

Vincent C. Müller
Editor

Philosophy and Theory of Artificial Intelligence 2017

 Springer

Editorial Note

The papers in this volume result from the 3rd conference on the “Philosophy and Theory of Artificial Intelligence” (PT-AI) 4–5 November 2017 which I organised in Leeds where I am a university fellow—for details on the conference, see <http://www.pt-ai.org/>.

For this conference, we had 77 extended abstract submissions by the deadline, which were reviewed double-blind by two to four referees. A total of 28 submissions, i.e. 36%, were accepted for presentation. We also accepted 18 posters to be presented. The invited speakers were Thomas Metzinger, Mark Sprevak, José Hernández-Orallo, Yi Zeng, Susan Schneider, David C. Hogg and Peter Millican. All papers and posters were submitted in January at full length (two pages for posters) and reviewed another time by at least two referees among the authors. In the end, we have 32 papers here that represent the current state of the art in the philosophy of AI. We grouped the papers broadly into three categories: “Cognition–Reasoning–Consciousness”, “Computation–Intelligence–Machine Learning” and “Ethics–Law”. This year, we see a significant increase in ethics, more work on machine learning, perhaps less on embodiment or computation—and a stronger “feel” that our area of work has entered the mainstream. This is also evident from more papers in “standard” journals, more book publications with mainstream philosophy presses and more philosophers from neighbouring fields joining us, such as philosophy of mind or philosophy of science. These are encouraging developments, and we are looking forward to PT-AI 2019!

We gratefully acknowledge support from the journal *Artificial Intelligence*, and the *IDEA Centre* at the University of Leeds.

May 2018

Vincent C. Müller

Contents

Cognition - Reasoning - Consciousness

| | |
|---|----|
| Artificial Consciousness: From Impossibility to Multiplicity | 3 |
| Chuanfei Chin | |
| Cognition as Embodied Morphological Computation | 19 |
| Gordana Dodig-Crnkovic | |
| “The Action of the Brain”: Machine Models and Adaptive Functions in Turing and Ashby | 24 |
| Hajo Greif | |
| An Epistemological Approach to the Symbol Grounding Problem | 36 |
| Jodi Guazzini | |
| An Enactive Theory of Need Satisfaction | 40 |
| Soheil Human, Golnaz Bidabadi, Markus F. Peschl, and Vadim Savenkov | |
| Agency, Qualia and Life: Connecting Mind and Body Biologically | 43 |
| David Longinotti | |
| <i>Dynamic Concept Spaces</i> in Computational Creativity for Music | 57 |
| René Mogensen | |
| Creative AI: Music Composition Programs as an Extension of the Composer’s Mind | 69 |
| Caterina Moruzzi | |
| How Are Robots’ Reasons for Action Grounded? | 73 |
| Bryony Pierce | |
| Artificial Brains and Hybrid Minds | 81 |
| Paul Schweizer | |

| | |
|--|-----|
| Huge, but Unnoticed, Gaps Between Current AI and Natural Intelligence | 92 |
| Aaron Sloman | |
| Social Cognition and Artificial Agents | 106 |
| Anna Strasser | |
| Computation - Intelligence - Machine Learning | |
| Mapping Intelligence: Requirements and Possibilities | 117 |
| Sankalp Bhatnagar, Anna Alexandrova, Shahar Avin, Stephen Cave, Lucy Cheke, Matthew Crosby, Jan Feyereisl, Marta Halina, Bao Sheng Loe, Seán Ó hÉigeartaigh, Fernando Martínez-Plumed, Huw Price, Henry Shevlin, Adrian Weller, Alan Winfield, and José Hernández-Orallo | |
| Do Machine-Learning Machines Learn? | 136 |
| Selmer Bringsjord, Naveen Sundar Govindarajulu, Shreya Banerjee, and John Hummel | |
| Where Intelligence Lies: Externalist and Sociolinguistic Perspectives on the Turing Test and AI | 158 |
| Shlomo Danziger | |
| Modelling Machine Learning Models | 175 |
| Raül Fabra-Boluda, Cèsar Ferri, José Hernández-Orallo, Fernando Martínez-Plumed, and M. José Ramírez-Quintana | |
| Is Programming Done by Projection and Introspection? | 187 |
| Sam Freed | |
| Supporting Pluralism by Artificial Intelligence: Conceptualizing Epistemic Disagreements as Digital Artifacts | 190 |
| Soheil Human, Golnaz Bidabadi, and Vadim Savenkov | |
| The Frame Problem, Gödelian Incompleteness, and the Lucas-Penrose Argument: A Structural Analysis of Arguments About Limits of AI, and Its Physical and Metaphysical Consequences | 194 |
| Yoshihiro Maruyama | |
| Quantum Pancomputationalism and Statistical Data Science: From Symbolic to Statistical AI, and to Quantum AI | 207 |
| Yoshihiro Maruyama | |
| Getting Clarity by Defining Artificial Intelligence—A Survey | 212 |
| Dagmar Monett and Colin W. P. Lewis | |
| Epistemic Computation and Artificial Intelligence | 215 |
| Jiří Wiedermann and Jan van Leeuwen | |

Will Machine Learning Yield Machine Intelligence? 225
Carlos Zednik

Ethics - Law

In Critique of RoboLaw: The Model of SmartLaw 231
Paulius Astromskis

AAAI: An Argument Against Artificial Intelligence 235
Sander Beckers

Institutional Facts and AMAs in Society 248
Arzu Gokmen

A Systematic Account of Machine Moral Agency 252
Mahi Hardalupas

A Framework for Exploring Intelligent Artificial Personhood 255
Thomas B. Kane

**Against Leben’s Rawlsian Collision Algorithm
for Autonomous Vehicles 259**
Geoff Keeling

Moral Status of Digital Agents: Acting Under Uncertainty 273
Abhishek Mishra

Friendly Superintelligent AI: All You Need Is Love 288
Michael Prinzing

Autonomous Weapon Systems - An Alleged Responsibility Gap 302
Torben Swoboda

Author Index. 315