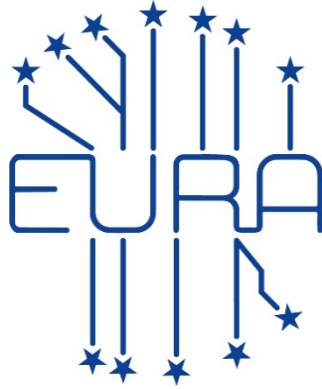




**Sant'Anna**  
School of Advanced Studies – Pisa

Co-funded by the  
Erasmus+ Programme  
of the European Union



# Human Robot Interaction between Trust and Deception

Pisa, October 31<sup>st</sup> 2019

Scuola Superiore Sant'Anna, Aula Magna  
Piazza Martiri della Libertà, 33  
56127, Pisa, Italy

## CONFERENCE DESCRIPTION

**The Conference.** The first EURA conference intends to deal with the emerging and potentially revolutionary field of social applications in robotics and AI. Advanced technologies promise to revolutionize our society, by allowing for the emergence of new goods and services, which are likely to strongly change our lives on an everyday basis. Potential applications are, indeed, countless. However, a paradigmatic shift – often accounted for in science fiction – is represented by applications that mimic human behaviour, being specifically designed to engage users in human-like interactions, becoming «your new best friend»<sup>1</sup>.

Indeed, machines are not alive, self-conscious, do not perceive their own existence, neither do understand the context, nor that a human is present before them, and that he is other-than-self. Yet machines could create such an appearance, inducing the user to believe that indeed they are alive, possess feeling and are capable of bonding. After all, the very definition of AI promoted by Alan Turing is grounded on the ability of the machine to mislead a human-judge.

Such traits are often actively pursued for it is commonly believed they would ensure good and agreeable design, and users might even enjoy such kinds of interactions, indeed perceiving the machine to be their new friend, something more than a thing.

In a social science and regulatory perspective, it is therefore essential to understand if such a trend is indeed admissible and unproblematic, so long as users are informed and freely consent or, to the contrary, whether some uses and applications ought to be prohibited. Said otherwise, if an ethical approach to AI requires applications to be «trustworthy»<sup>2</sup>, should machines be allowed to deceive their users?

The interplay between the fundamental principles of human dignity and freedom of self-determination – as emerging from the European Charter of Fundamental Rights – plays a role in such a complex balancing, yet tracing lines, despite necessary is neither easy nor obvious.

The complexity of the matter requires multiple perspectives to be taken into account, engineers, lawyers, ethicists, as well as psychologists are called in to contribute to the debate.

**The Centre of Excellence.** The EURA Jean Monnet Centre of Excellence constitutes a focal point of competence and knowledge on Robotics and AI, focusing on their ethical, legal, social and economic (ELSE) implications. EURA's mission is to promote innovative multidisciplinary research, offer advanced educational programs, and foster the dialogue with policy makers, increasing social awareness and promoting an informed debate. Through its activities, as well as its interdisciplinary and functional-based approach, EURA creates a worldwide network of experts, professionals, stakeholders, and policy makers, facilitating cross-fertilization among different fields and interest groups. Ultimately, EURA intends to identify, assess, discuss, and promote the European Approach to AI and advanced robotics, as defined by European Commission in its communication of the 25th April 2018<sup>3</sup>.

---

<sup>1</sup> <https://www.prnewswire.com/news-releases/meet-jibo-your-new-best-friend-introducing-the-first-social-robot-for-the-home-300542856.html>.

<sup>2</sup> «Trustworthy AI has three components, which should be met throughout the system's entire life cycle: (1) it should be lawful, complying with all applicable laws and regulations (2) it should be ethical, ensuring adherence to ethical principles and values and (3) it should be robust, both from a technical and social perspective since, even with good intentions, AI systems can cause unintentional harm»: High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, 2019).

<sup>3</sup> European Commission, *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Artificial Intelligence for Europe*, 2018.

