AI & SOCIETY (2019) 34:365–371
https://doi.org/10.1007/s00146-018-0833-9

OPEN FORUM

Why AI shall emerge in the one of possible worlds?

Ignacy Sitnicki[1]

Received: 1 October 2017 / Accepted: 20 February 2018 / Published online: 6 March 2018

Abstract
The aim of this paper is to present some philosophical considerations about the supposed AI emergence in the future. However, the predicted timeline of this process is uncertain. To avoid any kind of speculations on the proposed analysis from a scientific point of view, a metaphysical approach is undertaken as a modal context of the discussion. I argue that modal claim about possible AI emergence at a certain point of time in the future is justified from a temporal perspective. Therefore, worldwide society must be prepared for possible AI emergence and the expected profound impact of such an event on the existential status of humanity.

Keywords Artificial Intelligence · Superintelligence · Existential risk · Possible worlds · Modality · Contingency · Metaphysics · Anthropic principle · Entropy · Extropy · Human enhancement

1 Philosophical considerations of AI emergence

The philosophical debate of the emergence of AI takes place at various levels. The first level is tied to the mathematically and logically directed approach corresponding with Alan Turing's invention that resulted in the presentation of the so-called Turing machine (Penrose 2005)—the first theoretical model of computational system based on programmed set of algorithms. The second approach is that of the epistemic and phenomenological character, trying to implement the relevant ontological and cognitive description of AI. The third approach is plunged into ethical and anthropological context of possible AI emergence and its expected profound impact on society. The fourth and last, but not the least approach, is tied to metaphysics. This means that the general philosophical approach on AI is framed within the context of some logical and ethical investigations and their importance for AI development. The aim of this paper is to present some specific issues deriving from the fourth, i.e., metaphysical approach.

What is artificial intelligence (AI)? The notion of AI may itself depend upon the assumed definition of the intelligence itself, in general human intelligence. Thus, AI would possess at least similar (but not necessarily the same) properties as human intelligence. This leads to academic debate on AI either in terms of Weak AI and Strong AI. The former may simulate some human intellectual activities below the level of general human intelligence; the latter is similar or nearly equal to it. AI that may exceed the level of Strong AI is sometimes called superintelligence or a super-intelligent agent or an ultraintelligent machine. This leads to a possible conclusion that AI may emerge outside the human body as a super-intelligent agent, an autonomous machine especially when the possible emergence of AI is perceived in the virtual environment as a result of computation. This implies transferring human intelligence into a virtual digital system, in other words, its emulation as a programmed set of algorithms. This kind of emulation is in fact a simulation at various levels of complexity and perfection. However, it is not impossible that such AI may in return, be transferred back into human brain as a process of scheduled intellectual enhancement of humans and this may result in specific convergence of human intelligence with the artificial one. This may also be of mutual benefit, if humankind will acquire tools to make this process feasible, e.g., by invention of brain–computer interface. The said approach imposes a required level of relevant close symmetry of relations between human intelligence and AI. But at this point, a crucial question may be raised: is Strong AI actually feasible and under what conditions? One could argue that Turing

✉ Ignacy Sitnicki
igi0404@poczta.onet.pl

[1] Institute of Philosophy, University of Warsaw, Warsaw, Poland

test developed by Alan Turing is the criterion of evaluation, when AI is equivalent to general human intelligence, in other words, when human participant of the test cannot differentiate intelligent machine action from the human one. However, supposed outcomes of the Turing test as suggested by measures of machine intelligent behavior are the subject of ongoing academic debate.

Several specific ideas shall be recalled at this stage of considerations.

