Croatian Journal of Philosophy
Vol. XX, No. 60, 2020

Book Reviews

Vincent C. Müller (ed.), Philosophy and Theory of Artificial Intelligence 2017*, Springer 2018, 313 pp.*

The volume we shall review here contains over thirty articles from various authors, divided into three thematic sections: *Cognition – Reasoning – Consciousness, Computation – Intelligence – Machine Learning,* and *Ethics – Law*. In this paper we shall focus on several articles selected from the book. However, some articles will not fully feature in the review, which is due to length limitations and for which I offer my apologies to the authors.

The first section premieres with *Artificial Consciousness: From Impossibility to Multiplicity* by Chuanfei Chin, in which the author analyses various debates and viewpoints on artificial consciousness, as well as skeptical arguments such as Chalmers' Hard Problem and potential rebuttals to such arguments. The following four articles will be, with all due respect to their authors merely listed by titles: *Cognition as Embodied Morphological Computation* (Dodig-Crnkovic), *"The Action of the Brain": Machine Models and Adaptive functions in Turing and Ashby* (Greif), *An Epistemological Approach to the Symbol Grounding Problem* (Guazzini), and *An Enactive Theory of Need Satisfaction* (Human et al.).

Next article that we will delve deeper into is David Longinotti's *Agency, Qualia and Life: Connecting Mind and Body Biologically,* which presents a very interesting perspective on the Mind-Body Problem, claiming, against the Strong Artificial Intelligence thesis, that qualia are unique to biological agents and may only be produced and explained by naturally occurring neural networks. Furthermore, the article explains qualia as a form of energy, and functionally defines it as 'control signals in the regulatory processes'.

How Are Robots' Reasons for Actions Grounded? by Bryony Pierce overviews different possible kinds of grounding to conclude that a non-conscious artificial agent's actions can only ever be externally grounded in the affective responses of their users or creators, as consciousness is necessary for grounding. The author states that while a conscious robot would be capable of internal grounding, creation of such robot would not be morally permissible.

Anna Strasser's paper *Social Cognition and Artificial Agents* takes cue in increasing presence of ever-smarter technology in our daily lives and thus human society. She follows a minimal approach to socio-cognitive abilities, according to which artificial agents can be considered as having such abilities, in that they can – to a minimal extent – understand other social

agents, exchange social cues with them and show a sense of commitment. The paper certainly raises or reminds the reader of numerous questions on status of artificial agents as members of society, assuming, of course, the reader agrees with the author's conclusions.

Other articles in this first section of the book that we have not yet mentioned are *Creative AI: Music Composition Programs as an Extension of the Composer's Mind* (Moruzzi), *Artificial Brains and Hybrid Minds* (Schweizer), and *Huge, but Unnoticed, Gaps Between Current AI and Natural Intelligence* (Sloman). An article that especially deserves a separate mention is René Mogensen's *Dynamic Concept Spaces in Computational Creativity for Music*, which has, however, unfortunately proved to be too complex to separately describe in this short review.

Moving to the second section of the book, we pay some extra attention to Shlomo Danziger's *Where Intelligence Lies: Externalist and Sociolinguistic Perspectives on the Turing Test and AI.* This paper reinterprets Alan Turing's Imitation Game as a test which also includes the socio-linguistic dimension of how an agent is perceived by the society, focusing on human prejudice and our anthropocentric tendencies, which may present an obstacle in objectively evaluating artificial agents' intelligence. With this, the paper presents a rarely seen point-of-view towards human-AI relations, and, as the author concludes, teaches us 'quite a bit about human intelligence as well'.

Will Machine Learning Yield Machine Intelligence? by Carlos Zednik, published as the last article in the second section, considers how Explainable AIs might be able to solve the Black Box Problem, and what that means for the prospect of machine intelligence, concluding, after a short but concise analysis, that the answer to the question in title may well be affirmative, iff we acknowledge algorithmic similarity of AI and humans to be a sufficient criterion for machine intelligence.

Other articles in this section that merit a special mention are *Supporting Pluralism by Artificial Intelligence: Conceptualizing Epistemic Disagreements as Digital Artefacts* (Human et al.), and Yoshihiro Maruyama's paper *Quantum Pancomputationalism and Statistical Data Science: From Symbolic to Statistical AI, and to Quantum AI.*

Other articles in this section will, with all due respect to their authors, be omitted in this review to avoid the risk of ending up simply with an augmented table of contents.

The third section contains contributions on numerous ethical issues that have been subject of recent debates in AI development, such as autonomous vehicles in *Against Leben's Rawlsian Collision Algorithm for Autonomous vehicles* (Keeling), or less commonly discussed autonomous weapons in *Autonomous Weapons Systems – An Alleged Responsibility Gap* (Swoboda).

AAAI: An Argument Against Artificial Intelligence by Sander Beckers considers perils development of conscious AI presents both for humanity and such AI itself, examining the probabilities of potential conscious AIs' suffering, as well as risks a superintelligent AI could present to humans, concluding rather decisively that there should be an ethics-based ban on developing conscious artificial agents.

Another paper that caught additional attention is Abhishek Mishra's *Moral Status of Digital Agents: Acting Under Uncertainty,* which considers,

rather than the moral status of AIs as artificial agents, the moral status of agents created within simulations of such AIs. Considering grounds for moral status of such agents and the Decision Problem that arises, the author points out the numerous additional questions raised in the process of attempting to solve the Decision Problem, creating an intriguing cue for future research and discussion on the subject.

Overall, the volume offers, also within articles I have omitted in this review, for which I again extend my apologies to their authors, numerous valuable contributions to already ongoing as well as new discussions taking place on the topic of AI Theory, making it a must-read for anyone working in AI-related fields, and an intriguing mental exercise for anyone simply interested.

NIKO ŠETAR
University of Maribor, Maribor, Slovenia