HUMAN RIGHTS, ETHICS & ARTIFICIAL INTELLIGENCE

CHALLENGES FOR THE NEXT 70 YEARS OF THE UNIVERSAL DECLARATION OF HUMAN RIGHTS

November 29, to
December 1, 2018
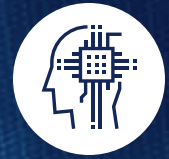
HARVARD Kennedy School
CARR CENTER
for Human Rights Policy

HARVARD UNIVERSITY
Edmond J. Safra Center for Ethics

BERKMAN
KLEIN CENTER
FOR INTERNET & SOCIETY
AT HARVARD UNIVERSITY

HUMAN RIGHTS, ETHICS & ARTIFICIAL INTELLIGENCE: CHALLENGES FOR THE NEXT 70 YEARS OF THE UNIVERSAL DECLARATION OF HUMAN RIGHTS

The Context

December 10, 2018, marks the 70th anniversary of the Universal Declaration of Human Rights. The human rights movement with its moral, legal and organizational ramifications has been one of the great innovations of the 20th century. Much has been achieved, but the world continues to fall short of human rights ideals in numerous ways. Now is also the time to look ahead to the next 70 years, an era likely marked by technological breakthroughs at an unprecedented scale. Artificial Intelligence is increasingly present in our lives, reflecting a growing tendency to turn for advice or turn over decisions altogether, to algorithms. The effectiveness of algorithms is increasingly enhanced through Big Data: the availability of an enormous amount of data on all human activity. The key idea behind human rights is that power is used in ways that protect and advance basic human interests, especially those of the most vulnerable and that there is a global responsibility to make sure of this. New technology greatly helps with these goals, but there are also dangers. In the short term, these dangers range from the perpetuation of bias in algorithmic reasoning to use of technology to create "alternative facts." In the longer term, we must worry about ownership of data, increasing power of tech companies, changes in the nature of work and perhaps eventually existential risk through the advent of entities that are not alive in familiar ways but might be sentient and intellectually and even morally superior to humans. This conference explores ways in which technological change will affect human life, especially the rights designed to protect that life, in the next 70 years and beyond.

In the larger picture, this conference stands at the intersection of closely connected emerging agendas at the Carr Center and the EJ Safra Center, as well as at the intersection of broader interests around Harvard in benefits and challenges from technological breakthroughs and artificial intelligence in particular. Alongside other human rights centers at Harvard, Carr responds to the Universal Declaration of Human Right's anniversary by reflecting on the past, present, and future of human rights. The bigger question behind that is how the institutional protection of the distinctive human life can be assured in an era of exploding technological innovation. EJ Safra seeks to explore how such innovation engages human values and a broad range of ethical issues that arise from our ways of using technology and perhaps eventually from sharing our lives with forms of technology that for now are merely imaginable. At stake is nothing less than the future of human life and its organizational possibilities, what that life will look like, and whether it might eventually be threatened at an existential level by its own innovations. Many initiatives are emerging in this domain at Harvard. It is time to bring them together and to make sure the ethical dimensions of these changes take center stage in Harvard's agenda.

Day 1 & 2

5:00 PM - 6:30 PM
Opening Keynote

6:30 PM - 9:00 PM
Reception & Dinner

8:15 AM - 9:00 AM
Registration & Light Breakfast

9:00 AM - 10:15 AM
From Ethical Challenges of Intelligent Systems to
Embedding Ethics in Computer Science Education

10:15 AM - 10:30 AM
Break

10:30 AM - 11:45 AM
Accounting for Artificial Intelligence: Rules, Reasons,
Rationales

11:45 AM - 12:00 PM
Break

12:00 PM - 1:15 PM
The End of Human Exceptionalism?

1:15 PM - 2:15 PM
Lunch

2:15 PM - 3:15 PM
Flash Presentations by Members
of the Inaugural Techtopia
Group

3:15 PM - 3:30 PM
Break

3:30 PM - 4:45 PM
How to Get Empowered, Not
Overpowered, by Artificial Intelligence

4:45 PM - 5:00 PM
Break

5:00 PM - 6:15 PM
What's Worse: Machine Learning That Doesn't
Work, Or Machine Learning That Does?