The first one is likely mechanical in character and has its roots in the Turing–Church approach (Krajewski 2003; Penrose 2005). It suggests that if human brain, considered as a source of human intelligence, works like a computer, it is possible to transfer its functionality to an artificial or a virtual environment, i.e., to a real computer or an autonomous agent. This approach relies on a supposed idea that any Turing machine-computable function may be computed by a more advanced Turing computational system and so forth, and in the end theoretically it would emulate whole brain functional map onto computer memory. However, this approach may be seen as a process of simulation leading to the level of human intelligence simulated by computation techniques. The conclusion seems to be incomplete, because if this simulation is going to work, then all functions of human brain must have been already discovered and quantified. However, we know that the phenomenon of human brain and its functions (e.g., consciousness, intuition, feelings), perceived as a finite set of all supposed functions up to now, are not fully discovered and even defined. So we are not able to transfer the whole brain map of neuron interactions onto a virtual environment. The Turing–Church approach (Krajewski 2003) maybe intuitively sound but practically uncertain. This may also be a consequence of the conclusions made by followers of Alan Turing and Alonzo Church (Sandberg and Bostrom 2008)—Nick Bostrom (ibid.) and Anders Sandberg (ibid.)—about the uncertainty of the feasibility of Strong AI.

The second approach has been suggested by Kurt Gödel (Krajewski 2003), who took a position that we cannot confirm or deny the possibility of AI compliance with general human intelligence. However, Gödel suggested that probably AI will never acquire the level of human intelligence, because machine cannot be equal to human mind, but on the other hand, we cannot find his clear statement that AI emergence is impossible. In spite of this, John Lucas has tried to find in Gödel's incompleteness theorem a justification of the argument that machine (i.e., Turing machine) will never reach the level of human intelligence (Lucas 1961). Roger Penrose (2005) is more moderate in his similar view, because he does not deny that maybe future breaking of discoveries will pave the way for AI emergence, but now Strong AI is not feasible. As a result, Lucas–Penrose position implies non-computable character of human mind.

The third position has been presented by Emil Post and has also been followed by John Lucas. Both deny possible emergence of AI equal to general human intelligence (Krajewski 2003). This may be considered as a radical negation of the emergence of Strong AI. Further, as it was mentioned above, Roger Penrose argues that the current computation model is not a convenient technology to create AI equal to general human intelligence and humankind shall discover (if it is possible) more advanced technology, e.g., "unsound algorithm" to make AI really feasible, otherwise like Lucas he admits that human thought is non-computable (Penrose ibid.). What is more controversial, Penrose suggests, is the idea that a quantum character of human brain activity can be perceived as neuronal interactions. But this conception has not yet been developed enough, it seems to be rather a supposition, unproved until now.

If we reject Lucas–Post strong negation of AI feasibility (Krajewski 2003) and the assumption that the so-called intelligent machine (Strong AI) will never be equal to human intelligence, we can take under consideration the three remaining positions, the Turing–Church argument about possible AI feasibility, Gödel's position about possibility to make AI feasible on nonmechanical grounds, and Penrose position about conditional feasibility of AI.

At this point of our inquiry, it is noteworthy to recall Hubert Dreyfus's position concerning substantial difficulties in making Strong AI effectively feasible. This position is grounded in the view that general intelligence is embodied and closely tied with human autonomous biological structure and consciousness. On the other hand, human intelligence in principle cannot be fully an object of symbolic representation primacy because in character it is also intuitive and instinctive, surpasses formal logic and information processing. Dreyfus rejected possibility of making AI symbolic architecture equal to general human intelligence (Haugeland 1996).

When we consider these particular positions, we may come to a conclusion that possibly a Strong AI may be feasible in the future. However, a reservation should be made that if AI means transfer of human intelligence onto a certain artificial autonomous system, it will not probably be the same general human intelligence, but rather a simulated form of general intelligence on a certain level of complexity. How similar is this intelligence—that is the question.

Also the next question may be raised as well: if similarity means that AI will never be entirely equal to human intelligence, but just nearly equal or that AI may surpass it—what may happen to humankind–AI relations if AI emerges as a dominant intelligent agent? The latter may be understood as the creation of superintelligence. But in spite of the ongoing debate about the character of AI, we can make an assumption that AI emergence is possible and this will be a real challenge for humanity.

