Ethical AI Challenges Workshop Programme
Friday, July 19, 9:00AM-1:00PM, Room: Sofitel Bellevue 2

9.20 – Welcome
9.30 – 10:30 Keynote and discussion, **Professor Jim Torresen, University of Oslo, Norway**
10:30 – 11:30 Keynote and discussion**, Professor Nigel Crook, Oxford Brookes University**
11.30 – 11:45 coffee break
11.45 – 12:45 Keynote and discussion**, Dr Matthias Rolf, Oxford Brookes University**

Key note presentation by

Professor Jim Torresen, University of Oslo, Norway

Jim Torresen is a professor at University of Oslo where he leads the Robotics and Intelligent Systems research group. He received his M.Sc. and Dr.ing. (Ph.D) degrees in computer architecture and design from the Norwegian University of Science and Technology, University of Trondheim in 1991 and 1996, respectively. He has been employed as a senior hardware designer at NERA Telecommunications (1996-1998) and at Navia Aviation (1998-1999). Since 1999, he has been a professor at the Department of Informatics at the University of Oslo (associate professor 1999-2005). Jim Torresen has been a visiting researcher at Kyoto University, Japan for one year (1993-1994), four months at Electrotechnical laboratory, Tsukuba, Japan (1997 and 2000) and a visiting professor at Cornell University, USA for one year (2010-2011).

His research interests at the moment include artificial intelligence, machine learning, reconfigurable hardware, robotics and applying this to complex real-world applications. Several novel methods have been proposed. He has published approximately 150 scientific papers in international journals, books and conference proceedings. 10 tutorials and a number of invited talks have been given at international conferences and research institutes. He is in the program committee of more than ten different international conferences, associate editor of three international scientific journals as well as a regular reviewer of a number of other international journals. He has also acted as an evaluator for proposals in EU FP7 and Horizon2020 and is currently project manager/principle investigator in five externally funded research projects/centres.

More information and a list of publications can be found here:
http://www.ifi.uio.no/~jimtoer

Title: Introduction to Different Challenges in Ethical AI and Possible Ways of Addressing Them

Summary:

Robots and artificial intelligence demonstrate to effectively contribute within an increasing number of different domains. At the same time, an increasing number of people – in the general public as well as in research – have started to consider a number of potential ethical challenges related to the development and use of such technology. This talk will give an overview of the most commonly expressed ones and ways being undertaken to reduce their impact using the findings in an earlier undertaken review. https://www.frontiersin.org/articles/10.3389/frobt.2017.00075/full (add link only if the hyperlink in the text don´t work)

Among the most important challenges are those related to privacy, safety and security. We are currently undertaking research in various projects where the challenges appear like in robots for elderly at home and mental health care technology. Robots would in the future be operating in closer interaction and collaboration with humans resulting in new technical and ethical research challenges to be addressed. This talk will introduce some examples from our work and how we address it both from a technical and human side.

Keynote Presentation by Professor Nigel Crook, Oxford Brookes University

Nigel Crook is professor in Artificial Intelligence and the Associate Dean Research and Knowledge Exchange at the Faculty of Technology, Design and Environment, Oxford Brookes University. He joined Oxford Brookes as a PhD student in 1985. Nigel studied for his undergraduate degree at Lancaster University, initially doing Maths and Philosophy before switching to Computing and Philosophy, being particularly attracted to the logic side of Philosophy. He moved to - as it was then- Oxford Polytechnic in 1985 to undertake a PhD in Medical Diagnostics Systems and upon completion in 1989, became a lecturer.

Title: Developing Robots with Moral Competence

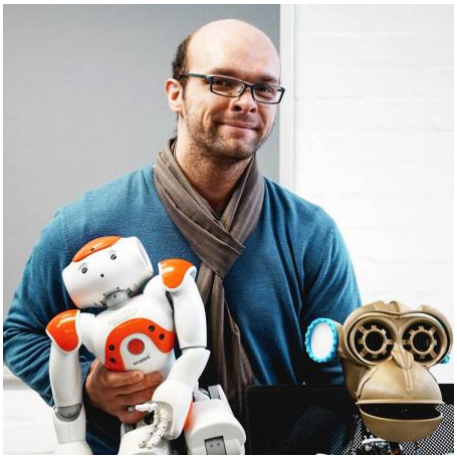
Summary:

The recent rapid and widespread deployment of AI and robotic systems across a broad range of application domains has raised considerable ethical concern in both public and academic arenas. This concern ranges from fears about the ethical consequences of creating so-called ‘super- intelligence’ to anxiety about the ability of autonomous vehicles to make the ‘right’ moral choice of who to kill in the conventional ‘trolley problem’ scenario. In this talk I will explore the motivations for developing machines with moral competence and

outline some current approaches. I will present a case study enabling an autonomous vehicle to recognise the good and bad driving behaviours of other vehicles. I will conclude by discussing how moral machines can be evaluated.

Keynote Presentation by Dr Matthias Rolf, Oxford Brookes University

Matthias Rolf is Senior Lecturer in Computing and Robotics at Oxford Brookes University, UK since 2016. He previously was Specially Appointed Researcher and Assistant Professor at Osaka University, Japan (2013-2016) and research assistant at Bielefeld University (2008-2013) where he obtained his PhD (with highest honors) in 2012.
His research interest spans robotics, machine learning, software engineering, and cognitive science, with a particular focus on developmental robotics. His work on "goal babbling" as an efficient way of sensorimotor learning lead to several research awards and international news coverage. His recent research focuses on machine learning approaches to agents' autonomous learning of own goals as well as ethical AI.



Title:
From social interaction to ethical AI: a developmental roadmap

Summary:

AI and robot ethics have recently gained a lot of attention because adaptive machines are increasingly involved in ethically sensitive scenarios and cause incidents of public outcry. Much of the debate has been focused on achieving highest moral standards in handling ethical dilemmas on which not even humans can agree, which indicates that the wrong questions are being asked.
While traditionally engineered artifacts, including AI, require the designer to ensure ethical compliance, learning machines that change through interaction with people after their deployment can not be vetted in just the same way. I will argue that in order to progress on this issue, we need to look at it strictly through the lens of what behavior seems socially acceptable, rather than idealistically ethical. Machines would then need to determine what behavior is compliant with social and moral norms, and therefore be receptive to social feedback from people.
I will discuss a roadmap of computational and experimental questions to address the development of socially acceptable machines, and emphasize the need for social reward mechanisms and learning architectures that integrate these while reaching beyond limitations of traditional reinforcement-learning agents. I suggest to use the metaphor of

“needs” to bridge rewards and higher level abstractions such as goals for both communication and action generation in a social context. We then suggest a series of experimental questions and possible platforms and paradigms to guide future research in the area.