XAILA 2020 - the Third Explainable & Responsible AI in Law (XAILA) Workshop at JURIX 2020 - the 33rd International Conference on Legal Knowledge and Information Systems
http://xaila.geist.re

Preface

Grzegorz J. Nalepa, Michał Araszkiewicz, Martin Atzmueller, Bart Verheij, Szymon Bobek

In the last several years we have observed a growing interest in advanced AI systems achieving impressive task performance. However, there has also been an increased awareness of their complexity and challenging consequences of their possibly limited understandability to humans. In response, a number of research directions have been initiated. These include humanized or human-centered AI, as well as ethically aligned, ethically designed, or just ethical AI. In many of these ideas, the principal concept seems to be the explanatory capability of the AI system (XAI), e.g. via interpretable and explainable machine learning, inclusion of human background knowledge and adequate declarative knowledge, that could provide foundations not only for transparency and understandability, but also for a possible value alignment and human centricity, as the explanation is to be provided to humans.

Recently, the term responsible AI (RAI) has been coined as a step beyond XAI. Discussion of RAI has been again strongly influenced by the "ethical" perspective. However, as practitioners in our fields we are convinced that the advancements of AI are way too fast, and the ethical perspective much too vague to offer conclusive and constructive results. We are also convinced that the concepts of responsibility, and accountability should be considered primarily from the legal perspective, also because the operation of AI-based systems poses actual challenges to rights and freedoms of individuals. In the field of law, these concepts should obtain some well-defined interpretation, and reasoning procedures based on them should be clarified. The introduction of AI systems into the public, as well as the legal domain brings many challenges that have to be addressed. The catalogue of these problems include, but is not limited to:

(1) the type of liability adequate for the operation of AI (be it civil, administrative of criminal liability);
(2) the (re)interpretation of classical legal concepts concerning the ascription of liability, such as causal link, fault or foreseeability; and
(3) the distribution of liability among the involved actors (AI developers, vendors, operators, customers etc.).

As the notions relevant for the discussion of legal liability evolved on the basis of observation and evaluation of human behavior, they are not easily transferable to the new and disputable domain of liability related to the operation of artificial intelligent systems. The goal of the workshop is to cover and integrate these problems and questions, bridging XAI and RAI by integrating methodological AI, as well as the respective ethical and legal perspectives, also

specifically with support of established concepts and methods regarding responsibility, and accountability. The workshop program included two presentations by invited speakers and eight by authors presenting their research.

Our first invited speaker was Philipp Hacker (Europa-Universität Viadrina, Frankfurt an der Oder) who delivered a lecture `AI and Discrimination: Legal Challenges and Technical Strategies'. The talk focused on the interaction between AI models and liability in the domain of non-discrimination. The author pointed out that the output of AI models may exhibit bias toward legally protected groups. In the past, various fairness definitions have been developed to mitigate such discrimination. Against this background, the talk presented a new model which allows AI developers to flexibly interpolate between different fairness definitions depending on the context of the model application. In the second step the talk inquired to what extent AI developers may risk liability under affirmative action doctrines if they seek to implement algorithmic fairness measures in their models.

The second invited speaker was Reinoud Baker (LexIQ) who delivered a lecture `Legal information systems in production'. The speaker presented LexIQ - a Dutch legal tech startup using data science for legal information services, endorsing the goal to serve citizens, governments and businesses, for instance by improved access to justice, efficient use of resources and enhanced compliance. The talk addressed lessons learned from the past 4 years and focused on the following questions: What can be achieved with modern software and algorithms? How to make innovative technologies available for legal professionals and even the wider public? Which challenges are being encountered?

Barbara Gallina presented the paper `Towards Explainable, Compliant and Adaptive Human-Automation Interaction' (coauthored with Görkem Pacaci, David Johnson, Steve McKeever, Andreas Hamfelt, Stefania Costantini, Pierangelo Dell'Acqua, and Gloria-Cerasela Crisan). The focus is on the responsible design of systems that interact with humans.

Youssef Ennali and Tom van Engers presented the paper `Data-driven AI development: an integrated and iterative bias mitigation approach'. They discuss bias that leads to discriminatory decisions, and the identification and prevention of such bias in an iterative approach aiming at an `unbiased-by-design' methodology.

Heng Zheng presented the paper `Precedent Comparison in the Precedent Model Formalism: Theory and Application to Legal Cases?' (cowritten with Davide Grossi and Bart Verheij). An approach to case comparison is presented in terms of propositional logic formulas, allowing for a generalization and refinement of existing approaches.

Bernardo Alkmim presented the paper `Reasoning over Knowledge Graphs in an Intuitionistic Description Logic' (with coauthors Edward Hermann Haeusler and Daniel Schwabe). The paper uses a natural deduction approach to reasoning over the information modeled in knowledge graphs, with examples in trust, privacy, and transparency.

Annemarie Borg presented the paper `Explaining Arguments at the Dutch National Police' (coauthored with Floris Bex). The paper addresses a basic framework for the argument-based explanation of system conclusions in order to give insight into the underlying decision models and techniques to police analysts and Dutch citizens.

Łukasz Górski, Shashishekar Ramakrishna and Jędrzej Nowosielski presented the paper `Towards Grad-CAM Based Explainability in a Legal Text Processing Pipeline'. Their approach adapts an image processing technique, Grad-CAM to explainability in the setting of legal texts, describing metrics and initial experiments.

Giovanni Sileno presented the paper `Like Circles in the Water: Responsibility as a System-Level Function' (cowritten with Alexander Boer, Geoff Gordon and Bernhard Reader). The paper sketches an approach addressing computational practices that take system environment and consequences of system use seriously.

Karl Branting presented the paper `Explanation in Hybrid, Two-Stage Models of Legal Prediction'. In the paper, core explanation tasks in legal decision support for adjudicators and litigants are identified, a legal prediction model is presented (addressing process initiation and assessment), and associated development requirements are discussed.

The workshop was concluded with a roundtable discussion in which the invited speakers were joined by Karl Branting and Enrico Francesconi as the panelists in a lively discussion with participants.

The workshop organizers would like to thank the Program Committee members for their work in the review process. We are also grateful to the Chairs of JURIX 2020 - the 33rd International Conference on Legal Knowledge and Information Systems for providing the venue for the third edition of the XAILA workshop, following the successful previous editions which accompanied the JURIX conferences in Groningen (2018) and in Madrid (2019), respectively. Finally, we would like to thank our inviting speakers, the authors of papers and all participants to the workshop for their stimulating contributions to the content of the XAILA2020 workshop. The distinctive feature of XAILA is its interdisciplinary character and the creation of a common forum for the exchange of results and opinions from the point of view of such disciplines as legal theory, ethics and Artificial Intelligence and to efficiently combine the theoretical insights with practical focus. The scope of topics covered at this third edition and the high level of the presented contributions and accompanying discussions create a firm basis for the continuation of relevant investigations at the forthcoming editions of the workshop.

Program committee

XAILA 2020 Organizing Committee