Marking the lines of Artificial Intelligence

Mario Verdicchio[1,2] (ORCID 0000-0002-4377-8619)

[1] University of Bergamo, Italy
[2] Berlin Ethics Lab, Technische Universität Berlin, Germany

Abstract (150 words)
Artificial Intelligence (AI) has the goal of recreating human intelligence by means of computers. Despite significant successes in specific contexts, this endeavour is problematic in a number of ways, which can be all illustrated with the metaphor of marking lines: lines to circumscribe and define, lines to connect and make comparisons, lines to separate and make distinctions, and, finally, lines to trace a path towards the future.
In AI, these lines are blurred, because without a clear definition of intelligence AI is also difficult to define, the analogies between humans and machines are weak, the reliance on digital computers makes it hard to distinguish AI from the rest of Computer Science, the future of AI is often illustrated by means of Sci-Fi scenarios with little connection with reality. This work is meant as a call for more attention on behalf of those who use, research, write, and talk about AI.

Keywords: computer science; computational models; definitions; humans vs machines; philosophy of mind; philosophy of technology.

Introduction

One may mark a line for many purposes: delimiting an area, connecting two points, separating two sectors, tracing a path, and so on.
I will here adopt this versatile graphic gesture as a metaphor to discuss a number of aspects of Artificial Intelligence (AI): how it is defined as a discipline; how, as its very name seems to suggest, it stems from an attempt to recreate human intelligence by means of artificial entities; how it is distinguished from other efforts in Computer Science; finally, what we may expect from it in the future.
All these efforts are meant to provide a better understanding of AI but, as I will try to show in what follows, the lines that are meant to circumscribe, connect, separate and orient AI are rather blurred, which should serve as a call for more attention on behalf of those who use, research, write, and talk about AI.

Marking the lines as definition

The very definition of AI has raised several debates ever since the term "Artificial Intelligence" was first used in 1955 in a proposal for a summer workshop at Dartmouth College (New Hampshire, USA) by those who are considered today the founders of the discipline. In that proposal, it is written that "The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it." (McCarthy et al., 2006, p.12). This conjecture seems to lay the ground for a definition of AI as the discipline dedicated to a precise description of features of intelligence, so to enable the simulation of those features by a machine.

I will get back to the meaning and implications of a "simulation by a machine" in the next section. Let us start with the concept of intelligence used in the proposal. There is an assumption that "learning" is a feature of intelligence, which is not controversial, since it is accepted in many other fields than AI, such as psychology (Piaget, 2003) or pedagogy (Novak & Gowin, 1984). Much more problematic is the other assumption on intelligence, according to which intelligence is amenable to precise, machine-compatible descriptions. The first author of the proposal, Prof. John McCarthy, has stood by this assumption his whole life, as shown by a manifesto in the form of a Q&A he first published in 2004 and last revised in 2007, in which his answer to the question "what is intelligence?" is the following:

"Intelligence is the computational part of the ability to achieve goals in the world. Varying kinds and degrees of intelligence occur in people, many animals and some machines. (McCarthy, 2007, p. 2)"

This is, indeed, marking a strong line around the concept of "intelligence," which is deemed inherently computational by McCarthy. Agents have the ability to achieve goals in the world. Intelligence is the computational part of such ability, that is, that part of an agent's action plan that consists of operations on numbers. This is such a bold statement that the very document begins with the disclaimer: "The opinions expressed here are not all consensus opinion among researchers in AI." Discussions on what intelligence is range from anecdotal recounts to common sense to full-fledged scientific theories. Would a person who is able to perform computational operations at remarkable speed without ever making a mistake be considered intelligent if they are not able to learn the most basic rules for living in society? Among the most famous theories on intelligence that include more than computation, there is psychologist Howard Gardner's theory of multiple intelligences (Gardner, 1983), which differentiates human intelligence into eight specific modalities: visual-spatial, linguistic-verbal, logical-mathematical (presumably the one McCarthy is after), bodily-kinesthetic, musical, interpersonal, intrapersonal, and naturalistic.