2 Can humankind survive without AI in a long-time perspective: existential *dilemmas*

It is rather a weak argument that humankind is able to continue its mission on the Earth and beyond for a far long period of time. Whatever is our opinion about soundness of disputed *Doomsday Argument* (Bostrom 2002), it is beyond any doubt that our civilization has not been given forever. It is not eternal and cannot last towards infinity. Even now there are plenty of actual and eventual existential risks that may endanger our world. One of the possible ways to avoid final existential disaster is colonization of the outer space and escape from our planet. However, our biology is closely tied with our biosphere; it is hard to believe that humankind in its actual physical condition is in fact able to explore remote Universe. This task requires different qualities of human body and mind, and also a new hypothetical tool that may help to implement such an enhancement is AI. It may be a sine qua non condition of human enhancement, no matter if we imagine AI as some human-like creature or just a set of information entities able to travel across unimaginable distances.

Let us accept the proposition that every individual human being is a contingent entity, simply because its nature is not absolutely perfect—in fact likely to be quite imperfect. We are mortal, fragile and not made for living long out of biosphere. However, such contingency cannot be applied to the entire humankind. But in spite of the latter unsolved question of contingency from the metaphysical point of view, every individual being is in opposition to *Ens perfectissimus*—a supposed absolute perfect being according to Aristotle. For Aristotle, this perfect being is a Divine Being or God. But the aim of this paper is not plunge into this very specific dispute about the existence of God or validity of various ontological arguments. Abolishing contingent being and necessarily perfect being antinomy means becoming as perfect as *Ens perfectissimum*, i.e., an absolute (abstract or real) being, containing all perfections, but this seems to be impossible for a human being. It is just the argument about human imperfections and limitations, which we can experience in our every-day life. To surpass these limitations as far as it is possible, human race must reshape its physical and intellectual condition. We can use digital technologies, genetic engineering, nanotechnology or even future technologies that have not been invented until now, as well as AI. The latter is a subject of our considerations. I argue that AI is an essential factor for enhancement of human intellectual and physical abilities. I argue that for implementing enhancement of human mind and body, a quasi-symmetry between human intellect and AI is required. This imposes a condition that AI cannot excel human intelligence if we do not want to initiate a collapse of humankind. If it happens, a supposed superintelligence would emerge and this may cause a real problem for our survival as we will not be able to control the ultraintelligent agent possesing the level of intelligence unavailable to us. This is a possible existential risk. And let us make it clear that if AI overcomes human intelligence level, it is very likely that humankind may extinguish like any other species losing their battle for being the relevant competitive creatures, which means losing a battle over intellectual supremacy.

The other crucial point we shall take under consideration is entropy, which from the philosophical standpoint means a certain level of disorder. The term is borrowed from physics (the second law of thermodynamics) but is adopted also for the purpose of philosophical discussion about the possibility to reverse the increasing level of entropy. It is quite an acceptable assumption that more complex systems create more increasing level of entropy. The antonym of entropy is extropy and today the latter term has been adopted within contemporary philosophical debate and may mean the process of increasing order within complexity of evolving systems. The problem from the point of view of complexity of human biological structure is how to stop and reverse the process of increasing entropy which leads to diseases and death. We can observe that inflammations, infections, diseases and other disorders result in an increasing level of entropy which in the end can produce final collapse of our biological structure. From the perspective of how to prolong significantly our life, the turning point is about the problem of exporting entropy from our biological structure to outside and importing more extropy into it. In fact, this process is developing through the implementation of medical and genetic treatments, but there is no breakthrough of the expected results up to now. In the near future, we can use more advanced procedures for implementing discoveries in genomics and nanotechnology. But it is possible that in the future AI will also be helpful in the process of decreasing entropy inside us and increasing extropy in our biological structure. I argue that AI development is a sine qua non condition and helpful tool for decreasing level of entropy in our human structure, which may lead to a radical life extension. Radical life extension means more potential of our body and mind in efforts to explore our galaxy and the more distant Universe. It may mean also a certain problem for society, but this is not a subject of this paper.

