logos_i_ethos_2020_(52), s. 33–48

DOI: http://dx.doi.org/10.15633/lie.3576

Paweł Polak

ORCID 0000-0003-1078-469X

Uniwersytet Papieski Jana Pawła II w Krakowie

Roman Krzanowski

ORCID 0000-0002-8753-0957

Uniwersytet Papieski Jana Pawła II w Krakowie

Ethics in autonomous robots as philosophy *in silico*: The study case of phronetic machine ethics

Paweł Polak – professor of philosophy at Pontifical University of John Paul II in Krakow, studied telecommunication at AGH University of Science and Technology in Kraków and philosophy at Pontifical Academy of Theology. Editor-in-chief of periodical "Philosophical Problems in Science (Zagadnienia Filozoficzne w Nauce)," member of Commission on the History of Science of the Polish Academy of Arts and Sciences, secretary of the Commission on the Philosophy of Science (the same Academy); published in "Studies in Logic, Grammar and Rhetoric," "Studia Historiae Scientiarum," "The Philosophy of Science/Filozofia Nauki." He published books about philosophical aspects of scientometrics and about philosophical reception of Special and General Relativity in Lwów. His interests in philosophy include the history and philosophy of computing/informatics, pancomputationalism, history of Polish philosophy, applied philosophy (e.g. wine philosophy).
Roman M. Krzanowski – has graduate degrees in engineering, philosophy and information sciences from universities in Poland, UK, Canada, and the USA. He is an expert in Ethernet networking technology spatial information systems and information processing. In his Ph.D. he developed the class of the spatial genetic algorithms. He published books in information science, network technology and the philosophy of Tao and TDK. He holds numerous international patents in information and networking systems. He also authored several networking standards with international standard organizations. His interests in philosophy include the philosophy of information and informatics, ontology and metaphysics of nature, ethics and ethical problems created in information society, the problem of infospehere, the history of ideas with the special attention to pre-Socratics, Plato, Aristotle and applied philosophy. He is currently working on the conceptualization of the nature of information and its ontology.

Introduction

This paper discusses the application of computer modeling (i.e., modeling *in silico*) to philosophical problems, especially ethical ones. Computer modeling has become a generally accepted method for researching, testing, and validating scientific theories, engineering models, and real-life scenarios and models (denoted as simulations)[1] in numerous *technological and scientific fields.*[2] Indeed, *in silico* methods are much cheaper and safer to implement than actual real-life studies. What is more, some studies are not possible without such methods because of their complexity or possible negative effects. *In silico* modeling has opened up new research venues, but it also has brought forward fresh epistemic challenges about how to build the "right" models and how to interpret the results of these studies.[3]

The idea of using mathematics (i.e., mathematical methods in general, of which computers are just one form) as a philosophical method predates the advent of computer systems. A prominent example of this approach is Isaac Newton's *Mathematical Principles of Natural Philosophy*,[4] where he formulated a general description of the physical world (which was conceived as a philosophical theory) in the language of mathematics. Thus, for Newton, mathematics served as a language to describe philosophical concepts. In the first half of the 20th century,

[1] Simulation (sometimes called computer-based simulation) in this context denotes the use of computer-based models to investigate the properties of physical, social, or other phenomena that can be represented meaningfully in computer symbolism. The computer models used in such simulations are mathematical abstractions, so their results only approximate the properties of the modeled objects. The relation between the simulated phenomenon and the simulation itself is called the epistemic distance between a computer model and the reality it represents.

[2] See, for example P. Thagard, *Computational Models in Science and Philosophy*, in: *Introduction to Formal Philosophy*, eds. S. O. Hansson, V. F. Hendricks, Cham 2018, p. 457–467, doi: 10.1007/978-3-319-77434-3_24; E. Winsberg, *Computer Simulation and the Philosophy of Science*, "Philosophy Compass" 4 (2009) no. 5, p. 835–845, doi: 10.1111/j.1747-9991.2009.00236.x.