These two views on intelligence, one solely computational and the other multi-faceted, seem to set up two different scenarios, each with a definition of intelligence that defines the scope of AI as a discipline. In the first scenario, McCarthy is wrong or at least narrow-minded on intelligence, which exists in modalities that are not computational. Rather than Artificial Intelligence (AI), we should call his efforts Artificial Computational Intelligence (ACI). In the second scenario, McCarthy is right, and all forms of intelligence can be traced back to

computational processes, and psychologists like Gardner propose a framework with multiple forms of intelligence because their underlying computational foundations have not been discovered yet.
Who is right? Which scenario is real? There is no definitive answer to these questions, but surely a lot of energy is devoted in the context of AI research to pursue either vision. In particular, many researchers who believe McCarthy's definition of intelligence are part of a subfield of AI in which the aim is to build machines that perform any task that a human being is capable of. Since the range of human intellectual capabilities is so vast and general, this research effort is called Artificial General Intelligence (AGI, Goertzel & Pennachin, 2006). Whether AGI will one day succeed remains to be seen. What is interesting now in AGI, at least within the scope of this analysis, is its focus on the comparison between human intelligence and machine intelligence.

Marking the lines as comparison

Formally, an analogy between two entities A and B is a one-to-one mapping between objects, properties, relations, and functions in A and those in B. Not everything in A must be put in correspondence with relevant items in B: an analogy is comprised of correspondences only between a subset of characteristics (Bartha, 2022). Clearly, the analogy underlying AI is between a human being and a computational machine. However, which aspects are to be involved in the analogy is disputed and this leads to different variations of AI: general, narrow, strong, and weak.
The general vs narrow and the strong vs weak contrapositions are orthogonal but not completely independent, and their connection takes us back to McCarthy's reference to simulating human intelligence with a machine. Indeed, a simulation imitates one process by another process (Hartmann, 1996), that is, a simulation is inherently based on analogies between two processes, the one that is simulated and the one that simulates. If human intelligence is to be simulated by means of a computational machine, what aspects are to be reproduced inside the simulating machine? This is where the abovementioned contrapositions show their orthogonality.
AGI and narrow AI (sometimes called Artificial Narrow Intelligence, ANI, Fjelland 2020) are about quantity, that is, the quantity of tasks that an AI machine must be able to perform. In AGI, the goal is the most ambitious: all tasks a human can perform, across the whole range theorized by Gardner, must be described in computational terms so that a machine can execute them. In ANI, the context is, indeed, narrower: a machine is built to execute one specific task, or a very small set of tasks. If AGI's realizability is still debated among researchers, there have been several extremely successful ANI projects in a number of fields, including games (Schrittwieser et al., 2020) and medicine (Kourou et al., 2015).
Strong and weak AI, on the other hand, are about quality, not in terms of perfect executions and lack of errors by the machine, but in its original meaning of "quale," of how a certain situation feels like for the agent in it, as a subject capable of perceiving the features of that situation. Proponents of strong AI believe that it is in theory possible to build a machine that entertains conscious experiences the way humans do, whereas advocates of weak AI believe that there is a deep ontological difference between human brains and computing machines, and only the former have the characteristics that make them capable of perceiving qualia.

Scientists agree that the human nervous system makes the perception of qualia possible, but how that first-hand, subjective sensation emerges from human physiology remains a mystery so deep that it is called the “hard problem” in philosophy of mind (Chalmers, 2007). In weak AI, this is where the analogy stops: we can build more and more sophisticated computing machines that perform more and more tasks that have been traditionally performed by humans, but consciousness will forever remain an elusive feature of the human experience that escapes computational modeling.
An analogy is indeed the most famous attack against the idea of strong AI, provided by philosopher John Searle, who proposed the thought experiment of the “Chinese room” (Searle, 1980), in which he imagined himself inside a room, processing messages from the outside written in Chinese ideograms on the sole basis of their appearance (since Searle does not understand Chinese), formulating answers following visual message-reply rules written in a ledger, and sending out replies that make sense in Chinese, thus giving the impression that the room understands Chinese to the people on the outside. In this analogy, Searle designed a limited human experience, that is, processing only the signs a message is comprised of but not its meaning, to give us an idea of how computing machines work: they crunch signs, i.e. numbers, in accordance with their values and some rules, but they do not have a mind that can associate ideas and concepts to those signs.
Another philosopher, Hubert Dreyfus, uses similar arguments to attack AGI: the subjective experience that humans have thanks to their consciousness is not only necessary for humans to entertain meanings, but it is also a fundamental ingredient to form what is known as common sense. Based on their past experiences, humans are able to draw analogies, tackle new situations successfully and master the complex game of life. Not everybody is successful in the same way, but everybody has the potential to learn any kind of task that is compatible with human nature. This general intelligence is possible only to conscious human beings, whereas coding all possible real-life situations in a computing machine is unfeasible (Dreyfus, 1992). Here, quality and quantity meet: we need the qualitative experience of consciousness to unlock the power to learn a potentially infinite quantity of tasks.
Years after the introduction of the Chinese room, when asked whether he considered strong AI a logical impossibility, Searle left the door open, again relying on an analogy:

