Studies in Applied Philosophy, Epistemology and Rational Ethics

Volume 63

Editor-in-Chief

Lorenzo Magnani, Department of Humanities, Philosophy Section, University of Pavia, Pavia, Italy

Editorial Board

Atocha Aliseda, Universidad Nacional Autónoma de México (UNAM), Mexico, Mexico

Giuseppe Longo, CNRS - Ecole Normale Supérieure, Centre Cavailles, Paris, France

Chris Sinha, School of Foreign Languages, Hunan University, Changsha, China

Paul Thagard, University of Waterloo, Waterloo, Canada

John Woods, University of British Columbia, Vancouver, Canada

Studies in Applied Philosophy, Epistemology and Rational Ethics (SAPERE) publishes new developments and advances in all the fields of philosophy, epistemology, and ethics, bringing them together with a cluster of scientific disciplines and technological outcomes: ranging from computer science to life sciences, from economics, law, and education to engineering, logic, and mathematics, from medicine to physics, human sciences, and politics. The series aims at covering all the challenging philosophical and ethical themes of contemporary society, making them appropriately applicable to contemporary theoretical and practical problems, impasses, controversies, and conflicts. Our scientific and technological era has offered "new" topics to all areas of philosophy and ethics–for instance concerning scientific rationality, creativity, human and artificial intelligence, social and folk epistemology, ordinary reasoning, cognitive niches and cultural evolution, ecological crisis, ecologically situated rationality, consciousness, freedom and responsibility, human identity and uniqueness, cooperation, altruism, intersubjectivity and empathy, spirituality, violence. The impact of such topics has been mainly undermined by contemporary cultural settings, whereas they should increase the demand of interdisciplinary applied knowledge and fresh and original understanding. In turn, traditional philosophical and ethical themes have been profoundly affected and transformed as well: they should be further examined as embedded and applied within their scientific and technological environments so to update their received and often old-fashioned disciplinary treatment and appeal. Applying philosophy individuates therefore a new research commitment for the 21st century, focused on the main problems of recent methodological, logical, epistemological, and cognitive aspects of modeling activities employed both in intellectual and scientific discovery, and in technological innovation, including the computational tools intertwined with such practices, to understand them in a wide and integrated perspective. **Studies in Applied Philosophy, Epistemology and Rational Ethics** means to demonstrate the contemporary practical relevance of this novel philosophical approach and thus to provide a home for monographs, lecture notes, selected contributions from specialized conferences and workshops as well as selected Ph.D. theses. The series welcomes contributions from philosophers as well as from scientists, engineers, and intellectuals interested in showing how applying philosophy can increase knowledge about our current world. Initial proposals can be sent to the Editor-in-Chief, Prof. Lorenzo Magnani, lmagnani@unipv.it:

- A short synopsis of the work or the introduction chapter
- The proposed Table of Contents
- The CV of the lead author(s).

For more information, please contact the Editor-in-Chief at lmagnani@unipv.it.

Indexed by SCOPUS, zbMATH, SCImago, DBLP.

All books published in the series are submitted for consideration in Web of Science.

Vincent C. Müller
Editor

Philosophy and Theory of Artificial Intelligence 2021

Springer

Editor
Vincent C. Müller
Friedrich-Alexander University
of Erlangen-Nuremberg (FAU)
Erlangen, Germany

ISSN 2192-6255 ISSN 2192-6263 (electronic)
Studies in Applied Philosophy, Epistemology and Rational Ethics
ISBN 978-3-031-09152-0 ISBN 978-3-031-09153-7 (eBook)
https://doi.org/10.1007/978-3-031-09153-7

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Dedicated to the memory of our colleague and friend
Ivica Crnkovic
(1955–2022)

Advisory Editors

Akinori Abe, Faculty of Letters, Chiba University, Chiba, Japan
Hanne Andersen, Centre for Science Studies, Aarhus University, Aarhus, Denmark
Otávio Bueno, Department of Philosophy, University of Miami, Coral Gables, FL, USA
Marcelo Dascal, Dept of Philosophy, Gilman Buildin, Tel Aviv University, Tel Aviv, Israel
Gordana Dodig Crnkovic, Dep. of Computer Science and Engineering, Chalmers University of Technology, Gothenburg, Sweden
Michel Ghins,Inst Supérieur de Philosophie, L3.06.01, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
Marcello Guarini, Department of Philosophy, Univeristy of Windsor, Windsor, ON, Canada
Ricardo Gudwin,Computer Eng. and Industrial Automation, State University of Campinas, Campinas, Brazil
Albrecht Heeffer, Sarton Ctr for the History of Science, Ghent University, Gent, Belgium
Gerhard Minnameier, Goethe-Universität Frankfurt am Main, Frankfurt, Hessen, Germany
Margaret C. Morrison, Trinity College, University of Toronto, Toronto, ON, Canada
Yukio Ohsawa, School of Engineering, The University of Tokyo, Tokyo, Japan
Sami Paavola, Faculty of Educational Sciences, University of Helsinki, HELSINKI, Finland
Woosuk Park, Humanities and Social Sciences, KAIST, Daejeon, Korea (Republic of)
Luís Moniz Pereira, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, CAPARICA, Portugal
Ahti-Veikko Pietarinen, Dept. of Philosophy, University of Helsinki, Helsinki, Finland
Demetris Portides, Dept of Classics & Philosophy, University of Cyprus, Nicosia, Cyprus
Athanassios Raftopoulos, Psychology, University of Cyprus, Nicosia, Cyprus

