

Program 15 July

MORNING SESSION

Chair: **David Billard** | *HES-SO University of Applied Sciences in Geneva, CH*

09:30	Chiara Gallese Nobile <i>Data Management & Library (DML), TU/e</i>	Opening
09:45	Carlo van de Weijer <i>General manager EAISI, TU/e</i> Iza Witkowska <i>Head of RDM and Privacy, TU/e</i>	Salutations
10:00	<u>Stefania Costantini</u> <i>University of L'Aquila, IT</i>	Introductory information of COST Action
10:15	<u>Marta Poblet Balcell</u> <i>RMIT University, AU</i>	AI governance and collective intelligence: opportunities and challenges
10:45	Coffee break	
11:00	<u>Vincent Müller</u> <i>Eindhoven University of Technology, NL</i>	Deep Opacity Undermines Data Protection and Explainable AI
11:30	<u>César Hidalgo</u> <i>University of Toulouse, FR</i>	How humans judge machines
12:00	Discussion among panelists	
12:30	Lunch break	



Carlo van de Weijer



Iza Witkowska



Stefania Costantini



Marta Poblet Balcell



Vincent Müller



César Hidalgo

Stefania Costantini | University of L'Aquila, IT

Introductory information of DIGforASP COST Action



ABOUT THE SPEAKER

Stefania Costantini is Full Professor in Computer Science at the Department of Computer Science and Engineering and Mathematics (DISIM). She is the Head of the research group AAI@AQ (Autonomous Agents and Artificial Intelligence at the University of L'Aquila). She has more than 150 publications. Her research interests are in (theory and practice of) Artificial Intelligence and Computational Logic, including Intelligent Software Agents and Multi-Agent Systems, Answer Set Programming, Non-Monotonic Reasoning, Knowledge Representation, Cognitive Robotics. She invented, defined and coordinated the first implementation of the DALI agent oriented logic programming language. She served in the Program Committee of the main Conferences of her fields of interest, and she is a member of the Editorial Board of the journal Theory and Practice of Logic Programming (Cambridge). She is currently the President of the Italian Association of Computational Logic (GULP), and Member of the Board of the Italian Association for Artificial Intelligence (AIxIA).

Marta Poblet Balcell | RMIT University, AU

AI governance and collective intelligence: opportunities and challenges



ABSTRACT

The EC Artificial Intelligence Act (AIA) proposal published on 21 April 2021 constitutes another step in the process of building a harmonised regulatory framework for AI in the European Union. The AIA proposal draws from previous legislative processes establishing a risk-based approach with extra-territorial effect. The proposal also sets a two-level governance system composed of a new European Artificial Intelligence Board and member state authorities with competences at the national level. This paper argues that the design of this top-down governance system remains state-centric and misaligned with the extra-territorial scope of the future legislation. Moreover, the two-tiered design falls short from leveraging both the expertise and stewardship that citizens and organisations can provide. The paper also outlines some principles to facilitate more distributed models seeking a balance between institutional stability and bottom-up innovation, thus contributing to further research towards decentralised AI governance models.

ABOUT THE SPEAKER

Marta Poblet is an Associate Professor at RMIT's Graduate School of Business and Law and the Director of RMIT's Social Change Enabling Capability Platform. She is one of the co-founders of the Institute of Law and Technology at the Autonomous University of Barcelona and past researcher at ICREA (Catalonia). She holds a JSD in law (Stanford University 2002) and a Master in International Legal Studies (Stanford University 2000). Her research interests cover different areas at the intersection of law, political sciences and sociology, and technology. She is also interested in the connections between technology developments (AI, blockchain, human-computer interaction) and the different theories of democracy, participation, and citizenship. She has been the PI of a number of national and international research projects and has published over 80 scientific articles on these topics in journals and books.

Vincent Müller | Eindhoven University of Technology, NL

Deep Opacity Undermines Data Protection and Explainable AI



ABSTRACT

It is known that big data analytics and AI pose another threat to privacy, and it is known that there is some kind of ‘black box problem’ in AI. In the paper, I propose that (1) the ‘black box’ becomes the ‘black box problem’ in a context of justification for judgments and actions, crucially in the context of privacy. (2) I suggest distinguishing two kinds of classic opacity and introducing a third: The subjects may not know what the system does (‘shallow opacity’), the analysts may not know what the system does (‘standard black box opacity’), or even the analysts cannot possibly know what the system might do (‘deep opacity’). (3) If the agents, data subjects as well as analytics experts, operate under opacity, then they cannot provide some of the justifications for judgments that are necessary to protect privacy – e.g. they cannot give “informed consent” or assert “anonymity”. It follows from (2) and (3) that agents in big data analytics and AI, often cannot make the judgments needed to protect privacy. So big data analytics makes the privacy problems worse, and the remedies less effective.

