



philosophies



an Open Access Journal by MDPI

The Perils of Artificial Intelligence

Guest Editor:

Dr. Roman V. Yampolskiy

Speed School of Engineering,
University of Louisville, Louisville,
KY 40292, USA

roman.yampolskiy@
louisville.edu

Deadline for manuscript
submissions:

1 December 2021

Message from the Guest Editor

The debate continues as to what degree of independence entities like robotics and AI (Artificial Intelligence) should have and how to make sure that they do not turn on their inventors. This Special Issue will be comprised of papers addressing different aspects of the AI control problem as it relates to the development of safe and secure AI and philosophical issues around this topic. We solicit papers ranging from issues like bias, transparency, and fairness to concerns such as value alignment, controllability, predictability, and explainability of superintelligent systems. We also welcome manuscripts on non-traditional and contrarian views of AI safety and security. We are interested in any aspects of AI safety and security, including: adversarial machine learning (ML); AI failures; bad actors/malevolence; boxing/confinement; controllability; drives/goals; ethics/law; explainability; friendliness theory; openness/secretcy; off-switch; predictability; reward engineering; self-improvement; security; singularity/superintelligence; validation/verification; value alignment; weaponized/military; wireheading

Dr. Roman V. Yampolskiy, *Guest Editor*



mdpi.com/si/57122

Special Issue



Editor-in-Chief

Prof. Dr. Marcin J. Schroeder

Global Learning Center, IEHE
(Institute for Excellence in Higher
Education), Tohoku University,
Sendai 980-8576, Japan

Message from the Editor-in-Chief

Philosophies searches for the syntheses of philosophical and scientific inquiries. It promotes philosophical work that is derived from the experience of diverse scientific disciplines and cultures. Multiple philosophies already exist — those of logic, information, computation, natural and artificial life, natural or artificial intelligence, complexity, technology, etc. Our mission is not to abandon philosophical roots and traditions of inquiry, but to promote the development of philosophical foundations and effective methodologies derived from diverse scientific explorations, and intended to enhance these explorations as to generate deeper and more holistic knowledge. Innovation may also be achieved through the cultural dimension. Other cultures can offer from their heritage a diversity of resources for exploration; these resources can also contribute to the emergent synthesis of philosophical inquiry.

Author Benefits

Open Access:— free for readers, with article processing charges (APC) paid by authors or their institutions.

High Visibility: indexed within Scopus, ESCI (Web of Science), PhilPapers, and many other databases.

Rapid Publication: manuscripts are peer-reviewed and a first decision provided to authors approximately 20.4 days after submission; acceptance to publication is undertaken in 6.2 days (median values for papers published in this journal in the first half of 2021).

Contact Us

Philosophies
MDPI, St. Alban-Anlage 66
4052 Basel, Switzerland

Tel: +41 61 683 77 34
Fax: +41 61 302 89 18
www.mdpi.com

mdpi.com/journal/philosophies
philosophies@mdpi.com