

Explainable Intelligent Systems (EIS)

Initiative: Künstliche Intelligenz – Ihre Auswirkungen auf die Gesellschaft von morgen

Ausschreibung: Künstliche Intelligenz – Ihre Auswirkungen auf die Gesellschaft von morgen - Full Grant (nur nach Aufforderung)

Bewilligung: 29.11.2020

Laufzeit: 3 Jahre

Artificially intelligent systems are increasingly used to augment or take over tasks previously performed by humans. These include high-stakes tasks such as suggesting which patient to grant a life-saving medical treatment or navigating autonomous cars through dense traffic—a trend raising profound technical, moral, and legal challenges for society. Against this background, it is imperative that intelligent systems meet central desiderata placed on them by society. They need to be perceivably trustworthy, support competent and responsible decisions, allow for autonomous human agency and for adequate accountability attribution, conform with legal rights, and preserve fundamental moral rights. EIS substantiates the widely accepted claim that meeting these desiderata presupposes human stakeholders to be able to understand intelligent systems' behavior and that to achieve this, explainability is crucial. EIS combines expertise from informatics, law, philosophy, and psychology in a highly interdisciplinary and innovative research agenda to develop a novel 'explainability-in-context' (XiC) framework for intelligent systems. The XiC framework provides concrete suggestions regarding the kinds of explanations needed in different contexts and for different stakeholders to meet specific societal desiderata. It will be geared towards guiding future research and policy-making with respect to intelligent systems as well as their embedding in our society.

Projektbeteiligte

Jun.Prof. Dr. Lena Kästner

Universität des Saarlandes
Philosophisches Institut
Saarbrücken

Prof. Dr. Eva Schmidt

Technische Universität Dortmund
Fakultät 14 Humanwissenschaften und Theologie
Institut für Philosophie und Politikwissenschaft
Dortmund

Prof. Dr. Georg Borges

Universität des Saarlandes
Rechtswissenschaftliche Fakultät
Professur für Bürgerliches Recht,
Rechtinformatik, deutsches und internationales
Wirtschaftsrecht sowie Rechtstheorie
Saarbrücken

Prof. Dr. Ulla Wessels

Universität des Saarlandes
Philosophisches Institut
Professur für Praktische Philosophie
Saarbrücken

Dr. Markus Langer

Universität des Saarlandes
Fakultät für Empirische Humanwissenschaften
Fachrichtung Psychologie
Arbeitseinheit Arbeits- und
Organisationspsychologie
Saarbrücken

Prof. Dr. Cornelius König

Universität des Saarlandes
Fachrichtung Psychologie
Professur für Arbeits- und
Organisationspsychologie
Saarbrücken

Prof. Dr. Holger Hermanns

Universität des Saarlandes
Fakultät für Mathematik und Informatik
Fachrichtung Informatik
Lehrstuhl für Zuverlässige Systeme
und Software
Saarbrücken

Open Access-Publikationen

From Responsibility to Reason-Giving Explainable Artificial Intelligence