Day 3

9:00 AM - 10:00 AM
Light Breakfast

10:00 AM - 11:15 AM
Transcending the Brain? AI,
Radical Brain Enhancement, and
the Nature of Consciousness

11:15 AM - 11:30 AM
Break

11:30 AM - 12:45 PM
Human Rights and Artificial
Intelligence: The Long
(Worrisome) View?

12:45 PM - 2:00 PM
Lunch

2:00 PM - 3:15 PM
Panel Discussion: Artificial
Intelligence and Human
Rights Law in the Here and
Now

3:15 PM - 3:30 PM
Break

3:30 PM - 4:45 PM
Contextualizing Calls for AI
Transparency and Fairness

4:45 PM - 5:00 PM
Break

5:00 PM - 6:15 PM
Responsible Business in the Digital
Economy: What's New

6:15 PM - 6:30 PM
Closing Remarks

HUMAN RIGHTS, ETHICS & ARTIFICIAL INTELLIGENCE: CHALLENGES FOR THE NEXT 70 YEARS OF THE UNIVERSAL DECLARATION OF HUMAN RIGHTS

Thursday, November 29, 5:00 PM – 6:30 PM

Speaker
- **Cynthia Dwork,** Gordon McKay Professor of Computer Science, John A. Paulson School of Engineering and Applied Sciences, Harvard University

SKEWED OR RESCUED?: THE EMERGING THEORY OF ALGORITHMIC FAIRNESS

Opening Keynote Address
Taubman Nye A, B, C

This talk also serves as the inaugural talk for the conference, followed by a dinner in Wexner Winter Garden.

6:30 – 9:00 PM
Wexner Winter Garden

DINNER & RECEPTION

End of Day 1

HUMAN RIGHTS, ETHICS & ARTIFICIAL INTELLIGENCE: CHALLENGES FOR THE NEXT 70 YEARS OF THE UNIVERSAL DECLARATION OF HUMAN RIGHTS

Friday, November 30, 8:15 AM – 9:00 AM

Rubenstein 414AB

REGISTRATION & LIGHT BREAKFAST

Friday, November 30, 9:00 AM – 10:15 AM

Speaker

- **Barbara Grosz,** Higgins Professor of Natural Sciences, John A. Paulson School of Engineering and Applied Sciences, Harvard University

FROM ETHICAL CHALLENGES OF INTELLIGENT SYSTEMS TO EMBEDDING ETHICS IN COMPUTER SCIENCE EDUCATION

First Session
Wexner 434AB

10:15 AM – 10:30 AM

BREAK

10:30 AM – 11:45 AM

Speaker

- **Solon Barocas,** Assistant Professor, Department of Information Science, Cornell University

ACCOUNTING FOR ARTIFICIAL INTELLIGENCE: RULES, REASONS, RATIONALES

Second Session
Wexner 434AB

HUMAN RIGHTS, ETHICS & ARTIFICIAL INTELLIGENCE: CHALLENGES FOR THE NEXT 70 YEARS OF THE UNIVERSAL DECLARATION OF HUMAN RIGHTS

11:45 AM – 12:00 PM

BREAK

12:00 PM – 1:15 PM

Speaker
- **Stephen Wolfram,** Founder and CEO, Wolfram Research

THE END OF HUMAN EXCEPTIONALISM?

Third Session
Wexner 434 AB

1:15 PM – 2:15 PM

LUNCH

2:15 PM – 3:15 PM

Panelists
- **Hannah Hilligoss | Moderator,** Project Coordinator, Cyberlaw Clinic, Berkman Klein Center for Internet & Society, Harvard Law School
- **Jenny Fan**, Master in Design Engineering, 2019, Harvard Graduate School of Design
- **Luis Valles,** Master in Business Administration, 2020, Harvard Business School
- **Joshua Simons**, PhD in Government, 2022, Harvard Graduate School of Arts and Sciences
- **Irene Solaiman,** Master in Public Policy, 2019, Harvard Kennedy School

FLASH PRESENTATIONS BY INAUGURAL TECHTOPIA GROUP

Fourth Session
Wexner 434 AB

HUMAN RIGHTS, ETHICS & ARTIFICIAL INTELLIGENCE: CHALLENGES FOR THE NEXT 70 YEARS OF THE UNIVERSAL DECLARATION OF HUMAN RIGHTS