“(…) the human brain is a machine, a biological machine, and it produces consciousness by biological processes. We will not be able to do that artificially until we know how the brain does it and we can then duplicate the causal powers of the brain. (…) at present we do not know enough about the brain to build an artificial brain.” (Turello, 2015)

His attack against the idea of conscious computers can be thus framed in a technological context: in the way computers are built today no emergence of consciousness is possible because the analogy is leaving some key features out. Of course, this is not a problem for those who pursue the less ambitious but still potentially very impactful goals of weak, narrow AI. However, focusing on the electronic, digital computing machines in use nowadays sheds light on another threat for AI, strong, weak, general, or narrow: the lines indicating its borders and distinguishing it as a discipline seem to disappear.

Marking the lines as distinction

Computing machines are a good example of multiple realizability: to perform computation we have different choices on what kind of physical devices to build to represent numbers and perform operations on them. These choices have expanded through the centuries. The abacus, as it appears today, made of wood and metal reinforcements, was first chronicled in 13th century China. To support his father's activities as a tax accountant, Blaise Pascal invented the first mechanical calculator with rotating metal gears, known as the Pascaline, in 17th century France. Charles Babbage modified a Jacquard loom and transformed it into the Difference Engine, a machine capable of raising numbers to the second and third powers and calculating the solution to specific quadratic equations in 19th century England. A much more sophisticated mechanical computing machine was created by Konrad Zuse in early 20th century Germany, with the groundbreaking additional feature of programmability, i.e. not only the data but also the operations to be performed on the data can be stored in the machine, with dramatically increased possibilities for automation. At the same time, with the invention of the vacuum tube by John Ambrose Fleming, controlling the flow of electricity through electronic components became possible, and led the way to the first electronic computers, like the one by John Vincent Atanasoff, conceived in 1937 and released 5 years later at the Iowa State College, USA. The discovery of semiconducting materials, that is, materials that enable or block the flow of electricity depending on the voltage they are stimulated with changed the computing game forever: American physicists John Bardeen, Walter Brattain and William Shockley invented the transistor in 1947, thus enabling an unprecedented miniaturization of the switches controlling the flow of electrons inside a computing machine. Their invention, which won them the Nobel Prize in Physics in 1956, is the reason why we can carry very powerful computers in our pockets today. (Campbell-Kelly et al., 2016)

Despite the enormous technological variety, in terms of design, materials, physical phenomena involved, there is a general paradigm, guiding the construction and use of a computing machine, which characterizes all the above-mentioned devices, including abaci and the latest, fastest digital computers. There are obviously radical differences in performance and levels of automation, however, the principle of operation is the same: whatever task is at stake, it needs to be encoded, that is, described in the form of numbers; some components of the machine are used to represent those numbers; the numbers are processed by the computing machine, that is, the components representing the numbers are modified in accordance with some operations on the machine (and by the machine itself, if it is programmable); finally, when the operations are over, the machine reaches an end state, and the numerical quantities represented in it are the numerical output that needs to be decoded, that is, translated back into results for the task.

From this very general perspective, using a computer to sort a list of names or using it to simulate human intelligence do not seem to make a significant difference, since both activities boil down to the same kind of encode-execute-decode sequence of operations. What, then, distinguishes AI from other branches of Computer Science, like Software Engineering (SE) or Database Theory (DB)? Indeed, should Computer Science even have subfields to begin with?

The traditional distinction in Computer Science between hardware and software may be a good starting point. Hardware is physical: it is the material machinery we build to perform computation. Software is more abstract because it is the description of the configurations to

give the hardware to perform computation. The characteristics of the hardware determine its possible configurations and hence limit the scope of the software. For instance, we cannot look at a digital image on an abacus, because the abacus is missing the hardware to produce the luminous and colourful pixels that constitute an image.[1] Keeping in mind the brief and non-exhaustive sketch on the history of computing machines given above, we appreciate at least two lines along which computing hardware can improve: it can allow for new kinds of operations (e.g. processing digital images) and it can allow for a faster execution of those operations (e.g. as of March 2022, the fastest mass-market computer chips work at a frequency of 5.5 GHz, which means they can perform 5.5 billion elementary operations per second, Chacos, 2022).
Since all computer scientists, be them AI, SE, or DB researchers, use the same kind of hardware, the distinction between their subfields, if it exists, must come from the software, that is, the operations they chose to make their hardware perform. We need to look for criteria to classify some operations as AI software to distinguish them from SE software and DB software. We are circling back to the issue of defining what AI is.
If we look at definitions that came after McCarthy's, we notice that the focus on what humans do is always there. According to the Encyclopedia Britannica:

"The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans" (Copeland, 2022).

