

# Proceedings of the 1st EXplainable AI in Law Workshop XAILA 2018

(<http://xaila.geist.re>)

## Preface

Explainable AI (XAI) has become a prominent research topic during the last few years. Unprecedented development of AI systems and increase of their impact on social life gives raise to important legal and ethical questions. AI algorithms are used to support decision processes, and sometimes they automatically make decisions that may in principle infringe the rights of individuals. The topic of explainability in AI is located on the intersection of computer science, ethics and jurisprudence and it naturally triggers interdisciplinary research questions. How should the notion of explainability (interpretability, transparency) of AI be understood? Is there a “right to explanation” of the work of algorithms and if so, how should it be regulated and implemented in practice? What engineering features should a system satisfy to count as “explainable” one? What is the relation between explanation, argumentation and justification in AI systems? How does the discussion on XAI affect the legal applications of machine learning? What values are relevant in the process of design and implementation of AI systems and how should they be protected? How the discussion on explainability influences the multi-faceted notion of AI itself?

The first edition of the XAILA (eXplainable AI and Law) workshop was held on December 12 2018 at the 31st International Conference on Legal Knowledge and Information Systems - JURIX 2018 (<http://jurix2018.ai.rug.nl>) in Groningen. The workshop was devoted to the discussion of the above mentioned and similar topics. The event attracted significant attendance (more than 30 participants) and 8 papers (6 regular ones and 2 short ones) that were accepted in the comprehensive review process. Upon invitation from the organizers, Professor Bart Verheij from the University of Groningen kindly agreed to deliver a keynote lecture entitled *Good AI and Law*, the abstract of which is published in this proceedings. The remaining part of the volume present revised versions of papers that were discussed during the workshop.

In his paper Harašta discusses legal technology specific regulations and their relation to the concept of Explainable AI. Next, Sileno et. al propose the concept of normware as a foundation of trustworthy and AI systems. Van Otterlo and Atzmueller discuss important requirements as well as design criteria for providing explainability of AI systems in the legal context. One of the solution for this problem are formalized argumentation systems discussed by Araszkievicz and Nalepa. Another formal model applied in the legal AI, based on description logics is presented by Alkmim et. al. Moreover, Yamin and Katt emphasize ethical and legal issues in the development of autonomous adversaries in the cyber domain. In their short paper Araszkievicz and Żurek present a dialogical framework for disputed issues in legal interpretation. Finally, Žolnerčíková discusses challenges of homologation of autonomous machines from a legal perspective.

The editors wish to thank the organizers of Jurix 2018 as well as the member of the international program committee for their support of XAILA!

**Editors:**

Grzegorz J. Nalepa, Jagiellonian University, AGH University of Science and Technology

Martin Atzmueller, Tilburg University

Michał Araszkiewicz, Jagiellonian University

Paulo Novais, University of Minho Braga

**Program Committee of XAILA 2018**

Martin Atzmueller, Tilburg University, The Netherlands

Michał Araszkiewicz, Jagiellonian University, Poland

Kevin Ashley, University of Pittsburgh, USA

Szymon Bobek, AGH University, Poland

Jörg Cassens, University of Hildesheim, Germany

David Camacho, Universidad Autonoma de Madrid, Spain

Pompeu Casanovas, Universitat Autonoma de Barcelona, Spain

Colette Cuijpers, Tilburg University, The Netherlands

Rafał Michalczyk, Jagiellonian University, Poland

Teresa Moreira, University of Minho Braga, Portugal

Paulo Novais, University of Minho Braga, Portugal

Grzegorz J. Nalepa, AGH University, Jagiellonian University, Poland

Tiago Oliveira, National Institute of Informatics, Japan

Martijn von Otterlo, Tilburg University, The Netherlands

Adrian Paschke, Freie Universität Berlin, Germany

Jose Palma, Univesidad de Murcia, Spain

Monica Palmirani, Università di Bologna, Italy

Radim Polčák, Masaryk University, Czech Republic

Marie Postma, Tilburg University, The Netherlands

Juan Pavón, Universidad Complutense de Madrid, Spain

Ken Satoh, National Institute of Informatics, Japan

Erich Schweighofer, University of Vienna, Austria

Piotr Skrzypczyński, Poznań University of Technology, Poland

Dominik Ślęzak, Warsaw University, Poland

Michal Valco, University of Presov, Slovakia

Tomasz Żurek, Maria Curie-Skłodowska University of Lublin, Poland