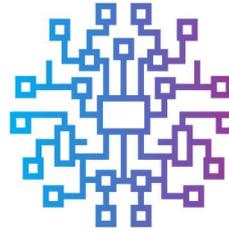


SPONSORED BY THE



Federal Ministry
of Education
and Research



GOAL
Governance von und durch Algorithmen

GOAL International AI Conference

– Governance of and by Algorithms –

Conference Report



On April 8th and 9th, 2021, the International AI Conference hosted by the GOAL project took place. Due to the special circumstances caused by the corona crises, the conference was held online. Both internal and external speakers discussed interesting issues revolving around the governance of algorithms.

Professor Johannes Wessels, Rector of the WWU Münster, and Professor Dr. **Thomas Hoeren**, head of the ITM at the WWU Münster, welcomed approximately 80 participants. In the opening lecture, Assoc. Professor Dr. **Sandra Wachter** of the Oxford Internet Institute spoke about how the extensive inequality and bias that characterize Western societies are inevitably embedded in the use of data for machine learning. Recognizing this problem, much work has emerged in recent years to test for bias in machine learning and AI systems using various bias metrics. Professor Wachter assessed the compatibility of technical fairness metrics and tests used in machine learning against the aims and purpose of EU non-discrimination law. She provided concrete recommendations including a user-friendly checklist for choosing the most appropriate fairness metric for uses of machine learning under EU non-discrimination law.

It emerged from the subsequent Q&A that AI should become fairer and more unbiased than human decision-makers. According to Professor Sandra Wachter, the problem are the current categories and comparison groups, which are not sufficient to make discrimination by AI fully visible. Furthermore, she insisted that making unbiased decision is not a fully achievable goal, but rather a path. The principal objective is the creation of an unbiased society. It was further noted that the EU has regulated only discrimination by humans, but that these regulations are also applicable to many current technologies. Discrimination by AI leads to unequal treatment of humans without them being aware of the discrimination (for example targeted advertising). In this respect, it is not the existing laws that needed to be adapted to these new characteristics of AI. But rather the current AI has to be developed in such a way that is compatible with the current legal system.

After the discussion, Assoc. Professor Dr. **Christian Djeffal** from the Munich Centre for Technology in Society at the Technical University of Munich gave a lecture on the topic: "Democracy's Twilight: Artificial Intelligence, Openness, and Design". He talked about how digital technologies are in the process of reconfiguring our democracy. While we look for orientation and guidance in this process, the relationship between technology and democracy is unclear and seems to be in flux. According to Professor Djeffal, it is time to challenge common narratives about digital technologies and in order to enable novel concepts and new practices

Afterwards, Professor **Katharina Zweig** from the AAL of the University of Kaiserslautern gave a presentation on the topic "Is Fairness by Unawareness fair?". In addition to the question whether protected characteristics should be included in decision-making processes, she used various practical examples to show how human discrimination is reflected in training data sets and how it affects ADM systems.

The topics presented by Professor Djeffal and Professor Zweig were addressed more intense through questions and suggestions from the participants. In this context, Professor Djeffal emphasized that it is necessary to see "justice" in relative terms. The various approaches to AI justice are a highly political decision. Asked how direct democracy treats political decisions and expert knowledge on AI, he explained that there is no simple boundary between direct and representative democracy. Every direct democracy has representative elements and every representative democracy uses direct

elements (e.g. polls). These elements could also be used in the decision-making process in AI regulation. Professor Zweig pointed out that most algorithms that are used in economics do not utilize personal data or make decision about individuals. She stressed the importance of keeping an incentive for further development, which is not possible if there is an over-regulation.

After a brief lunch break, **Reuben Binns**, Assoc. Professor of Human Centered Computing at Oxford University, gave a report on the topic: "Designing Regulation for Algorithms and Algorithms for Regulation". According to Professor Binns, from data protection to competition, consumer protection and equality, regulators are increasingly being called on to investigate and take enforcement action against harmful uses of algorithms. Many are also exploring the use of algorithmic methods to guide and assist in their regulatory activities. This talk will address some of the novel opportunities and challenges raised by the regulation of and by algorithms, including the design of algorithmic tools for regulatory uses; the potential for challenges to algorithm-driven regulation (e.g. judicial review); and the notion of 'algorithmic parity of arms' between regulators and the private/public sector.