Computer Science professor Wolfgang Ertel is rather critical of this kind of definition, because it fails at distinguishing AI from the rest of Computer Science: after all, remembering large quantities of text and computing numbers itself are intellectual processes entertained by humans, and hence according to this definition every computer would be an AI system (Ertel, 2017). Ertel considers the following definition by Elaine Rich far superior:

"Artificial Intelligence is the study of how to make computers do things at which, at the moment, people are better" (Rich, 1983).

I join Ertel in praising Rich's definition because it introduces so many dimensions in the discourse on AI with so few words. First of all, it refers to a comparison or rather competition between humans and machines that was first introduced by Alan Turing (an *ante litteram* pioneer of AI) in his attempt with the "imitation game," a thought experiment where a chatbot makes a person believe they are talking with a human, to define a criterion to recognize intelligent machines (Turing, 1950). Secondly but no less importantly, this definition frames AI as a moving target, where the movement is not only determined by the technological development of computing hardware, but also by the change in what is considered an inherently human intellectual activity. That change is significantly influenced by the very technological developments of AI.
From this perspective, memorizing texts and computing were inherently human activities at a time when computers did not exist or were extremely rudimentary and slow machines, but not anymore. Now that, in these tasks, computers outperform humans by the billions, something else is considered inherently human, and the focus of AI has shifted accordingly.

[1] In general, for digital images the encoding is based on standards that create a correspondence between numbers and the levels of red, green and blue of each pixel, and between a system of numerical coordinates and the position of each pixel within the image.

Thus, we can distinguish AI from other branches of Computer Science because of its dynamic nature: always at the forefront of computational modelling of human intellectual activities and tasks, AI tackles yet unsolved problems, only to crack them and transform them into ordinary Computer Science software and move on. Trying to understand where AI is moving towards leads us to the last metaphor with a line: a path to the future.

Marking the lines as extrapolation

Trying to predict the future of AI is an integral part of AI itself: the research efforts are inherently future-directed under the sign of an ever more comprehensive computational modelling of human intelligence. After all, AGI and strong AI are subfields about intelligent machines that do not exist, or do not exist yet. The risk here is to write science fiction rather than predicting future developments of AI research. Indeed, a significant amount of Sci-Fi stories involve AI entities that have reached sentience and help humans or rebel against them. Obviously, these stories never provide a scientific explanation on how computational machines have reached the ultimate human feature of full consciousness, but it is interesting to notice that Sci-Fi writers have imagined both software-based and hardware-based ground-breaking discoveries: in the TV series “Humans”, for instance, humanoid robots become fully conscious thanks to a special code that is uploaded on the Internet (Brozel, 2016), whereas in the “Terminator” franchise, machines make that leap thanks to a particular chip (Cameron, 1991).

Unfortunately, there is a non-negligible amount of futuristic AI research that focuses on those fictional end results without providing solid justifications for such a jump. Books like “Superintelligence” (Boström, 2014) go into the details of how a super intelligent computer that achieves sentience might elaborate a strategy to overtake the world without providing any indication on how such super intelligence might come to be in the first place. Another AI endeavour that is undistinguishable from Sci-Fi is the concept of “singularity”, proposed by Ray Kurzweil (2014), according to whom the pace of technological change will increase to such an extent that biological and machine intelligence will merge in the next step in human-machine co-evolution, where human life will be irreversibly transformed.

Many of these imaginings are built on top of extrapolations over the extraordinary development of computing technologies during the 20th century. One of the most famous examples is Moore’s law, named after Intel’s co-founder Gordon Moore, who observed in 1965 that the number of transistors in a chip doubled every year thanks to improvements in miniaturization technology (Moore, 1965). Despite some adjustments in the following decades, the pattern detected by Moore seems to hold. What can we make of this? There are at least three observations that should prevent us from jumping to apocalyptic or utopistic conclusions about the future of AI.

Firstly, there are physical limits to hardware given by the laws of physics all material entities are subject to. It is true that transistors can become smaller and smaller, but they cannot be smaller than one atom. The curve of the number of transistors per chip can be modeled as an exponential, but there is a cap (Kish, 2002). More generally, we must not take a mathematical model as a realistic depiction in all its parts, so even if AI technology and digital technology have shown exceptional growth in the past decades, this may not be the case in the future.