The last question, but worthy of consideration, is that of a specific Posthuman Paradox (PP). Let us assume that our goal is to colonize outer space. We do so because after all the Universe is today the actual observable environment we are exploring step by step. This exploration ability is limited by our biological imperfections. But the proposed Posthuman Paradox is closely tied with our biology, culture and religious beliefs because we are afraid of changing our life radically as we fear to create quite different posthuman condition of ourselves. All these desires and fears lead to a paradox because at the same time we want and do not want to excel our fundamental biological limitations. This specific paradox is like living in a golden cage of our human values built on the grounds of secular humanism, religion, imposed ethical principles and social order. This is, in fact, a kind of static captivity and a trap leading us to a doomed future. However, this Posthuman Paradox may disappear in the future when technology reaches the level that may help us to change safely our imperfect human condition. I argue that AI may play an important role in such a process. Additionally, I argue that safe passage from human to posthuman transformation of our body and mind requires nearly equivalent relation between human being and AI as well as quasi-reflexive, quasi-symmetric and quasi-transitive relations between human being and the intelligent agent. If this equilibrium collapses, the PP will not be abolished as the fear of something strange and unknown prevails in the positive approach to AI development and may cause the end either of humankind or AI. The debate regarding this specific equilibrium between humankind and AI may be the subject of research within theory decision investigations.

3 Metaphysics, modality and arrow of civilization evolution

Leibniz famously pronounced an idea that our actual world must be the best of all possible worlds (appeared in the mind of God). This is a clear supposition based on theological grounds and has very little to do with logical soundness and perhaps is rather logically undetermined (with no logical value tied). With some reservations it may be considered as a specific speculation within metaphysical context. But if we complete it with another famous question once raised by Leibniz: "why is there something rather than nothing" (in *De rerum originatione radicali*, 1697—*On the Ultimate Origination of Things*), we can assume the previous statement in some way justified as we are living in one of the possible worlds, which is a real one for us and thus may be considered the best one we can observe as an actual world. However, following Leibniz narration, it may be a contingent world. The latter supposition is additionally justified by the anthropic principle conclusion that our biosphere is fine-tuned and is the only known environment where the life emerged in the form we can observe and where the life is continued (Barrow and Tipler 2009). On the other hand, a certain reservation is required as this premise may collapse if we assume the soundness of David Lewis conception of the so-called modal realism (Stalnaker, 1976). However, the modal realism idea is far beyond the subject of this paper. But the extension of considerations mentioned above will lead us to a modal logic anyway and possible world semantics. As a rule, possible world semantics is about the provability of modal claims (propositions) where possibility $[\Diamond P \leftrightarrow \neg \Box \neg P]$ means truth in some possible world and necessity $[\Box P \leftrightarrow \neg \Diamond \neg P]$ means truth in every possible world. What is of significant importance for the question, why AI shall emerge in one of the possible worlds, is Hintikka's discovery that it is quite unnatural to assume that the multiplicity of possible worlds we shall consider within a set of possible worlds is limited (Hintikka 2014). Otherwise, the multiplicity of possible worlds we shall take under consideration in possible world semantics modal inquiry is unlimited. This of course is not a conclusion that is taken under consideration by the David Lewis proposition of modal realism, but rather a premise tied with Kripke's discoveries, i.e., a logical and metaphysical approach of possible worlds. Additionally, Stig Kanger's analytic approach (Lindström 1998) may be helpful as well. And if we assume Hintikka's premise about unlimited number of possible worlds in modal system of alternative sets (Hintikka 2014), then it is an efficient cause to conclude that it is possible that AI may emerge in the one of analyzed possible worlds. And if we assume that our world is a real alternative one within a set of all possible worlds, there is also a possibility that AI may emerge in our actual world. That is of essential metaphysical importance about AI question on how to deal with AI when AI will emerge as a feasible entity. There is no necessity of AI emergence in our actual world, but the possibility of it is logically justified, hence we must be prepared for a supposed profound impact of AI emergence on our society. Moreover, according to the anthropic principle, we must be prepared for our specific position as intelligent observers of the Universe. We also must be prepared for the observation of AI emergence in our actual world.

