



INTERNATIONAL CONFERENCE

EXISTENTIAL THREATS AND OTHER DISASTERS

HOW SHOULD WE ADDRESS THEM

Mediteran Hotel & Resort • Budva • Montenegro

May 30-31st 2024

SPEAKERS:

Vojin Rakić /The Center for the Study of Bioethics
Vardit Ravitsky /The Hastings Center
Roger Crisp /The Oxford Uehiro Centre for Practical Ethics
Peter Singer /Princeton University
Julian Saulescu /National University of Singapore
Arthur Caplan /New York University's Langone Medical Center
Josephine Johnston /The Hastings Center
Ingmar Persson /University of Gothenburg
Anders Sandberg /Oxford University's Future of Humanity Institute
Nicholas Agar /University of Waikato
James Hughes /University of Massachusetts

Keynote speakers:

- **Peter Singer:** “Existential Threats and Other Disasters: The Need for a Consequentialist Approach”
- **Julian Savulescu:** “AI as Human Moral Enhancement”
- **Vojin Rakić:** "Extraterrestrial Forms of Cognition and Morality: Existential Threats or Opportunities"
- **Vardit Ravitsky:** “Bioethics’ Capacity to Address Existential Threats: A Path Forward”
- **Roger Crisp:** “Human Extinction: Possibly Good?”
- **Ingmar Persson:** Is Pessimism about the Future of Humanity Justifiable?
- **Anders Sandberg:** “Civilizational Virtue, Civilizational Autonomy, and Existential Risks”
- **Arthur Caplan:** “The Eradication of Infectious Diseases and the Threat of Future Outbreaks in a World Brimming with Conflicts”
- **Josephine Johnston:** “Reframing Low Birthrates as an Existential Opportunity”
- **Slobodan Perović:** "Exponential Appetite for Energy: An Existential Risk Or Necessity of Life?"
- **Nicholas Agar:** “The Big Problem with Charismatic Extinction Threats”
- **James Hughes:** “Are Children The Future?: Longtermism, Epistemic Discounting and Axial Futurism”

AUTHOR

Marina Budić, Marko Galjak

INSTITUTIONAL AFFILIATION AND ADDRESS

Research Assistant at the Institute of Social Sciences,
Kraljice Natalije 45,
Belgrade, Serbia

TITLE OF THE PRESENTATION

Artificial Intelligence and Existential Risk: Deontological and Consequentialist Approach

ABSTRACT (MAX. 250 WORDS)

As we edge closer to the development of AGI, the potential for a superintelligent entity that surpasses human intelligence becomes real. While AGI promises innovation and progress, it also introduces existential risk. The argument that AGI constitutes an existential risk for the human species rests on two premises: (1) the Singularity claim that states AI may reach superintelligent levels, at which point humans lose control; (2) the Orthogonality thesis that states any level of intelligence can go along with any goal, i.e., intelligence doesn't necessarily correlate with benevolent goals. This presents a quandary: how does humanity proceed if an AGI's goals aren't aligned with ours? AGI's potential necessitates ethical consideration. We explore how deontological and consequentialist ethical theories can be used as a prism to view the existential risk of AI. We examine these ethical approaches in answering whether we should take that risk. This presentation delves into the ethical challenges AGI poses, exploring value alignment, control problems, and societal impacts. Drawing upon insights from the fields of AI ethics and normative ethics, we evaluate specific responses to the risk at hand: the alignment issue, the morality of boxing solutions – questioning the constraints placed on AI and its entitlement to rights and exploring governance structures, including possible regulatory interventions or stringent oversight mechanisms to curb unchecked AGI evolution, or total surveillance to prevent further AGI developments. We answer these questions from perspectives of deontological and consequentialist approaches in the realm of AGI and superintelligence, emphasizing the depth and complexity of ethical considerations in this field.