3:15 PM – 3:30 PM

BREAK

3:30 PM – 4:45 PM

Speaker
- **Max Tegmark,** Professor of Physics, Massachusetts Institute of Technology

HOW TO GET EMPOWERED, NOT OVERPOWERED, BY ARTIFICIAL INTELLIGENCE

Fifth Session
Wexner 434 AB

4:45 PM – 5:00 PM

BREAK

5:00 PM – 6:15 PM

Speaker
- **Jonathan Zittrain,** George Bemis Professor of International Law, Harvard Law School, Harvard Kennedy School, Faculty Director, Berkman Klein Center for Internet & Society

WHAT'S WORSE: MACHINE LEARNING THAT DOESN'T WORK, OR MACHINE LEARNING THAT DOES?

Sixth Session
Wexner 434 AB

6:30 PM – 9:00 PM

PRIVATE RECEPTION & DINNER

End of Day 2

HUMAN RIGHTS, ETHICS & ARTIFICIAL INTELLIGENCE: CHALLENGES FOR THE NEXT 70 YEARS OF THE UNIVERSAL DECLARATION OF HUMAN RIGHTS

Saturday, December 1, 9:00 AM – 10:00 AM
Rubenstein 414 AB

LIGHT BREAKFAST

10:00 AM – 11:15 AM

Speaker
- **Susan Schneider,** Director, AI, Mind and Society Group, The University of Connecticut, Distinguished Scholar, US Library of Congress, Institute for Advanced Study, Princeton, New Jersey

TRANSCENDING THE BRAIN? AI, RADICAL BRAIN ENHANCEMENT, AND THE NATURE OF CONSCIOUSNESS

First Session
Wexner 434 AB

11:15 AM – 11:30 AM

BREAK

11:30 AM – 12:45 PM

Speaker
- **Mathias Risse,** Lucius N. Littauer Professor of Philosophy and Public Administration, Harvard Kennedy School, Faculty Director, Carr Center for Human Rights Policy

HUMAN RIGHTS AND ARTIFICIAL INTELLIGENCE: THE LONG (WORRISOME?) VIEW

Second Session
Wexner 434 AB

#AIAtHKS

HUMAN RIGHTS, ETHICS & ARTIFICIAL INTELLIGENCE: CHALLENGES FOR THE NEXT 70 YEARS OF THE UNIVERSAL DECLARATION OF HUMAN RIGHTS

12:45 – 2:00 PM

LUNCH

2:00 PM – 3:15 PM

Panelists
- **Vivek Krishnamurthy | Moderator,** Counsel, Foley Hoag LLP, Affiliate, Berkman Klein Center for Internet & Society, Harvard Law School, Senior Fellow, Carr Center for Human Rights Policy
- **Michael Karimian,** Human Rights Program Manager, Microsoft
- **Amy Lehr,** Director of Human Rights, Center for Strategic and International Studies
- **Nani Jansen Raventlow,** Director, Digital Freedom Fund
- **Can Yeginsu**, Barrister, 4 New Square Chambers and Columbia Law School

ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS LAW IN THE HERE AND NOW

Panel Discussion
Wexner 434 AB

3:15 PM – 3:30 PM

BREAK

3:30 PM – 4:45 PM

Speakers
- **John Basl,** Assistant Professor of Philosophy, Northeastern University
- **Ronald Sandler,** Chair and Professor of Philosophy, Director, Ethics Institute, Northeastern University

CONTEXTUALIZING CALLS FOR AI TRANSPARENCY AND FAIRNESS

Third Session
Wexner 434 AB

HUMAN RIGHTS, ETHICS & ARTIFICIAL INTELLIGENCE: CHALLENGES FOR THE NEXT 70 YEARS OF THE UNIVERSAL DECLARATION OF HUMAN RIGHTS

4:45 PM – 5:00 PM

BREAK

5:00 PM – 6:15 PM

Speaker
- **Nien-Hê Hsieh,** Associate Professor of Business Administration, Harvard Business School

RESPONSIBLE BUSINESS IN THE DIGITAL ECONOMY: WHAT'S NEW?

Fourth Session
Wexner 434 AB

6:15 PM – 6:30 PM

CLOSING REMARKS

End of Day 3