Let us consider the following observations concerning a proposed "arrow of evolution" of our civilization as a metaphor arrow of time. In macroscale, it is not a disputable observation that arrow of time leads from one point in the past to another in the future, i.e., the arrow of time is moving forward, but in microscale. In the context of quantum physics theory, this flow may not follow the same rule, as it may not be observable moving only in one direction forward, and subsequent observable events may happen in different positions at the same time. The similar pattern can be observed in the process of civilizational evolution. In microscale, we can observe arrow of evolution unstable, going forward or backward or remaining in stagnation, but in macroscale we can observe a constant progress of the civilizational evolution initiated by *Homo sapiens*. The evolution of civilization in macroscale is not a kind of jumping forward and backward. We are moving forward, not living at the same time in Stone Age and in technologically developed society.

However, we can realize big differences in levels of development in various local societies. Of course, our civilization may collapse or even disappear like other past civilizations did, but in principle we are a progressing civilization. This may raise, as a kind of side effect, a disputable question that maybe time is a kind of illusion or our life is a simulation. But this controversial idea is not a subject of our considerations. On the other hand, if the arrow of evolution in macroscale is a developing process, we cannot deny a possibility that AI shall emerge in one of the possible worlds at a certain point of time. This claim is about temporal modality. If humankind will die out before this expected point of time comes true, we do not mind AI emergence. Considering our world, i.e., our planet and the observable Universe, as the only world we can experience until now, the consequence is that we cannot deny that AI may emerge in our actual world. The conclusion is that it is probable and we can presume that it is possible at a certain point of time on a scale of civilizational evolution that AI shall emerge in our world as the consequence of subsequent positions of arrow of time and we can predict this possible event from the standpoint of an intelligent observer.

4 What may possibly happen beyond AI existential risk horizon: society after AI emergence

The key point about possible AI emergence and its impact on the future of human society and order of social arrangements is the question of how dangerous for humankind is existential risk that is tied with AI emancipation and possible supremacy of superintelligence.This breaking moment of technological turning point is usually called a singularity. Singularity is a conception of various meanings, but as a rule, implies possible losing control over AI. Hence, the principal problem is about how to control AI development and make it a human-friendly entity in moral sense, free of all negative properties, that may endanger humanity. In addition, how to avoid the creation of a reality, where superintelligence excels human general intelligence level, causing a risk of AI domination over humankind. This perspective is a doomed future, creating an existential risk horizon beyond which there may be no place for humanity. This means a real catastrophic scenario for our fate. Of course, we do not know what may happen beyond the existential risk horizon as we probably will not be a subject of these unknown possible events. It looks like an existential "black hole" which may destroy a civilization made by humankind. This is a perspective about the disappearance of humankind, which means the existential nothingness from metaphysical point of view.

At this point, we can assume that a crucial task concerning possible AI emergence is to subject AI development to positive moral and aesthetical values, and this means a close convergence between morally enhanced humanity and AI. If the gap between human and AI moral values will be inconveniently big, the existential risk of AI supremacy will be increasing. With reduction of this gap to minimum, the risk will be decreasing. But all efforts to reach such a result mean moral enhancement of humankind itself as we assume that development of AI is about simulation of our own properties. This is, however, a question about metaethics and how to change our ethical values to a more perfect system. It seems to be quite acceptable to build such a system on moral principles that have their roots in humanism and open society principles, as well as, in bioethics and good practice (in moral sense) rules. I also think that a cooperation with religious groups all over the world is needed. We shall bear in mind Einstein's remark that science without religion is lame and religion without science is blind. This is not a question about science–religion controversy, but of a worldwide society, in the sense that we have to take under consideration a certain burden of our history and tradition.