ABOUT THE SPEAKER

Vincent C. Müller is Professor for Philosophy of Technology at the Technical University of Eindhoven (TU/e) - as well as University Fellow at the University of Leeds, Turing Fellow at the Alan Turing Institute, London, President of the European Society for Cognitive Systems and Chair of the euRobotics topics group on 'ethical, legal and socio-economic issues'. He was Professor at Anatolia College/ACT (Thessaloniki) (1998-2019), Stanley J. Seeger Fellow at Princeton University (2005-6) and James Martin Research Fellow at the University of Oxford (2011-15). He studied philosophy with cognitive science, linguistics and history at the universities of Marburg, Hamburg, London and Oxford.

Müller is known for his research on theory and ethics of disruptive technologies, particularly artificial intelligence (AI). He has published widely on the philosophy of AI and cognitive science, philosophy of computing, philosophy of language, applied ethics, etc. (citations >250/year). Müller edits the "Oxford handbook of the philosophy of artificial intelligence" (OUP), wrote the Stanford Encyclopedia of Philosophy article on Ethics of AI and Robotics and has a book forthcoming with OUP on "Can Machines Think?". [More information](#)

César Hidalgo | University of Toulouse, FR

How humans judge machines



ABSTRACT

How would you feel about losing your job to a machine? How about a tsunami alert system that fails? Would you react differently to acts of discrimination performed by a machine or a human? How about public surveillance? How Humans Judge Machines compares people's reactions to actions performed by humans and machines. Using data collected in dozens of experiments, this book reveals the biases that permeate human machine interactions. Are there conditions in which we judge machines unfairly? Is our judgment of machines affected by the moral dimensions of a scenario? Is our judgment of machines correlated with demographic factors, such as education or gender? Hidalgo and colleagues use hard science to take on these pressing technological questions. Using randomized experiments, they create revealing counterfactuals and build statistical models to explain how people judge A.I., and whether we do it fairly or not. Through original research, they bring us one step closer to understanding the ethical consequences of artificial intelligence.

How Humans Judge Machines can be read for free at <https://www.judgingmachines.com/> (in print with MIT Press).

ABOUT THE SPEAKER

Cesar A. Hidalgo directs the Center for Collective Learning at the Artificial and Natural Intelligence Institute (ANITI) at the University of Toulouse. Prior to joining ANITI, he directed the Collective Learning group at MIT. Hidalgo holds a PhD in Physics from the University of Notre Dame, and is the author of dozens of peer reviewed papers and three books. His latest book is How Humans Judge Machines (MIT Press, 2021).

Program 15 July

AFTERNOON SESSION

Chair: **Chiara Gallese Nobile** | *Eindhoven University of Technology, NL*

13:30	Chiara Gallese Nobile <i>Data Management & Library (DML), TU/e</i>	Introduction to the afternoon session
13:45	<u>Alexandros Kalousis</u> <i>HES-SO University of Applied Sciences, Western Switzerland</i>	AI's blind spots and societal challenges
14:15	<u>Francisco de Elizalde</u> <i>IE University, ES</i>	Automation of Justice. How law could determine technological success
14:45	Coffee break	
15:00	<u>Andrea Pin</u> <i>University of Padua, IT</i>	Rage Against the Machine? The Fear of ADM and Comparative Constitutional Law
15:30	Discussion among panelists	
16:00	Discussion with the public	



Alexandros Kalousis



Francisco de Elizalde



Andrea Pin

Alexandros Kalousis | HES-SO University of Applied Sciences, Western Switzerland

AI's blind spots and societal challenges



ABSTRACT

The progress in deep learning research has taken the world by storm, exhibiting super human performance in pattern recognition problems. The latest research results and models are now routinely deployed in the real world in settings as diverse as recommendation engines for e-commerce and social media, face recognition, ranging from social media settings to security, providing decision support to determine access to resources, such as social aide and support, even to assess the risk of recidivism to support sentencing decisions and many more. In fact to a considerable extend research is now led by private actors and the industry. As a result of the progress pace, we have lagged behind in the development of a solid understanding of these learning models; adversarial examples provide a clear demonstration of the blind spots of these models by making them fail in settings which are very easy for humans, providing evidence that the decision mechanisms of these models capture features that are not related to the semantics of their inputs, at least not in the way humans do. Things can get considerably more complicated when it comes to learning agents that need to learn to interact within a dynamic environment, as it is the case with reinforcement learning or whenever machine learning systems are used in feedback loops. Such configurations typically appear in the real world when it comes to personalised content recommendation systems which learn from users' actions. Such systems will maximize some measure of the user's interaction with the platform, as a proxy for revenue maximization from ad placement. The interaction of such models with humans in a very large scale produces some rather worrying results for society at large, and is associated with echo chamber effects polarisation, and radicalization. The use of recommendation systems can also raise serious questions concerning the loss of autonomy of the individual. The latter is not respected when recommendations respect third party interests prior to those of the individual.