[3] E. Winsberg, *Sanctioning Models: The Epistemology of Simulation*, "Science in Context" 12 (1999) no. 2, p. 275–292, doi: 10.1017/S0269889700003422.

[4] I. Newton, *Philosophiae Naturalis Principia Mathematica*, Londini 1687.

the deductive structures of formal logic were used to analyze classical concepts of philosophy.[5] Later, in the latter half of that century, different mathematical structures were used for the same type of philosophical analysis. For example, Raine and Heller applied the mathematical concept of differentiable manifold to analyze the properties of space-time structures and correct some interpretations of classical concepts for dynamics (e.g., the dynamics of Aristotle, Newton, and Special and General Relativity).[6]

The obvious question is then whether computational modeling, or *in silico* methods, could be useful for philosophy beyond the study of formal methods.[7] It seems that as computational modeling, or *in silico*, studies in science serve as an extension of the methodology of philosophy in science (Heller's methodology), adopting them in other areas of philosophy would only be a natural extension.[8]

So far, little work has applied *in silico* methods to classical and modern philosophical questions. The likely reason for this is that philosophy is often regarded as a mental exercise in theory with little or no practical import, so it does not lend itself to algorithmic, digital methods for *in silico* studies, excluding of course purely formal studies

[5] See, for example J. Salamucha, *Dowód "ex motu" na istnienie Boga. Analiza logiczna argumentacji św. Tomasza z Akwinu*, "Collectanea Theologica" 54 (1934) no. 1–2, p. 53–92; English translation: J. Salamucha, *The proof "ex motu" for the existence of God. Logical analysis of St. Thomas' arguments*, "The New Scholasticism" 32 (1958), p. 327–372; Salamucha's analysis could be conceived as fruitful for philosophical research, because it inspired formalization of the concept of change, see, for example, L. Larouche, *Examination of the axiomatic foundations of a theory of change I*, "Notre Dame Journal of Formal Logic" 9 (1968) no. 4, p. 371–384.

[6] D. J. Raine, M. Heller, *The Science of Space-Time*, Tucson, AZ 1981, p. 240; M. Heller, *Evolution of space-time structures*, "Concepts of Physics" 3 (2006), p. 117–131; M. Heller, *How is philosophy in science possible?*, "Philosophical Problems in Science (Zagadnienia Filozoficzne w Nauce)" 2019 no. 66, p. 231–249; P. Polak, *Philosophy in science: A name with a long intellectual tradition*, "Philosophical Problems in Science (Zagadnienia Filozoficzne w Nauce)" 2019 no. 66, p. 251–270.

[7] See, for example, P. Thagard, *Computational Models in Science and Philosophy*…, dz. cyt., passim.

[8] See, for example, A. Sloman, *The Computer Revolution in Philosophy: Philosophy, Science, and Models of Mind*, Atlantic Highlands, NJ 1978, passim; J. Vallverdú, *Thinking Machines and the Philosophy of Computer Science: Concepts and Principles*, Hershey, PA 2010, passim.

in ontology, logic, or linguistics. However, with philosophy working on the problems of representing the mind, the simulation of cognitive functions, and the modeling of ethical situations (e.g., the trolley problem and game theory), *in silico* methods may become a part of the philosophical toolbox. In addition, experimental philosophy or methods, such as reflective equilibrium[9] or conceptual analysis,[10] seem to be almost begging to be explored with the aid of computational methods.

Interesting studies into the computational modeling of philosophical problems have been undertaken in Kraków, namely in the work of Robert Janusz SI at the Vatican Observatory and Copernicus Center for Interdisciplinary Studies. Robert Janusz proposed using the object-oriented programming paradigm to model classical metaphysical[11] concepts and psycho-ethical[12] relations and formalize the concept of analogy.[13]

The study presented here investigates an application of the *in silico* method to the modeling of robotic ethics based on the concept of phronesis. Specifically, we wish to investigate whether *in silico* models can deliver a philosophical analysis of the phronetic ethics implemented in autonomous robots, determine how ethical research into autonomous robotic ethics may benefit from *in silico* methods, and establish how our understanding of ethics could benefit

[9] J. Rawls, *A Theory of Justice*, Cambridge 1971, passim.

