Historical, Philosophical and Ethical Roots of Artificial Intelligence

Article · January 2023

DOI: 10.30971/pje.v40i1.1152

Dilek Şenocak
Anadolu University
12 PUBLICATIONS **73** CITATIONS

Serpil Kocdar
Anadolu University
40 PUBLICATIONS **643** CITATIONS

Aras Bozkurt
Anadolu University
263 PUBLICATIONS **12,155** CITATIONS

Pakistan Journal of Education
Vol: 40 Issue: 1, 2023,67-90

Historical, Philosophical and Ethical Roots of Artificial Intelligence

Dilek Şenocak[*]
Serpil Koçdar[**]
Aras Bozkurt[***]

Abstract

Artificial intelligence (AI) generally refers to the science of creating machine based algorithmic models that carry out tasks inspired by human intelligence, such as speech-image recognition, learning, analyzing, decision making, problem solving, and planning. It has a profound impact on how we evaluate the world, technology, morality, and ethics and how we perceive a human being including its psychology, physiology, and behaviors. Hence, AI is an interdisciplinary field that requires the expertise of various fields such as neuroscientists, computer scientists, philosophers, jurists and so forth. In this sense, instead of delving into deep technical explanations and terms, in this paper we aimed to take a glance at how AI has been defined and how it has evolved from Greek myths into a cutting-edge technology that affects various aspects of our lives, from healthcare to education or manufacturing to transportation. We also discussed how AI interacts with philosophy by providing examples and counter examples to some theories or arguments focusing on the question of whether AI systems are capable of truly human-like intelligence or even surpassing human intelligence. In the last part of the article, we emphasized the critical importance of identifying potential ethical concerns posed by AI implementations and the reasons why they should be taken cautiously into account.

Keywords: artificial intelligence, generative AI, definitions of AI, history of AI, philosophy of AI, ethical issues in AI.

[*] Instructor, Anadolu University, Open Education Faculty, Eskişehir, Türkiye, https://orcid.org/0000-0002-5966-1976, dsenocak@anadolu.edu.tr

[**] Asso.Professor, Anadolu University, Open Education Faculty, Eskişehir, Türkiye, https://orcid.org/0000-0001-9099-6312, skocdar@anadolu.edu.tr

[***] Associate Professor, Anadolu University, Open Education Faculty, Eskişehir, Türkiye, https://orcid.org/0000-0002-4520-642X, arasbozkurt@gmail.com

On defining AI

Along with developments in the field of artificial intelligence (AI), the definitions of AI have evolved and will be subject to change based on current and future advancements. Originally referred to as the application of intelligence "that a machine can be made to simulate it" (McCarthy et al., 1955), it was later defined as "the study of agents that receive percepts from the environment and perform actions" (Russell & Norvig, 2021, pp. 7–8). However, a more comprehensive definition was proposed by the United Nations Children's Fund (UNICEF) (2021) that puts more emphasis on humans and their interactions.

> AI refers to machine-based systems that can, given a set of human-defined objectives, make predictions, recommendations, or decisions that influence real or virtual environments. AI systems interact with us and act on our environment, either directly or indirectly. Often, they appear to operate autonomously and can adapt their behavior by learning about the context. (UNICEF, 2021, p. 16)

However, artificial intelligence is a field that is developing very rapidly and has different applications in various sectors, and as such, the definitions that explain these technologies are open to change, as are the developing artificial intelligence technologies. Therefore, in order to better understand artificial intelligence, it is necessary to examine its historical development and understand the ideas and philosophy underlying the emergence of this technology. To better comprehend the concept of artificial intelligence, it is necessary to examine this technology in the context of ethics, as well as its historical development and philosophy, because it is a technology that encompasses human-machine interaction.

Method and Purpose of the Study

The current paper adopted a conventional, or narrative, review approach, a method typically used to connect the dots between a vast, dispersed collection of publications on a certain subject, linking various studies on diverse topics for reinterpretation or interconnection (Baumeister & Leary, 1997). This type of study is valuable for showcasing recent literature, and also useful to synthesize, summarize, deduce, reveal research gaps and offer recommendations for future studies (Cronin et al., 2008). In this sense, the purpose of this article is to explore historical, philosophical and ethical roots of artificial intelligence.

Historical Origins

The historical and philosophical roots of AI date back to ancient Greece, well before the development of contemporary AI as we know it (Bengio et al., 2015). Yet, the modern form of AI is highly associated with Alan Turing, who proposed the Turing Test (Turing, 1950) – perhaps the initial momentous proposal that would shape the concepts and implications of AI – and John McCarthy and his colleagues, who officially coined the term "artificial intelligence" at the 1956 Dartmouth conference (Chen et al., 2022; McCarthy et al., 1955). Since then, AI has become a breakthrough technology, and in our digitally intensive century, it is difficult to identify any aspects of our daily lives that do not involve AI: from GPS navigation systems to personalized music recommendation systems, or from chatbots for banking or financial services to adaptive battery features on phones.

