

2018 Uehiro-Carnegie-Oxford Conference
on
“Ethics and the Future of Artificial Intelligence”
May 16-18, 2018

Venue: Carnegie Council, 170 East 64th Street, New York City

Organizers: The Uehiro Foundation on Ethics and Education, Tokyo
Carnegie Council for Ethics in International Affairs, New York
Oxford Uehiro Centre for Practical Ethics, Oxford University

Background

As we enter an era set apart by unprecedented levels of automation and sophisticated artificial intelligence, it is essential that we come together to consider the ethical values and consequences at stake in this evolving aspect of human society.

From factory production lines to mail delivery services, businesses are requiring ever fewer workers to maintain and increase their productivity as a rising number of tasks are automated.

The scale and speed of the current automation trend has many concerned that the labor market will be unable to adapt to this structural shift. Yet responsive public policies remain controversial and elusive.

In addition to labor concerns, advancements in artificial intelligence and robotics could soon lead to an assortment of new actors exhibiting human-like qualities. As consumer electronics begin incorporating deep-learning systems meant to mimic human neural networks, the scope of AI will continue to expand. Computers could come to recognize human emotions, to speak in colloquial speech patterns, to interpret and create art.

What are the challenges and opportunities associated with automation and AI advancements? Are societies obligated to insure equitable public benefit from these technologies?

In such a future, could an artificial being be considered a moral agent? If so, could they be held to the same ethical standards as humans? If software could emulate moral codes, which values would be included?

Format

Each session will be a panel discussion. The goal is to facilitate an exchange of views rather than to present research papers.

Wednesday, May 16, 2018

6:00pm **Welcome Dinner** at the Yale Club of New York, Saybrook Room, 50 Vanderbilt Avenue between 44th and 45th Streets

Thursday, May 17, 2018

8:30AM Breakfast Available

9:00AM Opening Remarks
Joel Rosenthal, President, Carnegie Council for Ethics in International Affairs
Goro Takaguchi, Managing Director, The Uehiro Foundation on Ethics and Education

9:30AM Session I: Moral Machines
Chaired by Joel Rosenthal

Can machines be programmed with AI to make moral choices? For example, how do we think about responsibility in the driverless car, or the automated financial services transaction? Can AI be used to improve human moral decision-making?

- **Wendell Wallach**, Interdisciplinary Center for Bioethics, Yale University on “Moral Machines: From Machine Ethics to Value Alignment ”
- **Julian Savulescu**, Oxford Uehiro Centre for Practical Ethics on “Moral Artificial Intelligence”
- **Fumio Shimo**, Keio University on “Towards Establishing the Basic Principles for Underpinning AI and Robot Use”

10:30AM Coffee Break

11:15AM - **Carissa Veliz**, Oxford Uehiro Centre for Practical Ethics on “What would it take for an AI to be a moral agent?”
- **Kush Varshney**, IBM Research AI on “On the Safety of Machine Learning: Cyber-Physical Systems, Decision Sciences, and Data Products”

12:30PM **Lunch**

1:30PM Session II: AI, Robotics, and the Future of Work and Life
Chaired by Julian Savulescu

How will increasing automation affect the human need and expectation for meaningful work? How will AI change work and life? Is AI friend or foe?

- **Arisa Ema**, University of Tokyo and Visiting Researcher at RIKEN Center for Advanced Intelligence Project in Japan on “Designing the Workplace with AI and Robots”
- **Yoshinori Hiroi**, Kokoro Research Center, Kyoto University on “AI, Public Policy and Aging Society”

2:15PM Coffee Break

3:00PM - **Hannah Maslen**, Oxford Uehiro Centre for Practical Ethics on “AI, Predictive Policing and Preparatory Offences”
- **Julia Powles**, New York University School of Law and Cornell Tech, on “AI and Power”

4:30PM Live Demonstration from a Robotics Lab (TBD)

5:30PM Adjourn

6:00PM Reception & Dinner at Carnegie Council

Friday, May 18

9:00AM Breakfast Available

Session III: Living with Robots
Chaired by Wendell Wallach

Can robots engage with human beings as tutors, healthcare providers, and companions?
What are the possibilities and limitations of AI in this area?

- **Go Yoshizawa**, Associate Professor of Biomedical Ethics and Public Policy at Graduate School of Medicine, Osaka University on “Dark Science and Positive Engagement”
- **Ken Tomiyama**, Chiba Institute of Technology on “Virtual Emotion for Robot - Towards Human Support Robot”
- **Robert Sparrow**, Monash University on “What Robots Represent and Why it Matters”

11:00AM Session IV: Drones and Killer Robots
Chaired by Joel Rosenthal

How is AI-equipped technology changing long-standing norms of public security?

- **Peter Asaro**, co-founder of the International Committee for Robot Arms Control on “Automating Violent Force: From Battlefields to Policing”
- **William D. Casebeer**, Senior Research Area Manager, Human Systems and Autonomy, Lockheed Martin Advanced Technology Laboratories on “Building an Artificial Conscience: Prospects for Morally Autonomous AI”
- **Angela Kane**, former UN High Representative for Disarmament Affairs on “Lethal Autonomous Weapons Systems: Can the International Community Agree on an Approach?”

1:00PM Lunch and Concluding Remarks by **Julian Savulescu**, Oxford Uehiro Centre for Practical Ethics