## CONFERENCE PROGRAM

### Registration and Welcome 8:30-9:30

### Morning Session 9:30 – 12:30

#### Chair

Caterina Sganga

Associate Professor of Comparative Private Law, Scuola Superiore Sant'Anna, Pisa, Italy

#### Speakers

Robert Sparrow

Professor of Philosophy, Monash University, Australia

Alistair Isaac

Senior Lecturer of Philosophy, University of Edinburgh, UK

Giampaolo Ghilardi

Researcher in Moral Philosophy, Campus Biomedico University, Rome

Andrew McStay

Professor in Digital Media, Bangor University, UK

Andrea Bertolini

Assistant Professor of Private Law, Scuola Superiore Sant'Anna, Pisa, Italy

### Afternoon Session 15:00 – 18:00

#### Chair

Gianluigi Palombella

Professor of Applied Legal Theory, Scuola Superiore Sant'Anna, Pisa, Italy

#### Speakers

Massimo Bergamasco

Professor of Applied Mechanics, Scuola Superiore Sant'Anna, Pisa, Italy

Erica Palmerini

Associate Professor of Private Law, Scuola Superiore Sant'Anna, Pisa, Italy

Antonio Chella

Professor of Robotics, University of Palermo, Italy

Tobias Mahler

Professor, Norwegian Research Center for Computers and Law, Department for Private Law, University of Oslo, Norway

### Presentation of the “EURA Young Scholar Prize” Papers

### Social Dinner 20:00

#### Scientific Coordinator

Andrea Bertolini

Scuola Superiore Sant'Anna – DIRPOLIS Institute  
Piazza Martiri della Libertà 33, 56127 Pisa (PI) – Italy

Tel: 0039 050.88.19.47

[andrea.bertolini@santannapisa.it](mailto:andrea.bertolini@santannapisa.it)

#### Conference Committee

Andrea Bertolini

Francesca Episcopo

Sira Grosso

E-mail: [aura@santannapisa.it](mailto:aura@santannapisa.it)

<http://www.aura.santannapisa.it/>

## BIOS AND ABSTRACTS – Morning Session



**Caterina Sganga** holds a Ph.D. in Comparative Private Law from Sant’Anna, an LL.M. from Yale Law School, and an LL.B. and J.D. from University of Pisa, and is currently Associate Professor of Comparative Private Law at Sant’Anna. Her research focuses on EU copyright law, IP and new technologies (AI, IoT, cloud computing, cyberspace, digitization), IP and innovation policies, the interplay between IP and human rights, data ownership and management in the era of big data. She is a member of several international IP and property law associations (EPIP, ATRIP, ALPS) and a fellow of the EUI.



**Robert Sparrow** is a Professor in the Philosophy Program, a Chief Investigator in the Australian Research Council Centre of Excellence for Electromaterials Science, and an adjunct Professor in the Monash Bioethics Centre, at Monash University, where he works on ethical issues raised by new technologies. He is a co-chair of the IEEE Technical Committee on Robot Ethics and was one of the founding members of the International Committee for Robot Arms Control.

**Love and Robots** \_ It’s often suggested that if people were to behave towards a robot precisely as they would towards a human being with whom they were in love, then that would be sufficient to establish that they loved the robot. In this paper, I argue that this approach to love and other social emotions is deeply mistaken. The vocabulary we use to talk about emotions and social relationships is inherently critical. That is to say, it presumes — and thus relies upon — the distinction between genuine instances and false semblances of emotions and relationships. Moreover, the normative framework that we employ to identify and evaluate emotions and relationships includes judgements about their proper objects. A philosophically informed and suitably nuanced account of emotions and social relationships therefore suggests that what is required to show that someone loves someone or something is much more demanding than most of the literature on social robotics acknowledges.



**Alistair M. C. Isaac** is Senior Lecturer in Philosophy of Mind and Cognition at the University of Edinburgh. His research focuses primarily on the nature of representation in science and cognitive science. He was led to the ethics of deceptive robots by considering the question of deception-detecting robots in collaboration with Computer Scientist Will Bridewell (Navy Center for Applied Research in Artificial Intelligence). They have argued that the representational requirements are identical for deceiving and deception-detecting agents. Their work on these topics includes “White Lies on Silver Tongues: Why Robots Need to Deceive (and How)” in the anthology *Robot Ethics 2.0*.

**When to Trust a Liar** \_ Our folk understanding of “deception” combines two features: (i) actions aimed to induce false belief; and (ii) malicious intent. Yet these come apart in practice. (ii) is present without (i) in the case of “paltering,” or the utterance of truths with the intent to mislead; (i) is present without (ii) in the case of “white lies” that serve a pro-social function. I argue that, in order to engage socially with humans, robots must be capable of regularly producing (i). In fact, there are positive ethical reasons for wanting robots to have the capacity to lie in complex social interactions, as they may need to do so to satisfy Asimov’s Laws. The real requirement for ensuring trustworthy robots is to solve the problem of preventing malicious intent. This problem is much more pressing, as it already confronts us today, for instance in military drones.