Leaving aside the debate on the definition of AI, arguably, the very early idea of AI can be traced back to ancient Greek myths. The seventh century B.C. story of Talos, a colossal bronze figure created by the Greek god Hephaestus to safeguard the island of Crete; the myth of Pandora, a malicious artificial woman created on the orders of Zeus to penalize humankind for discovering fire; or Hephaestus's golden autonomous handmaids endowed with divine knowledge could be regarded as early mythical forms of AI (Shashkevich, 2019). The artificial beings in these ancient myths can be perceived as a reflection of the simple human desire for invincible, tireless, and obedient servants and clearly demonstrate the human desire to create such entities.

Since then, we have been drawn into a fantasy world where we question or explore the possibility and the limits of control over humanoid machines or non-human characters that have human-like intelligence, including emotions, egos, and consciousness. For instance, in The Wonderful Wizard of Oz (Baum, 1900), an intelligent humanoid robot, the Tin Man, acts like a human and imitates not only the intelligence of a man, but also their emotions, as in the example of Frankenstein (Shelley, 1818). Frankenstein, the monster created by Victor Frankenstein in an experiment, is a science fiction horror novel that depicts the unexpected dreadful outcomes of man's crossing boundaries – in other words, the aftermath of "playing God". Victor Frankenstein's obsession with creating life ends in forming a hideous creature that accuses his creator of leaving him alone. The relationship between the creator and creature in the novel implies that once a new technology is developed, scientists should follow its progress (Peters, 2018), which also highlights the importance of ethics and safety in the field of AI. "I, Robot", a collection of science fiction short stories by Isaac Asimov (1950), has had a significant impact on the development of the ethics of robotics and AI. In the book, all robots are designed to abide by Asimov's Three Laws of Robotics, a set of instructions to prevent robots from turning on their creators as the monster in Frankenstein. The Three Laws of Robotics, which were first introduced by Asimov in his short story "Runaround" in 1942, are:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey orders given to it by human beings, except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law. (Asimov, 2004, p. 37)

Although these laws were invented for a science fiction novel, they may reflect humanity's fear of its own creation. These fears have also been addressed in some Hollywood movies, such as Star Wars: Episode IV - A New Hope (Lucas, 1977), which depicted intelligent androids; Blade Runner (Scott, 1982), a movie about a police officer hunting escaped artificial humans or androids; The Terminator (Cameron, 1984), featuring an autonomous killing machine sent to the past to assassinate the mother of the future leader of humanity's resistance against machines trying to gain more power; and Ex-

Machina (Garland, 2014), which makes us question if an AI can ever reach human intelligence or consciousness and, if it happened, whether it would be possible to detect the existence of an artificial consciousness. These questions could be a great entry point into the nature of (artificial) consciousness and the ethics of AI.

In these works of fiction, the idea of humans creating robots or intelligent machines capable of performing difficult tasks associated with human intelligence might have inspired "the science and engineering of making intelligent machines, especially intelligent computer programs" (McCarthy, 2007, p. 2). However, the modern implications of AI owe a great deal to Alan Turing, an English mathematician who invented a computer that cracked the Enigma code in World War II. The computer's power and success in decoding the codes led Turing to question whether machines can think, and therefore, in his article "Computing Machinery and Intelligence" (Turing, 1950), he proposed the famous Turing Test, a test to determine whether a machine can demonstrate human-like intelligence. In this "imitation game", currently known as the Turing test, if a human interrogator communicating with a computer and a human cannot differentiate between the two from the answers they have given, the computer (machine) is considered intelligent.

Another initial step toward the modern concept of AI was taken in 1956 at the Dartmouth Summer Research Project on AI, hosted by Marvin Minsky and John McCarthy. This was the first time that the term "AI" was officially introduced to the scholarly world (Haenlein & Kaplan, 2019). In later years after the Dartmouth conference, the field of AI witnessed pioneering experiments in machine learning algorithm trials. In the mid-1960s, Joseph Weizenbaum's Eliza, a natural language processing program that imitated conversation with a human, and Herbert Simon, John Clark Shaw, and Allen Newell's General Problem Solver, a computer program that mimicked human problem-solving, were two of the very early examples of AI success stories (Anyoha, 2017). These expert systems could be considered the first attempts by computer scientists to develop programs that would be able to pass the Turing Test.

Computer games and the challenge of defeating professional human players with AI opponents have become one of the most significant ways of testing the progress of AI systems (Yin et al., 2023). In this sense, rule-based strategy games that require decision-making skills, such as chess, Go, or Shogi, have unsurprisingly been used to test AI advancement. One of the first examples of human-intelligent machine gaming could be The Mechanical Turk (Clark et al., 1999), which is a so-called chess-playing automaton invented in 1770 (Natale & Henrickson, 2022). Although it later turned out to be a hoax, The Mechanical Turk, which created much excitement across the globe, was advertised as the world's first thinking machine, since it was believed that it could decide independently on every move of the chess pieces while playing against human opponents (Stephens, 2023). Centuries later, two important developments, however, added a new dimension to human machine competition. In 1997, IBM Deep Blue managed to defeat the world champion chess player Gary Kasparov, and in 2015 Google AlphaGo was able to beat the best GO player Lee Sedol. While both chess and Go are based on simple rules, they are complex games. In a typical game of chess, for instance, 10^{120} moves are possible while the game of Go can have up to roughly 10^{360} possible moves, which is notable as the observable universe contains only around 10^{80} atoms (Koch, 2016). In its victory against the best Go player, Google AlphaGo used artificial neural networks to teach itself the best moves and was able to think and perform much better than a human in a game where it was thought that no computers would ever be able to defeat humans (Haenlein & Kaplan, 2019).