Gerhard Schurz, Dept of Philosophy, DCLPS, Heinrich Heine University, Düsseldorf, Nordrhein-Westfalen, Germany
Cameron Shelley, Centre for Society, Technology & Values, University of Waterloo, Waterloo, ON, Canada
Frederik Stjernfelt, Center for Semiotics, University of Aarhus, Aarhus C, Denmark
Mauricio Suárez, Logic and Philosophy of Science, Complutense University, Madrid, Spain
Jeroen van den Hoven, Fac of Technology, Policy & Management, Delft University of Technology, Delft, Zuid-Holland, The Netherlands
Peter-Paul Verbeek, Department of Philosophy, University of Twente, Enschede, Overijssel, The Netherlands
Mireille Hildebrandt, Erasmus MC, Rotterdam, The Netherlands
Michael H. G. Hoffmann, School of Public Policy, Georgia Institute of Technology, Atlanta, GA, USA
Alfredo Pereira, Institute of Biosciences, Universidade de São Paulo, São Paulo, Brazil
Dagmar Provijn, Centre for Logic and Philosophy, Ghent University, Ghent, Belgium
Joao Queiroz, Institute of Arts and Design, Universidade Federal de Juiz de Fora, Juiz de Fora, Brazil
Colin T Schmidt, Ensam ParisTech & Le Mans University, Laval, France
Nora Schwartz, Department of Humanities, Universidad de Buenos Aires, Buenos Aires, Argentina
Marion Vorms, Pantheon-Sorbonne University, Paris, France
Riccardo Viale, Economics, Management and Statistics, University of Milano-Bicocca, Milan, Italy

Preface

The papers in this volume result from the 4th conference on the 'Philosophy and Theory of Artificial Intelligence' (PT–AI) that we organised in Gothenburg on 27–28 September, 2021 (see http://www.pt-ai.org/). The local organisation was taken care of by Profs. Ivica Crnkovic and Gordana Dodig-Crnkovic, with support from Chalmers University in Gothenburg.

This conference had been planned for a much earlier date, but several things got in the way, most of all the COVID pandemic. Eventually, we decided that in some way or other we should run PT-AI in 2021, and for most people, this was the first conference where actual physical presence was possible again, though with significant safety measures. It was the first PT-AI conference that was run in a hybrid online/onsite fashion (earlier conferences had already allowed online live listening to keynote talks). As a result of all this, the conference was a bit smaller than usual, with 60 submissions, of which 25 were presented at the meeting. My thanks to the colleagues on the programme committee who worked hard on the double-blind reviewing and assured a very high academic level! The inspiring invited speakers were Virginia Dignum (Umeå, Sweden), Michael Levin (Tufts, USA), David Papineau (KCL, UK), and Shannon Vallor (Edinburgh, UK).

It was very good to see many new faces coming to the field, as well as some established philosophers from other areas showing an interest in the philosophy of AI—which is clearly moving into the mainstream now. The new people and the influences from many different directions are clearly enriching the field and I expect this to continue.

I have to end on a bitter note: In February 2022, our co-organiser Ivica died suddenly. It is extremely sad to see such a rich life cut short, and many people left behind with a huge gap in their lives—first of all, his wife and his children. At the same time, I am also grateful that I had the chance to know Ivica and learn from him, academically and as a human being.

Eindhoven, The Netherlands
April 2022

Vincent C. Müller
http://www.sophia.de/

Contents

Theory

Cognitive Architectures Based on Natural Info-Computation 3
Gordana Dodig-Crnkovic

Artificial Intelligence Systems, Responsibility and Agential Self-Awareness 15
Lydia Farina

Models, Algorithms, and the Subjects of Transparency 27
Hajo Greif

Autonomy of Attention 39
Kaisa Kärki

Toward Out-of-Distribution Generalization Through Inductive Biases 57
Caterina Moruzzi

Is There a Trade-Off Between Human Autonomy and the ‘Autonomy’ of AI Systems? 67
Carina Prunkl

Towards a Taxonomy for the Opacity of AI Systems 73
Alessandro Facchini and Alberto Termine

Validating Non-trivial Semantic Properties of Autonomous Robots 91
Jiří Wiedermann and Jan van Leeuwen

Ethics

Dignity in Digital Ethics 107
Marcel Becker

An Enactive Approach to Value Alignment in Artificial Intelligence: A Matter of Relevance 119
Michael Cannon

The Problem of AI Influence 137
Laura Crompton

Artificial General Intelligence and the Common Sense Argument 155
Olle Häggström

Moral Status of AI Systems: Evaluation of the Genetic Account 161
Leonhard Kerkeling

Deception by Default 171
András Kornai

Robot Rights in Joint Action 179
Guido Löhr

Is It Likely that We Are Living in a Computer Simulation? 193
Ralf Stapelfeldt

Human-AI Friendship: Rejecting the Appropriate Sentimentality Criterion 209
Dan Weijers and Nick Munn

AI Risk Skepticism 225
Roman V. Yampolskiy