach**

11:00 Break

11:30 Maximilian Berger: *Moral Judgement and Decision-making Under Epistemic Uncertainty*

12:00 Nathan Wood (online): *Explainable AI in the Military Domain*

12:30 Sebastian Schleidgen: *Trust Through Explanation?*

13:00 Yannick Fernholz: *Navigating Ethical Dimensions in Smart City Technologies*

13:30 Eugenia Stamboliev: *Can We Foster an AI Literacy That Does Not Only Explain?*

14:00 Wrap up and final words

14:30 Lunch (informal)