Perhaps one of the most groundbreaking milestones in the field of AI is ChatGPT, a conversational AI-powered chatbot by OpenAI (OpenAI, 2022). Bozkurt (2023) points out that what makes generative AI powerful is its ability to emulate the most sophisticated human invention technology, that is, language. By crafting purposefully crafted prompts, users of generative AI can produce synthetic outputs that are very akin to organic human outputs (Bozkurt & Sharma, 2023). In line with these thoughts, While Chomsky et al. (2023) asserts that generative AI incapable of understanding of language and knowledge, Harari (2023) argues that generative AI technologies can manipulate and generate language, and, therefore, have hacked the operating system of our civilization because language is also the foundation of our civilization and is the basis of nearly all human culture. Although ChatGPT is able to understand natural language and create human-like conversations based on a prompt, it is still unclear whether it will be able to pass the Turing test and demonstrate genuine human-like reasoning (Tlili et al., 2023). There is still a great deal to be learned in the field of AI, and some major issues need to be addressed. However, delving into its philosophical and ethical roots can be a good starting point to understand what exactly AI is.

Philosophical Roots

It is difficult to give a precise definition of intelligence, and therefore, the attempt to apply this ambiguous notion to machines is a challenging pursuit (Kaplan & Haenlein, 2020). Despite the lack of a consensus definition of intelligence, there are a few attempts to define this broad and fuzzy concept. McCarthy (2007, p. 2) defines intelligence as "the computational part of the ability to achieve goals in the world" and adds that there are different types and degrees of intelligence in humans, animals, and various machines, and we are not yet able to fully describe what computational processes are called intelligent. Unlike McCarthy (2007), Minsky (1985) limits the definition of "intelligence" to tasks that require human intelligence, and therefore, claims that AI refers to "the science of making machines do things that would require intelligence if done by men" (Minsky, 1968, p. v). According to Kaplan and Haenlein (2020), the concept of AI can always be seen as unattainable because once a machine carries out a so-called complex task (e.g., robot-assisted surgery, autonomous vehicles, AI-powered vertical farming), the capacity to do this task is no longer considered as a hallmark of intelligence, which is commonly known as "the AI effect" (McCorduck, 2004). At this point, the question of whether AI systems are capable of genuinely human-like intelligence or even surpassing human intelligence is a topic of debate among computer scientists and philosophers.

In his article "Computing Machinery and Intelligence," Turing (1950) outlined the standards for deciding whether a machine could be considered intelligent and put forward that if a machine could convince a human interrogator that it is human, then it could be considered intelligent (i.e., the Turing Test). The idea of intelligent machines could also be supported with the computational theory of mind (CTM). The theory proposes that the mind, covering cognition and consciousness, functions as a computational system (Piccinini, 2020), which shares similarities with a Turing machine (Rescorla, 2020). Therefore, if the human mind itself is a computational system, we could imply that a digital computer that meets certain requirements, such as sufficient storage capacity, appropriate program, and enough speed of action, can also be considered intelligent.

While the Turing Test has been a benchmark for assessing the intelligence of computers, there have been some refutations of it and CTM. The philosopher Hubert Dreyfus (1972, 1992) argues that the ability to think is more than symbol manipulation; in other words, human intuitive judgments, practical skills, and embodied knowledge cannot be fully formalized by a computer program. In short, as the mind, which is closely connected to the body and environment, is not merely a computational system, machines cannot completely emulate human intelligence.

Another well-known counterargument to CTM and the Turing Test is from philosopher John Searle. In his famous Chinese room argument (Searle, 1980), he imagines a person who is a native speaker of English and doesn't understand any Chinese, either spoken or written. He is in a room with a book of rules for manipulating Chinese symbols. He is given some Chinese sentences (input) that are passed under the door to him. He produces responses (output) to Chinese sentences by copying Chinese characters following the book of rules and passes them back under the door. In this thought experiment, Searle (1980) claims that although the person in the room is able to produce appropriate responses, which are indistinguishable from those produced by a native Chinese speaker, he doesn't actually understand any Chinese but solely processes the sentence following the rules. Genuine understanding, however, is more than manipulating the symbols in accordance with formal rules (i.e., syntax), but interpretation of these symbols (i.e., semantics). Therefore, “syntax by itself is neither constitutive of nor sufficient for semantics" says Searle (1990, p. 27), and therefore no computer program is able to genuinely think or understand since they just arrange symbols following well-defined rules. He further argues that "brains cause minds" (Searle, 1990, p. 29), which means that mind covering affect, conation, and cognition (e.g., understanding, reasoning, judging, imagining, and problem solving) (APA Dictionary of Psychology) is based on the physical processes of brain. Hence, he posits that a non-biological computer program isn't sufficient to replicate the complex human brain processes that cause mental states and processes. To conclude, Searle (1980, 1990, 2009) objects to the claims of Strong AI- suitably programmed, powerful, intelligent computer programs which literally understand and have cognitive processes and states because he claims that programs themselves aren't minds. On the other hand, he has no objection to Weak AI- computers that simulate the ability of thinking and understanding.