In this talk I will briefly review some of the weak points of the deep learning systems from a technical perspective. In addition I will discuss cases in which their deployment in real worlds settings had detrimental effects for individuals, groups of individuals and the society at large.

ABOUT THE SPEAKER

Alexandros Kalousis is a Professor at the University of Applied Sciences, Western Switzerland in the Geneva School of Business Administration, where he leads the Data Mining and Machine Learning group. He also holds a "Chargé de Cours" position in Department of Computer Science at the University of Geneva. He received his PhD from the University of Geneva in the area of machine learning. He has a BSc and MSc in Computer Science in the University of Athens, Greece.

He publishes regularly in the top venues of data mining and machine learning, such as ICML, NIPS, ICDM, KDD, SDM, ECML/PKDD. He is a regular program committee member of all main machine learning and data mining conferences.

Francisco de Elizalde | IE University, ES

Automation of Justice. How law could determine technological success



ABSTRACT

It is a common opinion that the use of automated decision making in the resolution of private disputes is optimal for small value claims. Therefore, effectiveness of ADM in the administration of justice would depend on the economic significance of claims. This presentation challenges that view as it posits that the drafting of legal rules (and not the value under dispute) is key for ADM to succeed in adjudication. Some laws are better suited than others for automation, which is an observation that results from the 2021 Legal Tech in Private Claims Survey. The presentation will explore the characteristics of laws that enhance or otherwise hinder the automated enforcement of law. An accurate understanding of those features would potentiate ADM in adjudication and would tackle better the difficult issues of transparency and the right to a fair trial in that context.

ABOUT THE SPEAKER

Francisco de Elizalde is an Associate Professor at IE University, Law School (Spain) and the current Chair of Legal Studies. He focuses on Comparative Private Law, especially Contracts and the Law of Property. He is a permanent Visiting Professor at Koç University (Turkey) and has held Visiting Professorship positions at City University of Hong Kong, FGV Sao Paulo (Brazil) and the Law Schools Global League. Prof. Elizalde was a Visiting Scholar to the University of Cambridge and has conducted research at Harvard Law School and the Max Planck Institute for Comparative and International Private Law (Germany). He is a member of the Madrid Bar Association, of the American Society of Comparative Law and of the European Law Institute. Prof. Elizalde is the head of the EU-financed Jean Monnet Module “Liability of Robots: a European Vision for a New Legal Regime”. His latest publication on this topic is the co-edited book *The Cambridge Handbook of Lawyering in the Digital Age*, CUP, 2021.

Andrea Pin | University of Padua, IT

Rage Against the Machine? The Fear of ADM and Comparative Constitutional Law



ABSTRACT

The ever-growing interest in ADM issues across jurisdictions is prompting broad reconsiderations on how to square the benefits of integrating AI-based tools in contemporary societies with the protection of rights. If we focus on the American and the European models, which both pay special attention to individual rights' concerns, we notice two diverging dynamics.

As is widely known, the U.S. approach has traditionally tried to leverage AI to boost the economy, whereas the E.U. has been more protective of individuals within the market, thereby making privacy rules the main dam to prevent unwarranted exploitation of personal data. The American-European divide on ADM is transitioning into a new phase. While the U.S. is driven by racial preoccupations and therefore mostly concerned with equality issues, the EU seems to be moving toward a new goal-oriented approach - one that prefers vetting the specific purposes of a given ADM, instead of implementing general bans on ADM protocols. These two patterns suggest that despite AI's global reach, its regulations will develop a patchwork of ever-changing legal solutions, which will be dictated by social sensitivity, political priorities, and institutional infrastructure.

Program 16 July

Chair: **Elena Falletti** | *Carlo Cattaneo University LIUC, IT*

09:15	Elena Falletti <i>Carlo Cattaneo University LIUC, IT</i>	Welcome and introduction to the second day
09:30	<u>Angela Müller</u> <i>Senior Policy & Advocacy Manager, Algorithm Watch, CH</i>	Automated Decision-Making, Ethical Implications, and Legal Approaches
10:00	<u>Sicco Verwer</u> <i>Delft University of Technology, NL</i>	Bias and discrimination in data mining
10:30	Coffee break	
10:45	<u>Susana Vieira</u> <i>University of Lisbon, PT</i>	Transparency and Reproducibility in Artificial Intelligence in Health Care Research
11:15	Discussion among panelists	
11:45	Discussion with the public	
12:15	David Billard <i>HES-SO University of Applied Sciences in Geneva, CH</i>	Closing remarks