Afterward, **Hanna Hoffmann** and **Johannes Kevekordes** from the ITM of the University of Münster explained in their presentation that the growing influence of algorithms on society makes it essential that those who are affected are able to understand and evaluate the decisions made by algorithms. A harmonious interplay between human and machine decisions could only be established if the latter were democratically and transparently controlled. They illuminated the existence of a possible Right to Explanation in the GDPR and the opportunities and challenges for implementing such a right.

In the ensuing discussion, the necessity and feasibility of such a Right to Explanation was intensively debated. Professor Binns also reported on his work experience with the Information Commissioner, the United Kingdom's Data Protection Commissioner and Lier of the British data protection authority.

The discussion was followed by **Catharina Rudschies** and **Matthis Jacobs** from the University of Hamburg, who reported on their latest research findings that ADM systems can be best understood as dynamic and evolving sociotechnical ecosystems. Recognition of these characteristics allows us to take a process perspective that helps to grasp the complexity and multidimensionality of ethical problems throughout the entire technological process – from conception and development to deployment, evaluation over time and continuous evolution.

Frank Pasquale, Professor of Artificial Intelligence Law, Algorithms and Machine Learning Law at Brooklyn Law School gave the final lecture on the first day of the conference. He presented how over the past decade, algorithmic accountability has become an important concern for social scientists, computer scientists, journalists, and attorneys. Exposés have sparked vibrant debates about algorithmic sentencing. Researchers have been exposed tech giants showing women ads for lower-paying jobs, discriminating against the aged, deploying deceptive dark patterns to trick consumers into buying things, and manipulating users toward rabbit holes of extremist content. Public-spirited regulators would have begun to address algorithmic transparency and online fairness, building on the work of legal scholars who have called for technological due process, platform neutrality, and nondiscrimination principles. He said, that this policy work was just beginning, as experts translate academic research and activist demands into statutes and regulations. Lawmakers were proposing bills requiring basic standards of algorithmic transparency and auditing. We would have started down on a long road toward ensuring that AI-based hiring practices and financial underwriting are not used if they have a disparate impact on historically marginalized communities. And just as this "first wave" of algorithmic accountability research and activism would targeted individual encounters with existing systems, an emerging "second wave" of algorithmic accountability would begin to address more structural concerns. Both waves would be essential to ensure a fairer, and more genuinely emancipatory, political economy of technology.

Second wave work would be particularly important when it comes to illuminating the promise & perils of formalizing evaluative criteria.

At the end of the first day, a panel discussion was held by Professor **Frank Pasquale**, Assoc. Professor **Sandra Wachter**, Professor **Joanna Bryson** (Hertie School of Governance) and Dr. **Fabian Niemann** (Partner at Bird & Bird LLP) on “Transparency in Algorithmic Decisions”. The discussion was moderated by Dr. **Nikolas Guggenberger** (Yale Law School). The panelists discussed, among other things, the opportunities and risks of AI, as well as the progress and obstacle that market participants face since the entry into force of the GDPR.

The following day started with a presentation by Dr. **Christina Timko** from the Ruhr University Bochum. She reports on the results of her field study among 150 students who had used a self-designed “news fed reader” application aimed at smartphone users to use the app more often and longer. Smartphone applications intentionally employ a number of behavioral elements, of which the presentation focused on profiling, filtering information, gamification and front-end design elements (e.g. push notifications, infinite scroll, visual cues). In addition to the behavioral elements, they also designed new self-control options and protective measures, such as customizable behavioral change settings.

This contribution was followed by Assoc. Professor **Linnet Tylor** from the Tilburg Institute for Law, Technology and Society, who spoke about the concept of fairness in relation to AI, which could be applied in a formalized way through guidelines and technology regulation. This would be criticized as this approach reduces the question of fairness to a question of requirements engineering. She raised the question of the value of formalizing fairness standards if the objective was to create fair technology governance and how to identify when formalization no longer served overall goal of fairness but turns into its opposite.