[10] M. Beaney, *Analysis*, in: *The Stanford Encyclopedia of Philosophy (Summer 2018 Edition)*, ed. E. N. Zalta, Stanford 2018, https://stanford.library.sydney.edu.au/archives/sum2018/entries/analysis/.

[11] R. Janusz, *Program dla Wszechświata: filozoficzne aspekty języków obiektowych*, Kraków 2002, passim; another interesting formalization of Leibniz's metaphysical concepts was done by Jarosław Strzelecki, see J. Strzelecki, *Monada = Monada() – interpretacja obiektowa*, in: *Filozofia i technika*, red. J. Sobota, G. Pacewicz, Olsztyn 2017, p. 35–52.

[12] R. Janusz, *Relacja etyczno-psychologiczna w ujęciu obiektowym*, in: *Philosophiae & musicae: księga pamiątkowa z okazji jubileuszu 75-lecia urodzin księdza profesora Stanisława Ziemiańskiego SJ*, red. R. Darowski, Kraków 2006, p. 375–380.

[13] R. Janusz, *O metodach wirtualnych w paradygmacie obiektowym*, "Zagadnienia Filozoficzne w Nauce" (2007) no. 41, p. 125–131.

from *in silico* studies. The field of robotic ethics seems particularly well suited to *in silico* methodology because it requires, on the one hand, a deep understanding of ethical issues that are always regarded as controversial and poorly articulated, while on the other hand, robotic ethics needs a very meticulous verification of computerized ethical solutions in real-life scenarios. The clarification and disambiguation of ethical concepts and the large-scale verification (implying a multitude of complex testing scenarios) of ethical solutions are some areas of ethical robotics where *in silico* methods may prove indispensable.

What is robotic ethics?

Ethics in autonomous robots (a-robots) is implemented in an abstract computational model of the Turing Machine (TM). Such a-robots can only be behavioral systems.[14] They are ethical zombies,[15] because TMs only implement behavioral features without the associated mental capacities.[16] The behavioral approach to morality has been long disavowed, however. Thus, if our aspiration is to have a-robots with human--like ethics, we could say that a-robots implement the wrong ethical model in the wrong computational paradigm. Of course, this may seem like a gross oversimplification of the research into ethical robotics.

[14] Contemporary psychology and philosophy largely share Hempel's conviction that any explanation of behavior cannot omit invoking a creature's representation of its world: "Psychology must use psychological terms. Behavior without cognition is blind. Psychological theorizing without reference to internal cognitive processing is explanatorily impaired" (C. G. Hempel, *Philosophy of Natural Science*, Prentice-Hall 1966).

[15] The TM paradigm limits the ethical capacities of TM-based systems. It would seem than that any extension to the TM model—such as non-deterministic TMs (NDTMs), c-machines, and o-machines—would not change the essence of the paradigm (i.e., eliminate its limitations regarding ethical capacities). See J. R. Searle, *Mind, Language and Society: Philosophy in the Real World*, New York 1998, passim.

[16] Behavior can be described and explained without making ultimate reference to mental events or internal psychological processes. The sources of behavior are external (i.e., in the environment) rather than internal (i.e., in the mind).

However, on close scrutiny, the various computerized ethics implementing emotional rules, virtues, and mixtures of deterministic and probabilistic algorithms (see later in the text) seem to be variations on the same theme.[17] When a-robots enter our social environment, we need them to think and make decisions like we do and share our social norms and values,[18] so we can accept them as social partners, particularly if we plan[19] to grant them some of our rights![20] A-robots need human-like ethics rather than zombie-like one.

Autonomous robots with ethical capacities (e-robots) have access to facts (i.e., information about the current environment at a given time), past ethical decisions, and past outcomes on an incomparably larger scale than we humans do. In addition, e-robots could potentially have vastly more powerful inference capacities than us. Yet so far, despite all the available computational power and past and present data, no robots have demonstrated superior ethics and better ethical decision-making than us humans, even if we do err so often!