Angela Müller



Sicco Verwer



Susana Vieira

Angela Müller | Senior Policy & Advocacy Manager, Algorithm Watch, CH

Automated Decision-Making, Ethical Implications, and Legal Approaches



ABSTRACT

The use of automated decision-making (ADM) systems can have a range of implications on individuals and society. In order to ensure that their use actually enhances individual autonomy, freedom, and the common good – instead of reducing it – we need a governance framework for such systems, which includes (but is not limited to) legal regulation. One approach to legal regulation has recently been proposed by the European Commission. The presentation will focus on AlgorithmWatch’s perspective on the implications of ADM systems, the necessary elements of such a governance framework, and our assessment of the Commission’s proposal.

ABOUT THE SPEAKER

Angela Müller leads the Policy & Advocacy team at AlgorithmWatch, a not-for profit organization committed to watch, unpack and analyze algorithmic decision-making (ADM) systems and their impact on society. She has M.A. in Political and Economic Philosophy and wrote her Ph.D. thesis in Law, where she focused on human rights—inter alia in light of the use of new technologies. Prior to her current position, Angela Müller was project leader at a foreign policy think tank, project manager for an innovation platform, and worked for the Swiss Foreign Ministry.

Sicco Verwer | Delft University of Technology, NL

Tba

ABSTRACT

About a decade ago, we started to research how to avoid bias and discrimination in data mining. We discovered that if one would naively apply classifiers to data involving people, the resulting predictions would correlate (negatively) with a person's race or gender. In fact, the bias that may be present in the data gets amplified. How to prevent this and perhaps even correct for the data's bias? In this talk, I will explain this problem and what (not) to do about it.

ABOUT THE SPEAKER

Dr. Sicco Verwer is associate professor in machine learning for software and cyber security at Delft University of Technology where he heads the cyber analytics lab. His research is focused on developing new learning algorithms for intrusion/anomaly detection and bug discovery for which he received VENI and VIDI grants from NWO. His work on discrimination-aware data mining got awarded a test-of-time award at the European Conference on Machine Learning (ECMLPKDD).



Susana Vieira | University of Lisbon, PT

Transparency and Reproducibility in Artificial Intelligence in Health Care Research



ABSTRACT

Clinical medicine relies on a strong research foundation in order to build the necessary evidence base to inform best practices and improve clinical care, however, large-scale randomized controlled trials (RCTs) are expensive and sometimes unfeasible. Secondary use of electronic health records (EHR) has been rapidly growing, so data contained in EHR is not only used for the patients' primary care but also for various secondary purposes such as clinical research, automated disease surveillance and clinical audits for quality enhancement. The healthcare industry has rapidly become computerized and digital. Most healthcare delivered in America today relies on or utilizes technology. Modern healthcare informatics generates and stores immense amounts of detailed patient and clinical process data. Today, real-world patient data is being used to further advance the field of health care. One of the large barriers to the utilization of these data was, and still is, the inaccessibility to researchers. Making these databases easier to access as well as integrating the data would allow more researchers to answer fundamental questions of clinical care. Nevertheless, reproducible and transparent studies using EHR are not as often as desirable. Data can be overwhelmingly complex or incomplete for any individual, and all the steps need to be communicated, otherwise becomes impossible to reproduce and validate the study. Further, only a few of these studies are medically validated and interpreted. We urge multidisciplinary research teams consisting of clinicians along with data scientists to unpack the clinical semantics necessary to appropriately analyse the data and understand the knowledge that can be achieved using Artificial Intelligence.

ABOUT THE SPEAKER

Susana Vieira received the MSc and PhD degrees both in Mechanical Engineering (Mec. Eng.) in 2005 and 2010, respectively, from Instituto Superior Técnico (IST), University of Lisbon, Portugal. She was a Teaching Assistant at IST, in Mec. Eng., from 2005 to 2006, in 2009, she was an Invited Teacher at the Erasmus University of Rotterdam, the Netherlands, and in 2014 she was a visiting Scholar at MIT, Boston, USA. From 2012 to 2018 she was an invited assistant professor at IST, University of Lisbon and since November 2020 she is an associate professor at IST, University of Lisbon and a Senior researcher at the Center of Intelligent Systems (CIS), IDMEC- IST since 2012.

Her main research area is Soft Computing and Artificial Intelligence, more specifically she works in machine learning, feature selection, fuzzy modelling, fuzzy optimization and metaheuristics. Her research focuses mainly on the development of computational intelligence methods for knowledge data discovery. Recently these methods are being used to identify important factors or features that lead to unfavourable or favourable clinical conditions of patients in Intensive Care Units, and design specific decision models that support clinicians' decisions.