**Giampaolo Ghilardi** graduated cum laude in Philosophy in 2002 at the Università Cattolica of Milan, and obtained a Phd in Bioethics in 2011 at the Università Campus Bio-Medico Rome. After 4 years of post-doctoral fellowship in the same University he was appointed Assistant Professor in Moral Philosophy in 2017. He got the Associate Professor qualification in 2018. He is a member of the Institute of Philosophy of Scientific and Technological Acting, the Italian Society of Moral Philosophy, the Interdisciplinary Group of Clinical Bioethics and Ethical Consulting

in Health Care, the Sciences, Literature, and Arts Academy of Bergamo (moral sciences section), in 2018 he won the Michele Erede Prize for a paper on Precision Medicine.

**Title: Is deceiving always wrong?** \_ Although deception is normally conceived as a bad action, entailing the idea of concealing true intentions from another person in order to mislead her, human robot interaction is bringing up new scenarios that need to be addressed. Consider the case of highly advanced prosthetics, where new “robotic” limbs in order to work properly need to “deceive” the amputee neural system and make him believe/think that the pieces of information gathered by the mechanical limb are truly yours. Can we consider this human machine interface under the category of deception? Is there an embodied intelligence different from the artificial intelligence paradigm we need to pay attention to? Body deception in order to get better perception has to be considered deceiving? Finally, isn’t the very idea of interface kind of deceptive itself? After all, interfacing literally means to put two faces in relation whereas, strictly speaking, there is just one real face.



**Andrew McStay** is Professor of Digital Life at Bangor University, UK. He publishes on issues central to the digital economy, including privacy and artificial intelligence. His recent book, *Emotional AI: The Rise of Empathic Media*, examines the impact of technologies that make use of data about emotional life. Non-academic work includes IEEE membership (P7000/7014) and ongoing advising roles for start-ups, NGOs and the UK’s Information Commissioner’s Office. He has also made appeared and made submissions to the United Nations Office of the High Commissioner on

the right to privacy in the digital age, the UK House of Lords AI Inquiry and the UK Department for Culture, Media and Sport Inquiry on fake news and reality media.

**Emotional AI and the rise of empathic technologies** \_ Once off-limits, the boundaries of personal space and the body are being tested by emotional AI and affect-sensitive applications in worn, domestic and public capacities. As cities, workplaces, homes, devices and media content transform through novel sensing and data usage, this raises questions about privacy, civic life, influence, regulation, moral limits and the desirability of profiling emotional life. Introducing his most recent book, ‘*Emotional AI: The Rise of Empathic Media*’, McStay will discuss the nature and scope of emotion-sensing and empathic technologies. Largely focusing on retail, marketing and advertising, McStay will assess the rise of emotional AI in public spaces. This will be explored in reference to: 1) interviews with stakeholders (industry, policy makers, data protection NGOs and the legal community); 2) UK survey work on how citizens feel about these developments; and 3) whether legal classifications of personal and sensitive data are sufficient, especially in out-of-home situations.



**Andrea Bertolini** is Assistant Professor of Private Law at the DIRPOLIS Institute, Scuola Superiore Sant'Anna, and adjunct professor in private law at the University of Pisa. His research ranges from private law (contracts, torts, law of obligations) to the regulation of robotics and AI and bioethics, with a comparative and law and economics approach. Since 2015 he works, either as coordinator or WP leader, in many national and European projects on the regulation of robotics and AI (the SIR project 'RELIABLE', the Jean Monnet Module 'Europe Regulates Robotics', the H2020 project INBOTS), and is currently the director of the Jean Monnet Centre of Excellence on the Regulation of Robotics and AI "EURA". Dr Bertolini holds a jointed degree from the Scuola Superiore Sant'Anna and the University of Pisa, a PhD in private law from the Scuola Superiore Sant'Anna, as well as a LL.M. from Yale Law School. He is an attorney licensed to practice in Italy and New York.

**A Neo-Kantian Approach to Human Dignity and Robot Design. Setting the Framework for an Acceptable Human-Robot Interaction** \_ Specific applications, namely robot companions, are researched and designed primarily for the purpose of interacting with humans for the widest range of purposes, to deliver services, to cure and provide care, to teach and entertain. After discussing the notion of «robot» to exclude the possibility of deeming existing applications «agents», the issue of deception is addressed, starting with the very definition of «artificial intelligence» emerging from the Turing test. If current machines are not truly intelligent, do not possess a conscience, nor are aware of their very existence, still they may simulate those characteristics, eliciting emotional reactions on the side of human beings. Engineering studies are presented that address the use of such applications with children and the elderly for a wide range of purposes, actively pursuing some form of emotional engagement that might amount to deception. The ethical framework is briefly introduced, as well as the legal, to deny validity to purely utilitarian stances as well as to the self-sufficiency of the criterion of freedom of self-determination.