This presentation triggered a lively round of questions and discussions among the participants. Professor Judith Simon remarked that there are countless fairness concepts and initiatives, which are often funded by public actors and at the same time do not lead to legislative decisions. Professor Taylor added that, at the moment, many discussions were merely a so-called “ethics theater” that created a space for the economic status quo but did little to change it.

There were clear differences of opinion among the participants regarding the concept of human-centered AI. The discussion was triggered by a comment by Professor Ingrid Schneider, who asked whether Professor Taylor would assess these demands as “ethics washing” or on the contrary as “ethics buildings” (e.g. by making companies more accountable). The question was raised with regard to the EU White Paper and the demands for trustworthy and human-centered AI. While Professor Taylor emphasized that instead of relying on human-centered ethics education, she was in favor of normative concepts, Professor Joanna Bryson advocated a more human rights-oriented and less political debate. However, Professor Taylor pointed out that human rights could not be invoked against companies and that therefore legislative action would be the focus. Towards the end of the discussion, Mr. Marc Hauer raised the exciting question whether ethical approaches for a more equal AI were the right way to go or would be ignored anyway. According to Professor Taylor, work has to be done both at the individual ethical level (micro-ethics: individual responsibility) and at the collective level (macro-ethics: political responsibility). Ethics should not be produced as dogma but should be able to deal with moving goals and be able to consider movements as part of the ethical challenge.

Afterwards, Professor **Joanna Bryson** of the Hertie School in Berlin spoke about what she believes to be an untruth in the political discourse on AI regulation: That AI is necessarily opaque. In fact, artificial

intelligence is not necessarily more opaque than natural intelligence and AI can be made much more transparent by design. She highlighted at the technological, sociological and economic barriers to transparency, how these are affected by AI and the digital revolution and what governance policies should be used to address them.

Dr. Carsten Orwat from the ITAS of the Karlsruhe University concluded his presentation with the challenges associated with the risk regulation of AI and ADM. Those would arise mainly from the characteristics of the risks considered, i.e. the risks to fundamental rights and values. This included normative ambiguities in the interpretation and operationalization of fundamental rights and values. In addition, there are types of risks that go beyond the individual level. Ubiquitous value conflicts and uncertainties about risks also challenged the determination of acceptable levels of risk. Such risk characteristics would have an impact on the design of risk regulation.

After the lunch break, Andrew Burt, managing partner at bnh. ai, a boutique law firm focused on AI and analytics, and Chief Legal Officer at Immuta Visiting fellow at Yale Law School's Information Society Project, gave a presentation. He spoke about the deep privacy and security gaps in AI systems and the three implications of this trend, which he summarized as follows: "1. Privacy is dead. 2. So is trust. 3. And you are not who you think you are." After a brief overview of these trends, he gave specific suggestions on how to address them - from the perspective of lawyers, politicians and citizens.

In the final round of discussion, the participants focused on the proposals put forward to counteract these trends. The idea of slowing down AI developments was criticized. This is unrealistic, noted one conference participant, as market innovation is driven by the consumer, to which the response was that the standards that are constantly on goods need to be applied more to software, so that the recommendation for slowing down is not fundamentally at odds with the market. Following on from his proposals for restoring privacy, the participants continued to discuss data protection by the GDPR with Mr. Burt. They agreed that the GDPR, in addition to its good approaches, presents great difficulties in terms of implementation. On the one hand, Mr. Burt feared that new laws would be implemented in such a way that they were either too resource-intensive and only large companies could implement and enforce them. On the other, they could be too theoretical so that they would not be implemented in practice.

After the discussion, Hanna Hoffman closed the conference with a big thank you to all the speakers and participants.

The contributions of the external speakers as well as the controversial discussions led to new impulses and provided detailed insights into the opportunities and risks of artificial intelligence. In particular, the differentiated perspectives of the participants in the course of the international dimension of the conference opened up new perspectives and broadened horizons. Already during the conference, some of the experts agreed to continue their exchange in the future.

The GOAL team would like to take this opportunity to say thank you: At the BMBF and DLR for making the event possible; all speakers for the great presentations; to the participants of the panel discussion for the lively debate; and to all conference participants who enriched the event with their interested questions and contributions to the discussion.