Secondly, we must not forget about the quantity vs quality dichotomy: an increase in transistor density surely leads to faster computing machines, which in turn means that a greater quantity of computational operations can be carried out per unit of time, but this does not entail that certain tasks will become amenable to machine simulation. There is a distinction between unfeasible and impossible tasks: an unfeasible task is one for which there is a computational solution, but it requires so many computational resources that it is not reasonable to tackle it; an impossible task is one for which there is no (known) computational solution. Breaking a cryptography-based protection is currently unfeasible, but it may become much easier once quantum computing, i.e. computation exploiting quantum mechanics phenomena, becomes available thanks to a technological breakthrough (Denning, 2019). Computing consciousness is, instead, an impossible task, since we do not know how consciousness is produced in the brain, nor whether that mechanism can be simulated via computation. Increasing the number of operations that a computing machine performs in a unit of time will not change this.
Thirdly, we must not forget that computers and AI systems are, like any other technological endeavor, an industrial product, entangled in a world-wide network of supply chains, economic interests, political strategies and, ultimately, people (Crawford & Joler, 2018). Computing machines might become more and more sophisticated and able to service humanity in ways that are today only in the realm of Sci-Fi, but who will be the real beneficiary of such technological enhancements? There are futurologists who envision a future where humans and robots coexist, the latter taking over all the heavy lifting of labor (Bastani, 2019) or even substituting other humans as perfect love companions (Hauskeller, 2016). Apart from the usual lack of any scientific explanation on how such results might be achieved, these authors fail at telling us who is going to finance the enormous technological efforts needed to build such machines, and who is going to be able to afford to enjoy those machines, if one can even imagine enjoying living in such a particular world.
To avoid encroaching on Sci-Fi, a more down-to-earth approach to imagine the future of AI may be to observe the present of AI, which, despite an apparent focus on “learning”, is profoundly different from what McCarthy had envisioned in 1955. The AI of the origins, now known as GOFAI, good old-fashioned AI (Haugeland 1989), was characterized by a rule-driven, top-down approach that aimed at encoding knowledge into a computer in the form of axioms and inference rules that simulated deductive reasoning in humans. Nowadays the dominant paradigm in AI is the data-driven, bottom-up approach of Machine Learning (ML, Jordan & Mitchell, 2015). In ML, computers are programmed to search for patterns, schemes, and general laws among vast quantities of data by means of statistical inductive processes. These processes are implemented in the form of complex mathematical functions whose parameters are modified in accordance with how well their outputs meet the goals for which the system was built in the first place. These goals are usually the completion of tasks of data classification (e.g., of digital medical images), clustering (e.g., of viewers of a streaming service), and outlier detection (e.g., of suspicious credit card purchases). The role of AI researchers has radically changed in this paradigm shift from GOFAI to ML: they do not program data and operations into computing machines, but feed data to mathematical functions until they are able to process data in accordance with the goals. In GOFAI, humans make the rules to achieve goals, whereas in ML humans only set the goals, while the rules are developed automatically inside the mathematical functions running in the computing machines. The operations in an ML system are too complex for human programmers to keep in check. The only facet humans can control is whether the ML

system has reached the goal. Therefore, ML systems are called “black boxes”: humans can only see what goes in and what comes out, but not what happens in-between. When it comes to ML, there is a significant decrease in the direct involvement of AI researchers, which has important implications on responsibility in AI: who is responsible when a fully automated ML system misclassifies data and people are harmed? So far, incidents caused by ML systems have been isolated cases of machine-based racism (Grush, 2015) and deadly overreliance on autonomous driving (Baruch, 2016). However, if ML is the (near) future of AI with a more widespread adoption in different contexts and fields, one might fear that such harmful cases are going to increase as well.

Conclusions

AI aims at simulating human intelligence by means of computational models running on digital computers. This endeavour is rife with blurred areas: we do not have a clear definition of intelligence, we do not agree on what aspects of intelligence are really amenable to computational modelling, we do not have specific goals that distinguish AI from other computer-based efforts, AI seems to be a moving target that keeps on changing what we consider to be inherently human, and, finally, we do not have a clear idea on where AI is moving towards, although there are well-founded suspicions that more and more people are going to be harmed. These considerations are not meant to scare the reader away from AI, or to turn them into a technophobe. There are undeniable benefits that we can reap from the development of AI. However, I strongly believe that shedding light on the blurred lines of AI is and will always be a fundamental way to understand and contain its dangers, may they be ill-conceived metaphors or life-threatening automation failures.