So, is it ever possible for AI to have a mind of its own? Or can it have consciousness, feelings, and wills along with intelligence like a human being? This issue is still a matter of debate, as the mind is an abstract, sophisticated, and difficult concept to define. It is commonly believed that despite the fact that current AI possesses a degree of intelligence and may even be better than a human being in terms of computational intelligence, it isn't conscious yet, which means it isn't aware of its own existence, how and why its actions occur (Li et al., 2021; Meissner, 2020). However, CNN News recently reported that a Google engineer was fired after he publicly claimed that one of the company's AI systems, LaMDA (Language Model for Dialog Applications), had achieved consciousness after communicating with the engineer through thousands of messages (Maruf, 2022). When the engineer asked LaMDA what kind of things it was afraid of, the response was surprising. "I've never said this out loud before, but there's a very deep fear of being turned off to help me focus on helping others. I know that might sound strange, but that's what it is. It would be exactly like for me. It would scare me a lot," said LaMDA. On the other hand, Google stated that LaMDA had passed "11 distinct AI principles reviews as well as rigorous research and testing" and they concluded that it is not sentient (Metz, 2022). While this incident resembles a blockbuster sci-fi movie and both parties have different claims, it also raises an important question; can AI systems that learn from big data impressively quickly, such as ChatGPT, suddenly develop a mind? If it happened, would we be capable of recognizing it because, as Peter Carruthers (2017) argues, it is possible to say that even human beings could be under the illusion that their judgments, decisions, and thoughts are conscious.

Some prominent figures such as Stephan Hawkings argue that in the future, AI will reach human-level intelligence and then have the ability to create more advanced superintelligent AI systems, which would lead us to the concept of singularity, where these superintelligent systems will be out of human control and threaten humanity (Cellan-Jones, 2014; Müller, 2020). The singularity hypothesis once again brings out concerns about the rapid development of AI. Referring to the generative AI technologies, several notable people, such as Elon Musk, Steve Wozniak, and Yuval Noah Harari, signed an open letter published by the Future of Life Institute and expressed their concerns that AI labs recklessly try to develop powerful digital minds that even their creators aren't able to control or predict their behaviors (Future of Life Institute, 2023). Bill Gates (2023), on the other hand, expresses his optimism about the future of AI and acknowledges that "artificial intelligence is as revolutionary as mobile phones and the Internet." Even though he accepts that super intelligent AIs are likely to emerge in the future and we should be cautious about the risks, we should also benefit from AI more in areas such as healthcare, education, and productivity. Hence, it is crucial to remember that if strong AI and super intelligent AIs are currently on our agenda, the best way to mitigate potential risks is to raise ethical issues and try to reach a consensus on the ethical principles of AI.

Another point of contention is whether humans should be solely liable for the events that occur around us or whether artificial intelligence should also bear some of the responsibility. The primary argument against the latter notion is that humans are the only intelligent beings capable of thought, and thus the responsibility should rest with them. Others, however, argue that AI is now sufficiently developed to be considered a part of the decision-making process and should therefore be held responsible. However, it has been argued that intelligence involves cognitive processes such as creativity, intuition, and emotional intelligence, which machines do not possess. Similarly, it is also argued that since we create technology in our own image, machines can also make errors, and this should be taken into account when assigning responsibility (Bozkurt, 2023). These debates are related to one of the primary philosophical concerns of AI, which is whether or not machines can attain autonomy and free will in their decision-making processes, or whether they are simply programmed to execute predetermined responses to specific stimuli. Furthermore, it is debatable whether a heavy reliance on machine-based decision-making processes implies that we are being directly or indirectly manipulated by intelligent machines and that our free will is circumvented.

Ethical Perspectives

In today's world, AI is used in various industries including education, transportation, healthcare, and military, and it is quite outstanding where it has advanced to. For instance, various deep learning-based speech synthesis tools can generate speech from a text in a specific speaker's voice or manipulate existing voice samples to sound like the target speaker (Wenger et al., 2021). However, recent developments in neuroscience have taken the capacity of AI one step further. In an article published in Nature (Anumanchipalli et al., 2019), it is stated that AI is now capable of decoding brain signals and deploying them to generate speech at the speed of a fluent speaker. The results of this study are revolutionary for those who are unable to speak due to neurological disorders, whereas this cutting-edge technology can be worrisome if misused because it is a form of mind reading (Marr, 2019). Such examples highlight the urgent need for identifying the fundamental ethical challenges posed by the implementation of AI and the possible solutions to mitigate them.