[17] See M. Anderson, S. L. Anderson, *Machine Ethics*, Cambridge 2011; M. Anderson, S. L. Anderson, *GenEth: General Ethical Dilemma Analyzer*, in: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, Québec 2014, p. 253–261; L. Floridi, J. W. Sanders, *On the morality of artificial agents*, "Minds and Machines" 14 (2004) no. 3, p. 349–379, doi: 10.1023/B:MIND.0000035461.63578.9d; W. Wallach, C. Allen, *Moral Machines: Teaching Robots Right from Wrong*, Oxford 2009; W. Wallach, S. Franklin, C. Allen, *A conceptual and computational model of moral decision making in human and artificial agents*, "Topics in Cognitive Science" 2 (2010) no. 3, p. 454–485, doi: 10.1111/j.1756-8765.2010.01095.x.

[18] See R. van Oers, E. Wesselman, *Social Robots*, in: KPMG Advisory 2016; I. Leite, C. Martinho, A. Paiva, *Social Robots for Long-Term Interaction: A Survey*, "International Journal of Social Robotics" 5 (2013) no. 2, p. 291–308, doi: 10.1007/s12369-013-0178-y.

[19] „Creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently" (European Parliament, *Civil Law Rules on Robotics European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))*, Rule 56.f., https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html).

[20] F. Alaieri, A. Vellino, *Ethical decision making in robots: Autonomy, trust and responsibility*, in: *Social Robotics*, eds. A. Agah et al., vol. 9979, Cham 2016, p. 159–168, doi: 10.1007/978-3-319-47437-3_16.

We make the claim that a critical, although not the only, difference in ethics between a-robots and humans is in the "inference/ascend method." In other words, this is a way of ascending to an ethical decision based on the facts present, the objectives of an action, and past experiences and their outcomes. Thus, to improve the moral capacities of a-robots, we need to change the way we "compute" ethical decisions. In other words, we need a paradigm shift in "roboethics." The new model that we propose is based on the Aristotelian concept of phronesis.[21]

What are phronetic robotic ethics?

What is phronesis or ethical wisdom? Phronesis, according to Aristotle, is not an exact science (like *episteme* in the Aristotelian sense) or art (like *techne*, craft or practical skills). Unlike exact sciences, it is not based on ultimate (necessary) principles.[22] The general principles of an exact science (for Aristotle) are "absolute," because they express abstract, nominal truths. Ethics, meanwhile, deal with "enmattered things" and with the changeability and variability inherent in concrete embodied facts. The objective of phronesis is therefore *eudaimonia*, a realization of a specific concept of good for a person (or actor). The focus of phronesis is the difference presented by a specific case from ultimate principles rather than its fit with "general" rules.[23] Phronesis cannot be taught like mathematics, for example, because there are no rules to teach in phronesis.[24] Phronetic expertise can only be gained through direct experience.

[21] P. Polak, R. Krzanowski, *Phronetic ethics in social robotics: A new approach to building ethical robots*, (2019), doi: 10.13140/rg.2.2.16802.79049.

[22] Aristotle, *The Nicomachean Ethics*, trans. J. A. K. Thomson, London–New York 2004; C. D. C. Reeve, *Practices of Reason: Aristotle's Nicomachean Ethics*, Oxford 1992.

[23] "The decision in this case is not from universal principles to the case but from the case to the universal principles (arguing to the first principles [101a37-b4], not from the first principles, as exact sciences do)" (J. H. Randall, *Aristotle*, New York 1965, p. 267).

[24] I.e., ethics is not Euclidian geometry.

We will not provide a detailed explanation of what phronesis is, because others have done this for some time already.[25] However, we will summarize the principles informing phronetic decisions:

1. Phronetic ascend cannot be encapsulated in a set of rules, because it deals with the specificity of particular cases (i.e., it is not procedural).