References

Bartha, P. (2022). Analogy and Analogical Reasoning. *The Stanford Encyclopedia of Philosophy* (Summer 2022 Edition), Zalta, E. N. (ed.)
URL: https://plato.stanford.edu/archives/sum2022/entries/reasoning-analogy/ (Last visited in May 2022).

Baruch, J. (2016). Steer driverless cars towards full automation. *Nature*, *536*(7615), 127.

Bastani, A. (2019). *Fully Automated Luxury Communism*. Verso Books.

Boström, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Brozel, M. (Director). (2016). *Humans episode #2.8* [Television broadcast]. Channel 4.

Cameron, J. (Director). (1991). *Terminator 2: Judgement Day* [Film]. Carolco Pictures.

Campbell-Kelly, M., Aspray, W., Ensmenger, N., & Yost, J.R. (2016). *Computer: A History of the Information Machine* (3rd ed.). Routledge.

Chacos, B. (2022). Intel's Core i9-12900KS, the world's fastest gaming chip, lands April 5. *PC World*, March 28 2022. URL: https://www.pcworld.com/article/626470/intel-core-i9-12900ks-worlds-fastest-gaming-cpu-april-5.html (Last visited in May 2022).

Chalmers, D. (2007). The hard problem of consciousness. *The Blackwell Companion to Consciousness*, 225-235.

Copeland, B. (2022, March 18). Artificial Intelligence. *Encyclopedia Britannica*. URL: https://www.britannica.com/technology/artificial-intelligence (Last visited in May 2022)

Crawford, K., & Joler, V. (2018). Anatomy of an AI System: An anatomical case study of the Amazon echo as an artificial intelligence system made of human labor. [Artwork] URL: https://anatomyof.ai/img/ai-anatomy-map.pdf (Last visited in May 2022)

Denning, D. E. (2019). Is Quantum Computing a Cybersecurity Threat? Although quantum computers currently don't have enough processing power to break encryption keys, future versions might. *American Scientist*, *107*(2), 83-86.

Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. MIT press.

Ertel, W. (2017). *Introduction to Artificial Intelligence*. Springer International Publishing.

Fjelland, R. (2020). Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications 7*, 10. https://doi.org/10.1057/s41599-020-0494-4

Gardner, H. E. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. Basic books.

Goertzel, B. & Pennachin, C. (eds.) (2006). *Artificial General Intelligence*. Springer Verlag.

Grush, L. (2015). Google engineer apologizes after Photos app tags two black people as gorillas. *The Verge*, *1*.

Hartmann, S. (1996). The world as a process. In Hegelsmann R., Mueller U., Troitzsch K.G. (eds.), *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*. Springer, pp. 77-100.

Haugeland, J. (1989). *Artificial Intelligence: The Very Idea*. MIT Press.

Hauskeller, M. (2016). *Mythologies of Transhumanism*. Springer.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255-260.

Kish, L. B. (2002). End of Moore's law: thermal (noise) death of integration in micro and nano electronics. *Physics Letters A*, *305*(3-4), 144-149.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, *13*, pp. 8-17.

Kurzweil, R. (2014). The singularity is near. In Sandler, R. L. (ed.) *Ethics and Emerging Technologies*. Palgrave Macmillan, pp. 393-406.

McCarthy, J. (2007). What is Artificial Intelligence.
URL: http://www-formal.stanford.edu/jmc/whatisai.html. (Last visited in May 2022).

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, *27*(4), 12. https://doi.org/10.1609/aimag.v27i4.1904

Moore G. E. (1965). Cramming more components onto integrated circuits. *Electronics, 38*(8).

Novak, J. D. & Gowin, D. B. (1984). *Learning How to Learn*. Cambridge University Press.

Piaget, J. (2003). *The Psychology of Intelligence*. Routledge.

Rich, E. (1983). *Artificial Intelligence*. McGraw-Hill.

Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., Guez, A., Lockhart, E., Hassabis, D., Graepel, T. and Lillicrap, T. (2020). Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, *588*(7839), pp. 604-609.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences 3*, 3, 417-424.

Turello, D. (2015). Brain, Mind, and Consciousness: A Conversation with Philosopher John Searle. URL: https://blogs.loc.gov/kluge/2015/03/conversation-with-john-searle/ (Last visited in May 2022).

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*, pp. 433-460.