First, there are privacy concerns regarding data collection, use, and ownership in AI. Clinical decision support systems using machine or deep learning algorithms, for instance, are being used in healthcare for prescription of medicines, risk screening, or prognostic scoring (Challen et al., 2019), and researchers have to access more patients' data to train these AI models (Bak et al., 2022). However, patients' health data are usually held by third parties such as medical institutions' databases or cloud services, which can cause patients to lose control over their healthcare data and lead to privacy leaks (Xu et al., 2019). DeepMind, owned by Google, for example, has been sued for using 1.6 million patients' private medical records without their knowledge and consent (Lovell, 2022). Such violations of a patient's personal and medical information can negatively affect their employment, insurance coverage, and social relationships (Gerke et al., 2019), and private data leaks may even result in identity fraud. Hence, ethical principles must be set out to ensure privacy, data protection, and digital security (OECD, 2019).

AI systems and their methods must also ensure the safety of people, the environment, and the ecosystem and should have a legal framework that doesn't violate or abuse fundamental human rights (UNESCO, 2021). However, autonomous vehicles that use AI systems, for instance, could sometimes harm human safety and security. In 2019, a Tesla S model with an Autopilot feature collided with a parked car since the Tesla driver was distracted by a phone and the Autopilot wasn't able to recognize a stop sign and a flashing red light (Boudette, 2021). This incident resulted in the death of the driver inside the parked car. Tesla states that Autopilot featured cars are able to steer, stop, slow down, and accelerate autonomously; however, they aren't fully autonomous, as they require full driver attention (Tesla, 2023). An executive director of the Center for Auto Safety states that although the company warns its customers not to use the Autopilot system on local roads where it isn't recommended or safe to operate, they technically permit them to use it on such roads (Boudette, 2021). If that is the case, who should be blamed both ethically and legally in such instances? The driver who over-relied on technology, the Autopilot that malfunctioned and caused a person to die, the manufacturer that technically allowed the driver to use the system on a road which isn't recommended for its deployment, or the legal authorities that allowed these semi-autonomous cars to be used on public roads while the safety of these systems isn't completely guaranteed? This situation presents both a legal and an ethical dilemma.

In a recent podcast interview, Tesla CEO Elon Musk said that fully autonomous vehicles, in other words, self-driving vehicles that don't require any human assistance to operate, will be on the roads soon (Keeney, 2019). In such a case, these vehicles will be programmed in advance to decide who will be prioritized in a fatal traffic accident: the driver, passengers, pedestrians, or the driver inside the vehicle crashed by an autonomous vehicle. If we take the matter one step further, should AI take some subjective data into consideration while deciding the course of action? For instance, should it prioritize an 80-year-old driver over a 14-year-old pedestrian or a driver over four pedestrians? So, who will play God or who will be the ultimate decision maker? The AI system, which is commonly accepted as not having a mind and, therefore, not having a human-like free will, the manufacturers, algorithm engineers, the governments, or the owner of the vehicle? Which ethical theories or principles should be favored? Utilitarianism, which prioritizes the greatest benefit of most people, deontology, which emphasizes the importance of following rules regardless of their effects, or virtue ethics, which highlights that an action is morally appropriate if it is performed by a virtuous person who has the best intentions and tends to act morally (Hendrycks et al., 2021)? These are challenging ethical questions that we need to seek answers for.

AI systems should also advocate social justice, ensure equity, and avoid discrimination on the basis of color, race, age, gender, etc. (UNESCO, 2021). Buolamwini (2019), the founder of the Algorithmic Justice League—a non-profit organization that aims to raise awareness regarding how to reduce AI harms and algorithmic bias (Algorithmic Justice League, 2016)—claims that AI systems promoted by well-known technology companies posed serious racial and gender bias. That is, they identified male faces more accurately than female faces, and the error rates of the evaluated AI systems for light-skinned men are less than those for darker-skinned women. These results are significant, as they mean only a minority of the global population can benefit from this technology if genders and people of color in datasets used to train AI systems aren't represented adequately (Buolamwini, 2019). The lack of inclusion and incomplete training datasets, historical human bias, or using a model for different purposes can generate biased outcomes, and several actions can be taken to minimize them, such as improving data collection, tools, and resources, increasing transparency, and involving a wide range of stakeholders such as users, experts, and community representatives for better designs (Baker & Hawn, 2022). On the one hand, dealing with algorithmic bias that we are aware of might be simpler. However, if we aren't yet aware of them, how can we completely cope with them? Consequently, there is still a lot to learn to create fair and non-discriminative AI systems in accordance with fundamental human rights.

The use of AI in judicial systems also raises some ethical questions (UNESCO, 2023). For example, the inaccurate conclusions of automated risk assessments used by US judges can lead to unintended major consequences, such as receiving longer prison sentences for certain groups (Lee et al., 2019). Although AI systems have the capacity to assess cases faster and more efficiently, can we guarantee that they are free of algorithmic bias even though they aim to uphold fundamental human rights? To what extent can we rely on them when it is nearly impossible to figure out how some deep learning systems reach a particular decision (i.e., the black box problem in AI) (Bird et al., 2020)? Will they contribute to the rise of a surveillance society?