2. Phronetic ascend is case-specific (i.e., the focus of the phronetic decision is a particular case).

3. The decision for the specific case is not achieved through logical analysis (which originally meant Aristotelian logic) or reasoning through principles (like in science) but rather through intuitively grasping the outcome or having "foresight of consequences."

The challenge faced by philosophers and computer scientists is to find a way to translate these principles into a robotic decision-making system. This is at the junction where philosophical analysis and *in silico* methods come to meet and interact.

Phronesis seems to be at the nexus of personal ethical sensitivity, intuition, experience, and ethical wisdom accumulated through individual and community experiences, whether communicated implicitly or explicitly through social interactions. How much of this complex collection of ideas, experiences, and unwritten rules may be translated into logic, and therefore programmed, is obviously a challenge to establish.

What can *in silico* models offer robotic ethics research?

There are several aspects of roboethic research, and specifically phronetic roboethics, to which *in silico* studies may be helpful. The many aspects of phronetic ethics, and ethics in general, have always been ambiguous and poorly understood. These include the problem of ethical ascend, the problem of ethical choice, the meaning of the *telo*s of ethics (the ultimate goal of ethics) and *eudaimonia* (the best good), the problem of generalizing ethical cases, and the problem of learning from

[25] J. H. Randall, *Aristotle*…, op. cit. passim; C. D. C. Reeve, *Practices of Reason*…, op. cit., passim.

experience what information is relevant to the ethical context, so we know what to retain and what to discard.

We may wonder how *in silico* studies may help. An *in silico* environment (as a computer model) requires a precise definition of terms, dependencies, and methods. *In silico* models will not accept poorly defined concepts, unclear specifications, or simple hand waving for the difficult aspects of a problem, all of which may appear in philosophical works and work on ethics (including ethics in robotics). While some philosophical problems cannot, by virtue of their nature, be exactly translated into an *in silico* environment (e.g., intuitions, moral values, hopes, and preferences), the effort to translate them into some computer-compatible form obliges researchers to perform a very detailed analysis of concepts and clarify the ideas. This mandatory specificity and clarity for the problem sometimes requires a more complex and accurate study than a typical conceptual analysis of analytic philosophy would demand. Questions about ethical ascend, intuition, decision-making in complex scenarios, and the important aspects of a case all have to be laid down in detail.

Creating a physical model of an autonomous robot and then experimenting with its capacities in a real environment is very costly, impractical, and possibly even dangerous. It is much safer to test such a robot in a virtual environment (i.e., *in silico*) using virtual scenarios. This is cheaper, faster, and operationally more efficient. *In silico* models can test more complex scenarios, expose the *in silico* ethical system to a much larger spectrum of situations,[26] and do it much faster. Table 1 presents the principles of phronesis, roboethics, and their implications for *in silico* studies.

[26] It is much cheaper and safer to teach pilots on airplane simulators than on real airplanes. Engineers also understand this already.

Table 1. Proposed interpretation Phronesis, roboethics, and *in silico* methods

Phronesis	Roboethics	*In silico* modeling
Ethics is not an exact science, so one cannot define generalized universal rules to guide behavior. (i.e., decisions are taken in particular cases, but general rules are always generalizations [sic], so they will sometimes be wrong in particular cases.)	This implies that the development of algorithmic rules to solve ethical problems will not improve the quality of ethical decision-making.	Disambiguation of ethical concepts; Clarification of the ethical decision-making process.
Ethical decisions are all case-specific and guided by the weight of experience.	This means that phronetic capacities are not "programmed" a priori but learned.	Learning from experience, formalizing ethical ascend, and testing and verifying ethical rules on a large scale
Ethical decisions always go from the specific case to the universal principle, not from the universal principle to the specific case.	This means that the focus of an ethical decision is the ethical aspect of the specific case rather than universal rules that generalize cases.	Clarifying generalization rules
The *telos* of ethics (the ultimate goal of ethics) is *eudaimonia* (the best good).	The *telos* of ethical decisions (in the case of a-robots) is the best good of the human actors involved, not the a-robots themselves. Other formulations of *telos* for a-robots would also be possible.	Disambiguation of ethical concepts

Conclusions

It may seem that *in silico* studies—while applicable to problems in science, technology, and mathematics—are of no import to philosophical research. This is a rather incomplete and dated perspective, however. With philosophy tackling increasingly more practical problems, *in silico* modeling may become an accepted philosophical methodology. From the clarification of ideas, through detailed conceptual analysis and the verification of proposed concepts, *in silico* modeling may enhance and enrich philosophical study. What is more, let us also remember that there is a mutual dependency between *in silico* modeling and philosophy.