If the required level of ethical perfection is achieved, the possible AI may be considered as an equal partner to humankind, but under one fundamental condition: the level of AI cannot surpass the level of general human intelligence. If it surpasses, the human race will extinguish as it had happened in the past when intellectually weak humanoids disappeared paving the way for more intellectually developed *Homo sapiens*. This is a real danger of superintelligence emergence. We cannot exclude this possibility, but it may be possible that humankind will find a path for controlling AI emancipation. This is also a question of finding relevant convergence between human being and AI. If we create this phenomenon as an alien agent, the existential risk will be very serious. To avoid this danger, the model of equal relations between humans and AI is required.

But it is not impossible to adopt AI by humans as a friendly tool for enhancement of our body and mind. The Russian philosopher and futurist Nikolai Fedorov suggested a supposed human enhancement as self-enhancement, the kind of engineered autotrophy (Fedorov 2008). If this idea might have a little sense, it may be possible to use AI as a tool for achieving this goal on the grounds of using human brain–computer interface. Further to this interface, nanotechnology, synthetic biology, genomics or other currently unknown possible technologies may be invented in the future. Fedorov's futuristic vision assumes the supposed autotrophic self-enhancement as a result of technology revolution, which will allow the use of scientific methods to eliminate our limitations and make our bodies and minds more perfect. These hypothetical methods may be supported by AI from outside and also from inside as a part of posthuman phenomenon. It is a possible claim that we cannot exclude that the supposed posthuman being may be perceived as quite a new form of human evolution created by the convergence between human intelligence and AI.

The last, but once again not least issue, is moral enhancement of human being as an individual being and the humankind as a society. If humankind will be able to achieve not only physical, but also moral perfection, as a significant enhancement, it will be possible to transfer these positive values into the intelligent machine (Bostrom 2014). It will be possible that two ethically enhanced phenomena, posthuman being and AI, may be fine-tuned with each other, creating quite new positive properties in a future augmented life. But this is a kind of disputed futuristic philosophy, far beyond the subject of this paper. However, we can try to develop such possible reasoning using metaethics and deontic logic as well as decision theory investigations.

5 Conclusions

The conclusions of the above-proposed metaphysical inquiry about possible (Strong) AI emergence are as follows:

- Multiplicity of possible worlds is unlimited.
- AI can possibly emerge in one of the possible worlds.
- Our actual world is the real possible world.
- As a humankind we must be prepared that AI may emerge in our actual world.
- AI may possibly be a real existential risk for human civilization.
- AI, if it emerges as a Strong AI, it may possibly eliminate humans from biosphere and the Universe.
- It is possible to create AI as a human-friendly entity.
- It is possible to build fine-tuned positive convergence between human being and the intelligent agent.
- The emergence of posthuman—as a result of human evolution powered by AI—is contingent.
- Allied and collaborative humanity as the worldwide society is obliged to create human-friendly AI in compliance with positive ethical values to avoid existential risk of the elimination of human race by AI.

Acknowledgements The concept of the paper is a result of discussion held during Decision Theory & the Future of Artificial Intelligence workshop organized by The Centre for the Study of Existential Risk (CSER) in summer 2017 at Trinity College, Cambridge, especially during coffee breaks and the dining time. Regarding decision theory as a significant discipline helping to develop AI, I have tried to find some philosophical aspect of discussed issues. I would like to express my heartily thanks to Professor Huw Price, Professor Stephan Hartmann and Dr Yang Liu for giving me the opportunity of participation in the most interesting academic event. For inspiring discussion helping me to clarify my position about AI phenomenon, I am grateful to the entire group of workshop participants, among them: Professor Karamjit S Gill, Professor A. R. Briggs, Professor Thomas Icard, Dr Arif Ahmed, Dr Adrian Currie, Dr Pedro Ortega, Dr Ramana Kumar, Dr Jiji Zhang, Dr George Nguyen, Dr Rune Nyrup, Dr Mostafa Mohajeri and Tsvi Benson-Tilsen.