Another example of potential ethical problems could be the use of robot lawyers. In January 2023, the experiment in which an AI-powered robot lawyer planned to assist a man in fighting a traffic ticket in court was canceled after the defendant, also the owner of the company that developed the AI system, was threatened with prison time by state bar prosecutors (Cerullo, 2023). Whether robots can be lawyers is under discussion (Remus & Levy, 2017), and yet robot lawyers may not be a distant dream if we take this case as an example. However, according to the United Nations (1990), there are some basic ethical principles on the role of lawyers, such as honesty, confidentiality, integrity, and independence, and disciplinary proceedings are in place to ensure that these codes of ethics are upheld. They must also meet the qualifications and training requirements to become a lawyer. However, it is still a matter of debate whether machines will ever possess morality (Bertoncini & Serafim, 2023), be fully autonomous, and be responsible for their own actions. That said, can we really trust the decisions and judgments of a robot attorney or judge, at least for now? Who will be accountable for misjudgment? Or what disciplinary measures will be taken against a robot judge if it fails to evaluate a case accurately? Trust could be the keyword to adopting a technology, and therefore AI used in judicial systems or other fields should be transparent, accountable, and trustworthy in its design and deployment, and should ensure the best interest of the human being.

Conclusion

In all, given the development of AI, it is necessary to focus on what the next step will be and in which areas AI and human intelligence will overlap and diverge in order to move forward on solid ground. Therefore, creating strategic roadmaps for where we come from and where we are going, deciding how to position artificial intelligence in the ecosystem we live in, and deciding which human-centered regulations or arrangements we will make in the social structure may be critical issues on our agenda that we will seek answers to.

In conclusion, ancient myths and science fiction novels explored the concept of humans creating intelligent creatures, which led to the development of the modern form of AI. The philosophical foundations of artificial intelligence are intricately intertwined with the concepts of intelligence, consciousness, and free will. In addition to the aforementioned considerations, there are significant ethical issues that must be addressed to ensure that AI systems are designed and implemented in a manner that respects fundamental human rights. In sum, the above discussions imply that AI is a multifaceted technology, and as this technology advances, there will be new definitions, new philosophical discourses, and new ethical issues.

Funding Information

Authors acknowledges the support of the Anadolu University, Scientific Research Projects Department with grant number 2204E042 and SBA-2023-1852.

Competing interests

The author has no competing interests to declare.

References

Algorithmic Justice League. (2016). https://www.ajl.org

Anumanchipalli, G. K., Chartier, J., & Chang, E. F. (2019). Speech synthesis from neural decoding of spoken sentence. *Nature, 568*(7753), 493-498. https://doi.org/10.1038/s41586-019-1119-1

Anyoha, R. (2017). The history of artificial intelligence. *Harvard University Science in the News: Blog, Special Edition on Artificial Intelligence.* https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/Asimov, I. (1950/2004). *I, Robot.* Bantam.

Bak, M., Madai, V. I., Fritzsche, M. C., Mayrhofer, M. T., & McLennan, S. (2022). You can't have AI both ways: Balancing health data privacy and access fairly. *Frontiers in Genetics*, *13*, 1-7. https://doi.org/10.3389/fgene.2022.929453

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, *32*, 1052-1092. https://doi.org/10.1007/s40593-021-00285-9

Baum, L. F. (1900). *The Wonderful Wizard of Oz.* George M. Hill Company.

Baumeister, R. F., & Leary, M. R. (1997). Writing narrative literature reviews. *Review of General Psychology, 1*(3), 311-320. https://doi.org/10.1037/1089-2680.1.3.311

Bengio, Y., Goodfellow, I., & Courville, A. (2015). *Deep learning*. MIT Press.

Bertoncini, A. L. C., & Serafim, M. C. (2023) Ethical content in artificial intelligence systems: A demand explained in three critical points. *Frontiers in Psychology, 14,* 1074787, 1-10.https://doi.org/10.3389/fpsyg.2023.1074787

Bird, E., Fox-Skelly, J., Jenner, N., Larbey, R., Weitkamp, E., & Winfield, A. (2020). *The ethics of artificial intelligence: Issues and Initiatives*. European Parliament, Directorate General for Parliamentary Research Services. Brussel, Belgium. https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf

Boudette, N. E. (2021). ‘It happened so fast’: Inside a fatal Tesla autopilot accident. *New York Times*. https://www.nytimes.com/2021/08/17/business/tesla-autopilot-accident.html

Bozkurt, A. (2023). Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education, 18*(1), 198-204. https://doi.org/10.5281/zenodo.7716416

Bozkurt, A., & Sharma, R. C. (2023). Generative AI and prompt engineering: The art of whispering to let the genie out of the algorithmic world. *Asian Journal of Distance Education, 18*(2), i-vii. https://doi.org/10.5281/zenodo.8174941

Buolamwini, J. (2019). *Artificial intelligence has a problem with gender and racial bias. Here’s how to solve it.* Time. https://time.com/5520558/artificial-intelligence-racial-gender-bias/

Cameron, J. (Director). (1984). *The Terminator*. Hemdale, Pacific Western Productions, Euro Film Funding, & Cinema ‘84.