In this study, we show how certain aspects of robotic ethics could benefit from the application of *in silico* methodology. Implementing phronetic ethics in autonomous machines would force a deep conceptual analysis of Aristotelian ethics. Several aspects of such ethics cannot be easily formalized, so we need to reformulate and rethink some aspects of phronetic ethics before we can use them in non-human ethical agents. This leads to new philosophical questions and problems and opens the way for new discussions about classical concepts. For example we may ask the question, what are the properties of a human actor as an ethical agent in Aristotelian ethics, and why are they unique?

A review of Aristotelian ethics reveals hidden premises that are implicitly assumed to be self-evident for a human ethical agent. Extending ethical agency to include autonomous systems, however, makes most of these assumptions problematic, so they need reframing. By explicating these tacit assumptions hidden within the classical concepts, we deepen the analysis of classical knowledge, leading not just to new solutions but also a better understanding of the old ideas. The analysis imposed by *in silico* methodology also gives us also a new philosophical perspective on the differences between humans and animals, humans and autonomous robots, and animals and autonomous robots. Attempts to formalize and implement phronetic ethics could also reveal fundamental connections between distant regions of philosophical thought (e.g., the role of semiotics in ethics). Phronetic roboethics and *in silico* methodology force the philosopher to clarify the problem of ethical ascend, elucidate the issue of ethical choice, and explicate the meaning of the *telos* of ethics (the ultimate goal of ethics). *In silico* studies force us to define the problems of generalizing ethical cases and learning from experience, as well as how to determine what information is ethically relevant to retain and what can be discarded, which in *in silico* terms is called information reduction. The *in silico* approach to ethical problems has also some heuristic potential, because it encourages new research into classical concepts. For example, the need to adapt Aristotle's ethics suggests that some fundamental concepts, such as *eudaimonia*, could be generalized.[27] Computational modeling in ethics also brings possibilities for the quasi-empirical testing of hybrid ethical concepts, like in Weronika Wojtanowska's proposition of enrichment for Thomas Nagel's ethics through some elements of virtue ethics.[28] A question remains, however: Would an ethical robot, even with some form of phronesis, display the normal range of human cognitive faculties, or would it display some of the symptoms that characterize disorders like autism spectrum disorders (ASD) and other dysfunctions? We are not entirely sure what causes these disorders in human subjects, so while programming robots, we could inadvertently create dysfunctional artificial minds. In short, we claim that *in silico* methodology for ethical research in robotics will open up new areas of discussion around the limits of machine ethics.

Ethical issues, including those of Aristotle's phronesis, always present an intellectual challenge and are therefore fertile grounds for deep philosophical studies, but they are usually lacking in praxis. The disconnect between the theoretical and practical, for an acute mind, is one of the great paradoxes of Aristotelian ethics. The philosophy of ethical thought as a practical philosophy has never produced clear and unambiguous practical solutions and guidelines. Is this a nature of the problem? Or is there an intellectual weakness from our side? *In silico* studies may help answer this question.

Now, we need to ask whether we should leave the concepts of phronesis, ethical ascend, and related issues as vague and undefined as they are in Aristole's ethics, or could we, benefit from actually applying them to quasi-real-life (i.e., virtual) scenarios through *in silico*

[27] For example, the issue of generalizing the concept of *eudaimonia* was analyzed by Weronika Wojtanowska in her book *Próba rozwinięcia filozofii moralnej Thomasa Nagela o elementy etyki cnót* (forthcoming). Wojtanowska started by analyzing Thomas Nagel's concept and tried to develop this abstract idea, but the value of this philosophical consideration was unclear until it was applied in an *in silico* approach.