## BIOS AND ABSTRACTS – Afternoon Session



**Gianluigi Palombella** is Professor of Applied Legal Theory, Coordinator of the PhD program in Law, J.D and PhD at the Scuola Sant'Anna. He was “F. Braudel” Professorial Fellow at the European University Institute, Invited Research Fellow at the UNSW in Sydney, “MacCormick” Fellow at the Law School of the University of Edinburgh. He was awarded the Inaugural Fellowship of the Straus Institute of Law and Justice at the NYU Law School (2009-2010), and the Distinguished Global Governance Professorial Fellowship at the Schuman Center (2013)

Along with his background research on the rule of law and fundamental rights, his latest scientific projects relate to regulation of environmental sustainability, the protection of rights and access to justice, the authoritarian undermining of the rule of law, the theory of law as challenged by the intertwined character of transnational normative legalities (inter-legality).



**Massimo Bergamasco** is Professor of Theory of Mechanisms and Machines at the Scuola Superiore Sant'Anna, Pisa, Italy. At present he is the President of the ARTES 4.0 (Advanced Robotics and enabling digital Technologies & Systems 4.0) Competence Center in the framework of the Industry 4.0 of the Italian Ministry for Economic Development. He has been the founder of the Perceptual Robotics Laboratory. His research activity deals with the study and development of haptic interfaces and wearable robots for the control of the interaction between humans

and Virtual Environments. His present research is focused on general aspects of perception and cognitive processes in the field of embodiment and humanoid robotics.

**Virtual Humans**\_ The concept of physical space has been recently juxtaposed to other digital entities like Augmented or Mixed Realities. Presumably, such new spaces will be occupied not only by digital objects but also by the digital equivalent of human beings, i.e. virtual humans. How present research in robotics and AI is going to define the behaviour of humanoid robots, androids and virtual humans is something that can be formulated also in terms of ethical and social terms. The presentation provides a general picture on how the interaction between human beings and artificial entities can be foreseen on the basis of present technology development.



**Erica Palmerini** is Associate Professor of Private Law. A graduate in Law at the Pisa University, and a PhD in Private Law at the Scuola Sant'Anna, her main research interests cover law and technology, medical law, tort law and bioethics. She coordinated the two-year European (FP7) project RoboLaw ([www.robolaw.eu](http://www.robolaw.eu)) – a World Technology Award 2013 winner – on the ethical, social and legal implications of robotic technologies, and published on issues at the intersection of law and technology, and on technologies' regulation. She carried other research projects

funded by the Italian Ministry of Science and Education (Genetic identity and disabilities: the protection against discrimination in civil law”, 2007-2009) and by private institutions (Droit – Droni in Toscana, Fondazione Cassa di Risparmio di Firenze, 2015-2016). She is member of the Association of Italian Civil Lawyers and the European Law Institute, where she partakes in the Special Interest Group on Digital law.

**Social care robots and the issue of legal personhood** \_ This talk focuses on the hypothesis to recognize legal personhood to robots, and especially to robots involved in health and social care. Regulatory proposal and doctrinal suggestions in this direction will be examined and discussed with regard to their underlying rationale, the conceptual significance of such an acknowledgment and its potential practical consequences.



**Antonio Chella** is full professor of robotics at the University of Palermo where he directs the Robotics Laboratory he founded in 1997. He actively collaborates with ICAR-CNR, Palermo. He coordinated several projects in Social Robotics, including Cicerobot (a museum robot guide); Robotics and Autism; Robotics and ALS; RoboDanza; Robot Orchestra Conductor. In 2017 he was awarded the "James Albus Medal" by the BICA Scientific Society. He is a member of the Academy of Sciences, Letters and Arts of Palermo. His research concerns the study of consciousness in robots and machines. He is the author of more than 200 international publications.

**Experiments in Trustworthy Human-Robot Interactions** \_ The presentation will discuss the problems related to how we can correctly trust a robot when the robot itself is a part of our team. A conceptual model of trust and the central psychological studies related to the issue of robot trust will be briefly reviewed. Then, the talk will present experiments carried out at the RoboticsLab concerning the role of some high-level cognitive capabilities that would be needed for a robot to establish trustworthy human-robot interactions. The robot self-monitoring capability, the ability for a robot to present an inner speech, and the capacity of the robot to generate a theory of mind of the user will be outlined. Finally, an experiment related to a cheating robot will be presented.