References

Aristotle (Arystoteles) (1984) Metafizyka (metaphysics), Warszawa

Barrow JD, Tipler FJ (2009) The anthropic cosmological principle, Oxford

Bostrom N (2002) Anthropic bias—observation selection effects in science and philosophy, New York

Bostrom N (2014) Superintelligence: paths, dangers, strategies, Oxford

Chalmers DJ (2010) The singularity: a philosophical analysis. J Conscious Stud 17:7–65. http://consc.net/papers/singularity.pdf. Accessed 12 Sept 2017

Dreyfus HL (1965) Alchemy and artificial intelligence. https://www.rand.org/content/dam/rand/pubs/papers/2006/P3244.pdf. Accessed 31 Jan 2018

Fedorov NF, (Федоров Н. Ф.) (2008) Философия общего дела (The Philosophy of the Common Task), Москва

Gödel K (1995) Collected works, volume III, unpublished essays and lectures, Oxford

Haugeland J (1996) Body and world: a review of What Computers Still Can't Do: A Critique of Artificial Reason (HubertL.Dreyfus). Artificial Intelligence 80 http://philosophy.uchicago.edu/faculty/files/haugeland/dreyfus.pdf. Accessed 31 Jan 2018

Hintikka J (1969) *Eseje logiczno-filozoficzne* (logico-philosophical essays), Warszawa 2014 (extended Polish edition of Models for Modalities: Selected Essays, Dordrecht)

Krajewski S (2003) Twierdzenie Gödla i jego filozoficzne interpretacje. Od mechanicyzmu do postodernizmu (Gödel's Theorem And Its Philosophical Interpretations: From Mechanism To Postmodernism), Warszawa

Kripke S (2013) Philosophical Troubles, Oxford

Kurzweil R (2016) The singularity is near, London

Leibniz GW (1967) On the ultimate origination of things http://www.leibniz-translations.com/ultimateorigination.htm. Accessed 22 Dec 2017

Leibniz GW (1985) Theodicy. [Open Court], Peru

Leibniz GW (2001) Nowe rozważania dotyczące rozumu ludzkiego (new essays on human understanding), Kęty

Leslie J (1996) The Anthropic Principle Today [in: Final Causality in Nature and Human Affairs], Washington

Lindström S (1998) An exposition and development of Kanger's early semantics for modal logic. In: Humphreys PW, Fetzer JH (eds) The new theory of reference—Kripke, Marcus, and its origins, Kluwer

Lucas JR (1961) Minds, machines and Gödel, Oxford

Łukasiewicz J (1957) Aristotle's syllogistic from the standpoint of modern formal logic, Oxford

Penrose R (2005) Shadows of the mind: a search for the missing science of consciousness, London

Sandberg A, Bostrom N (2008) Whole brain emulation: a roadmap (http://www.fhi.ox.ac.uk/reports/2008-3.pdf). Accessed 1 Oct 2017

Solomonoff R (1985) The time scale of artificial intelligence: reflections on social effects. Human Syst Manag 5:149–153. http://world.std.com/~rjs/timesc.pdf. Accessed 14 Sept 2017

Stalnaker RC (1976) Possible Worlds, No&ucirs:s, Vol. 10, No. 1, Symposium Papers to be Read at the Meeting of the Western Division of the American Philosophical Association in New Orleans, Louisiana, April 29–May 1, 1976, pp 65–75

Wang H (1996) A logical journey. From Gödel to philosophy. MIT Press, Cambridge

Yudkowsky E (2008) Artificial intelligence as a positive and negative factor in global risk. In. Global catastrofic risks, Oxford, pp 308–345