[28] W. Wojtanowska, *Próba rozwinięcia filozofii moralnej Thomasa Nagela o elementy etyki cnót* [to be published], Kraków 2020.

methods. After all, the confrontation with reality (or "trial by fire," although in our case, the fire is virtual) was something that Aristotelian ideas were originally designed to face. Will philosophy be open to new approaches, or will it choose to limit itself to the theoretical study of ideas, afraid of being confronted with reality, even if it is only a reality *in silico*?

Bibliography

Alaieri F., Vellino A., *Ethical decision making in robots: Autonomy, trust and responsibility*, in: *Social Robotics*, eds. A. Agah, J.-J. Cabibihan, A. M. Howard, et al., vol. 9979, Cham 2016, p. 159–168, doi: 10.1007/978-3-319-47437-3_16.

Anderson M., Anderson S. L., *GenEth: General Ethical Dilemma Analyzer*, in: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, Québec 2014, p. 253–261.

Anderson M., Anderson S. L., *Machine Ethics*, Cambridge 2011.

Aristotle, *The Nicomachean Ethics*, trans. J. A. K. Thomson, London–New York 2004.

Beaney M., *Analysis*, in: *The Stanford Encyclopedia of Philosophy (Summer 2018 Edition)*, ed. E. N. Zalta, Stanford 2018, https://stanford.library.sydney.edu.au/archives/sum2018/entries/analysis/.

European Parliament, *Civil Law Rules on Robotics European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))*, https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html.

Floridi L., Sanders J. W., *On the morality of artificial agents*, "Minds and Machines" 14 (2004) no. 3, p. 349–379, doi: 10.1023/B:MIND.0000035461.63578.9d.

Heller M., *Evolution of space-time structures*, "Concepts of Physics" 3 (2006), p. 117–131.

Heller M., *How is philosophy in science possible?*, "Philosophical Problems in Science (Zagadnienia Filozoficzne w Nauce)" 2019 no. 66, p. 231–249.

Hempel C. G., *Philosophy of Natural Science*, Englewood Cliffs 1966.

Janusz R., *O metodach wirtualnych w paradygmacie obiektowym*, "Zagadnienia Filozoficzne w Nauce" 2007 no. 41, p. 125–131.

Janusz R., *Program dla Wszechświata: filozoficzne aspekty języków obiektowych*, Kraków 2002.

Janusz R., *Relacja etyczno-psychologiczna w ujęciu obiektowym*, in: *Philosophiae & musicae: księga pamiątkowa z okazji jubileuszu 75-lecia urodzin księdza profesora Stanisława Ziemiańskiego SJ*, red. R. Darowski, Kraków 2006, p. 375–380.

Larouche L., *Examination of the axiomatic foundations of a theory of change I*, "Notre Dame Journal of Formal Logic" 9 (1968) no. 4, p. 371–384.

Leite I., Martinho C., Paiva A., *Social Robots for Long-Term Interaction: A Survey*, "International Journal of Social Robotics" 5 (2013) no. 2, p. 291–308, doi: 10.1007/s12369-013-0178-y.

Newton I., *Philosophiae Naturalis Principia Mathematica*, Londini 1687.

Oers R. van, Wesselman E., *Social Robots*, https://assets.kpmg/content/dam/kpmg/pdf/2016/06/social-robots.pdf.

Polak P., *Philosophy in science: A name with a long intellectual tradition*, "Philosophical Problems in Science (Zagadnienia Filozoficzne w Nauce)" 2019 no. 66, p. 251–270.

Polak P., Krzanowski R., *Phronetic ethics in social robotics: A new approach to building ethical robots* (2019), doi: 10.13140/rg.2.2.16802.79049.

Raine D. J., Heller M., *The Science of Space-Time*, Tucson, AZ 1981.

Randall J. H., *Aristotle*, New York 1965.