**Tobias Mahler** teaches law at the faculty of law at the University of Oslo. He is a professor at the Norwegian Research Center for Computers and Law (NRCCL). His research interests cover a broad range of legal issues arising in the context of (i) robots, particularly with artificial intelligence capabilities, (ii) Internet governance (especially the domain name system), as well as (iii) cybersecurity and privacy. He teaches primarily robot regulation, cybersecurity regulation, legal tech and artificial intelligence, as well as Norwegian and German law of obligations. He is the deputy director of the NRCCL and the director of the centre's LLM programme.

**Human-robot impact assessments**\_ An impact assessment (IA) is a structured process for considering the implications of proposed actions. IAs can support decision-making about actions at a variety of levels, such as when adopting new policies or contemplating projects. Impact assessments have found their way into the regulatory toolbox, e.g. in the context of data protection IAs (DPIAs). A human-robot IA (HRIA) could extend DPIAs with additional elements, to take into account interests such as human dignity and autonomy, which go beyond data protection. This talk discusses open questions regarding proposals for HRIA. What would be an appropriate methodology for HRIA? DPIAs usually have a strong focus on risk, but the distinction between IA and risk assessment methods is often blurry. Should there be an obligation to assess the impact of some forms of human-robot interaction? If yes, who should be the addressee of such a requirement, and how would it be delimited?



## “EURA YOUNG SCHOLARS PRIZE” PAPERS

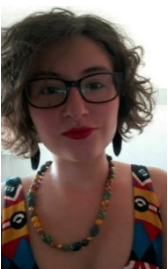


**Eduard Fosch-Villaronga** is a Marie Skłodowska-Curie Postdoctoral Researcher at the eLaw Center for Law and Digital Technologies at Leiden University and the co-leader of the Ethical, Legal, and Societal Aspects Working Group at the H2020 Cost Action 16116 on Wearable Robots. He investigates legal and regulatory aspects of robots and AI, with a special focus on healthcare, and is interested in human-robot interaction, responsible innovation, and the future of law. He worked as a Post doc at the Microsoft Cloud Computing Research Center (Queen Mary University of London) and the University of Twente. He holds an Erasmus Mundus Joint Doctorate in Law, Science, and Technology coordinated from University of Bologna, an LL.M. from University of Toulouse, an M.A. from the Autonomous University of Madrid, and an LL.B. from the Autonomous University of Barcelona.



**Lexo Zardiashvili** is a Ph.D. candidate at the eLaw Center for Law and Digital Technologies at Leiden University (NL) conducting research on Dignity and Artificial Intelligence. Lexo has also worked on the project ‘SCALES: Designing a Regulatory and Institutional Framework to Balance Public Interests and Individual Liberties in the Use of Data Analytics’ covering ‘Responsible use of Artificial Intelligence in the Dutch Police’. Previously, he was an Anti-Money Laundering Specialist at the National Bank of Georgia, and consultant at Compliance Development Center also in Georgia. Lexo holds an LL.M. *cum laude* from Leiden University (2018, NL) where he was part of the Advanced Master’s Degree in Law and Digital Technologies, and an LL.B. from Caucasus Law School (2014, GE).

**Exploring the use and development of healthcare robot technology through the lenses of ‘human dignity’** \_ Healthcare robots enable practices that seemed far-fetched in the past, such as supporting medical interventions, helping people in their neurorehabilitation process, or their activities of daily life. Such practices raise the question to what extent society will allow developments in robots to influence human nature. In this presentation, we explore the concept of human dignity as a way to address the impacts associated with the use of robots for healthcare purposes.



**Irene Sucameli** is a Ph.D. student in Computer Science at the University of Pisa (IT). Her research interests focus on Natural Language Processing and Understanding for conversational agents and Artificial Intelligence. She is a member of AILC (Associazione Italiana di Linguistica Computazionale) and, in 2018, she received her M.Sc. in Digital Humanities from the University of Pisa with a thesis correlated to the field of computational linguistics. Her present research is related to the development of a conversational agent for the Italian language and to the study of the social and ethical issues connected with the use of dialog systems.

**How to improve the level of trust in human-machine conversation** \_ The aim of this paper is to contribute to the debate about ethics in AI focusing particularly on the social and ethical issues which emerge during conversational agent-human interaction. Those issues will be approached adopting both the perspective of the user and the one of the machine. Finally, it will be proposed a short set of moral guidelines that may produce a social beneficial impact and increase the level of trust within the human-computer interaction.