Rawls J., *A Theory of Justice*, Cambridge 1971.

Reeve C. D. C., *Practices of Reason: Aristotle's Nicomachean Ethics*, Oxford 1992.

Salamucha J., *Dowód ex motu na istnienie Boga. Analiza logiczna argumentacji św. Tomasza z Akwinu*, "Collectanea Theologica" 54 (1934) nr 1–2, p. 53–92.

Salamucha J., *The proof ex motu for the existence of God. Logical analysis of St. Thomas' arguments*, "The New Scholasticism" 32 (1958), p. 327–372.

Searle J. R., *Mind, Language and Society: Philosophy in the Real World*, New York 1998.

Sloman A., *The Computer Revolution in Philosophy: Philosophy, Science, and Models of Mind*, Atlantic Highlands, NJ 1978.

Strzelecki J., *Monada = Monada() – interpretacja obiektowa*, in: *Filozofia i technika*, red. J. Sobota, G. Pacewicz, Olsztyn 2017, p. 35–52.

Thagard P., *Computational Models in Science and Philosophy*, in: *Introduction to Formal Philosophy*, eds. S. O. Hansson, V. F. Hendricks, Cham 2018, p. 457–467, doi: 10.1007/978-3-319-77434-3_24.

Vallverdú J., *Thinking Machines and the Philosophy of Computer Science: Concepts and Principles*, Hershey, PA 2010.

Wallach W., Allen C., *Moral Machines: Teaching Robots Right from Wrong*, Oxford 2009.

Wallach W., Franklin S., Allen C., *A conceptual and computational model of moral decision making in human and artificial agents*, “Topics in Cognitive Science” 2 (2010) no. 3, p. 454–485, doi: 10.1111/j.1756-8765.2010.01095.x.

Winsberg E., *Computer Simulation and the Philosophy of Science*, “Philosophy Compass” 4 (2009) no. 5, p. 835–845, doi: 10.1111/j.1747-9991.2009.00236.x.

Winsberg E., *Sanctioning Models: The Epistemology of Simulation*, “Science in Context” 12 (1999) no. 2, p. 275–292, doi: 10.1017/S0269889700003422.

Abstrakt

Implementacja etyki autonomicznych robotów jako przykład filozofii *in silico*: Studium przypadku fronetycznej etyki maszyn

W artykule przestawiono zastosowanie narzędzi informatycznych do modelowania koncepcji etycznych. Komputerowe modelowanie nazywane jest modelowaniem *in silico*. Metody tego typu mają zastosowania m.in. w biologii, chemii, kosmologii, socjologii. Rzadko jednak stosuje się to podejście do modelowania problemów filozoficznych (jak etyka). Obecnie wydaje się ono obiecujące nie tylko dla uproszczenia rozwoju etycznych robotów, ale i dla głębszego wglądu w istotę filozoficznych problemów uwikłanych w kwestie związane z etyką maszyn (poprzez ukazanie ich wewnętrznej struktury). Artykuł ukazuje również zwięzły przegląd koncepcji modelowania w kontekście historycznym, jak i we współczesnym.

Słowa kluczowe

modelowanie komputerowe, cyfrowa humanistyka, etyka maszyn, roboty autonomiczne, metodologia filozofii, metodologia etyki

Abstract

Ethics in autonomous robots as philosophy *in silico*: The study case of phronetic machine ethics

The paper explores the application of computing science to the modeling of the ethical concepts. The modeling in computers is denoted as in silico modeling. The in silico method has found applications in biology, chemistry, cosmology, sociology among others. The applications of in silico modeling to philosophical problems (like ethics) are rather infrequent. Yet, the approach discussed in the paper holds the promise of not only facilitating the development of ethical robotics but it also may provide the insights into the philosophical problems themselves (by explicating their implicit structures). The paper provides also a brief overview of the concept of modeling in silico in historical and current contexts

Keywords

computational modeling, digital humanities, machine ethics, autonomous robots, methodology of philosophy, methodology of ethics