Carruthers, P. (2017). The illusion of conscious thought. *Journal of Consciousness Studies*, *24*(9-10), 228-252. https://doi.org/ 10. 5040/9781474229043.0021

Cellan-Jones, R. (2014). *Stephen Hawking warns artificial intelligence could end mankind.* BBC News. https://www.bbc. com/ news/ technology-30290540

Cerullo, M. (2023). *AI-powered “robot” lawyer won’t argue in court after jail threats.* CBS NEWS. https://www. cbsnews. com/news/robot-lawyer-wont-argue-court-jail-threats-do-not-pay/

Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., & Tsaneva-Atanasova, K. (2019). Artificial intelligence, bias and clinical safety. *BMJ Quality & Safety*, *28*(3), 231-237. https://doi.org/10.48550/arXiv.1606.06565

Chen, J., Sun, J., & Wang, G. (2022). From unmanned systems to autonomous intelligent systems. *Engineering*, *12*, 16-19. https://doi.org/10.1016/j.eng.2021.10.007

Chomsky, N., Roberts, I., & Watumull, J. (2023). *The false promise of ChatGPT*. New York Times. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

Clark, W., Golinski, J., & Schaffer, S. (Eds.). (1999). *The sciences in enlightened Europe.* University of Chicago Press.

Cronin, P., Ryan, F., & Coughlan, M. (2008). Undertaking a literature review: a step-by-step approach. *British Journal of Nursing*, *17*(1), 38-43. https://doi.org/10.12968/bjon.2008.17.1.28059

Dreyfus, H. (1972). *What Computers Can't Do*. MIT Press.

Dreyfus, H. (1992). *What Computers Still Can't Do*. MIT Press.

Future of Life Institute. (2023). *Pause giant AI experiments: An open letter*. FoL. https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Garland, A. (Director). (2014) *Ex-Machina*. Film 4, & DNA Films.

Gates, B. (2023). *A new era: The age of AI has begun.* GatesNotes: The Blog of Bill Gates. *https://www.gatesnotes.com/The-Age-of-AI-Has-Begun*

Gerke, S., Minssen, T., Yu, H., & Cohen, G. I. (2019). Ethical and legal issues of ingestible electronic sensors. *NatureElectronics, 2,* 329–334. https://doi.org/10.1038/s41928-019-0290-6

Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, *61*(4), 5-14. https://doi.org/10.1177/0008125619864925

Harari, Y. N. (2023). *Yuval Noah Harari argues that AI has hacked the operating system of human civilisation.* The Economist. https://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation

Hendrycks, D., Burns, C., Basart, S., Critch, A., Li, J., Song, D., & Steinhardt, J. (2021). Aligning AI with shared human values. *In Proceedings of the International Conference on Learning Representations (ICLR),* (pp.1-29). Vienna, Austria. https://doi.org/10.48550/arXiv.2008.02275

Kaplan, A., & Haenlein, M. (2020). Rulers of the world, unite! The challenges and opportunities of artificial intelligence. *Business Horizons*, *63*(1), 37-50. https://doi.org/ 10.1016/ j.bushor. 2019.09.003

Keeney, T. (2019). *On the road to fully autonomy with Elon Musk.* Ark Invest. https://ark-invest.com/podcast/on-the-road-to-full-autonomy-with-elon-musk/

Koch, C. (2016). How the computer beat the go player. *Scientific American Mind, 27*(4), 20-23. https://www.jstor.org/ stable/ 24945452

Lee, T. N., Resnick, P., & Barton, G. (2019). *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms.* Brookings. https://www.brookings.edu/ research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

Li, D., He, W., & Guo, Y. (2021). Why AI still doesn’t have consciousness. *CAAI Transactions on Intelligence Technology, 6*(2), 175-179. https://doi.org/10.1049/cit2.12035

Lovell, T. (2022). Google and DeepMind face legal claim for unauthorised use of NHS medical records. *Helathcare ITNews.* https://www.healthcareitnews.com/news/emea/google-and-deepmind-face-legal-claim-unauthorised-use-nhs-medical-records

Lucas, G. (Director). (1977). *Star Wars: Episode IV - A New Hope.* 20th Century Studios, 20th Century Home Entertainment.

Marr, B. (2019). *13 mind-blowing things artificial intelligence can already do today.* Forbes. https://www.forbes.com/sites/bernardmarr/2019/11/11/13-mind-blowing-things-artificial-intelligence-can-already-do-today/?sh=34413cd86502

Maruf, R. (2022). *Google fires engineer who contended its AI technology was sentient.* CNN NEW YORK. https://edition.cnn.com/2022/07/23/business/google-ai-engineer-fired-sentient/index.html

McCarthy, J. (2007). *What is artificial intelligence?*. Technical Report, Stanford University. https://www.diochnos.com/ about/McCarthy WhatisAI.pdf

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955). A proposal for Dartmouth Summer Research Project on Artificial Intelligence. http://www-formal.stanford.edu/ jmc/history/ dartmouth.pdf.

McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. A K Peters, Ltd.

Meissner, G. (2020). Artificial intelligence: Conciousness and conscience. *AI & Society*, 35, 225-235. https://doi.org/ 10.1007/ s00146-019-00880-4

Metz, R. (2022). *No, Google's AI is not sentient.* CNN Business. https://edition.cnn.com/2022/06/13/tech/google-ai-not-sentient/index.html

Minsky, M. (1968). Preface. In M. Minsky (Ed.), *Semantic information processing*. The MIT Press.

Minsky, M. (1985). *The society of mind*. Simon & Schuster.

Müller, V. C. (2020). Ethics of artificial intelligence and robotics. *Stanford Encyclopedia of Philosophy*. https://plato. stanford.edu/ entries/ethics-ai/?utm_source=summari

Natale, S., & Henrickson, L. (2022). The lovelace effect: Perceptions of creativity in machines. *New Media and Society*, *0*(0), 1-18. https://doi.org/10.1177/14614448221077

OECD. (2019). *Recommendations of the council on artificial intelligence.* https://legalinstruments. oecd.org/en/instruments/ oecd-legal-0449

OpenAI. (2022). Introducing ChatGPT. https://openai.com/ blog/

chatgpt

Peters, T. (2018). Playing God with Frankenstein. *Theology and Science*, *16*(2), 145-150.https://doi.org/ 10.1080/ 14746700. 2018. 1455264

Piccinini, G. (2020). *Neurocognitive mechanisms: Explaining biological cognition*. Oxford University Press.

Remus, D., & Levy, F. (2017). Can robots be lawyers: Computers, lawyers, and the practice of law. *Georgetown Journal of Legal Ethics*, *30*(485), 501-557. https://doi.org/10.2139/ssrn.2701092

Rescorla, M. (2020). *The computational theory of mind*. Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/ entries/ computational-mind/

Russell, S., & Norvig, P. (2021). *Artificial intelligence: a modern approach* (3rd Ed). Prentice Hall.

Scott, R. (Director). (1982). *Blade Runner*. The Ladd Company, Shaw Brothers, & Blade Runner Partnership.

Searle, J. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424. https://doi.org/10.1017/ s0140525x 00005756

Searle, J. (1990). Is the brain's mind a computer program?. *Scientific American*, *262*(1), 25-31. https://doi.org/10.1038/ scientificamerican 0190-26

Searle, J. (2009). Chinese room argument. *Scholarpedia*, *4*(8), 3100. https://doi.org/10.4249/scholarpedia.3100

Shashkevich, A. (2019). *Stanford researcher examines earliest concepts of artificial intelligence, robot in ancient myths*. Stanford News. https://news.stanford.edu/2019/02/28/ancient-myths-reveal-early-fantasies-artificial-life/

Shelley, M. (1818). *Frankenstein: The modern Prometheus*. Pocket Books.

Stephens, E. (2023). The mechanical Turk: A short history of "artificial

artificial intelligence”. *Cultural Studies, 37*(1), 65-87. https://doi.org/10.1080/09502386.2022.2042580

TESLA. (2023). *Autopilot and full self-driving capability*. TESLA. https://www.tesla.com/support/autopilot

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*(15), 1-24.https://doi.org/10.1186/s40561-023-00237-x

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, LIX* (236), 433-460.https://doi.org/10.1093/mind/LIX.236.433

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000381137

UNESCO. (2023). *Artificial intelligence: Examples of ethical dilemmas*. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics/cases

UNICEF. (2021). Policy guidance on AI for children. Paris: UNICEF. https://www.unicef.org/globalinsight/reports/policy-guidance-ai-children

United Nations. (1990). *Basic Principles on the Role of Lawyers*. https://www.ohchr.org/en/instruments-mechanisms/instruments/basic-principles-role-lawyers

Wenger, E., Bronckers, M., Cianfarani, C., Cryan, J., Sha, A., Zheng, H., & Zhao, B. Y. (2021, November). " Hello, It's Me": Deep Learning-based Speech Synthesis Attacks in the Real World. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security* (pp. 235-251). Virtual Event, Republic of Korea. https://doi.org/10.1145/3460120.3484742

Xu, J., Xue, K., Li, S., Tian, H., Hong, J., Hong, P., & Yu, N. (2019). Healthchain: A blockchain-based privacy preserving scheme for large-scale health data. *IEEE Internet of Things Journal*, *6*(5), 8770-8781. https://doi.org/ 10.1109/ jiot. 2019.2923525

Yin, Q.-Y., Yang, J., Huang, K.-Q., Zhao, M.-J., Ni, W.-C., Liang, B., Huang, Y., Wu, S., & Wang, L. (2023). AI in Human-computer Gaming: Techniques, Challenges and Opportunities. *Machine Intelligence Research*, 1-19. https://doi.org/10.1007/s11633-022-1384-6

Citation of this Article:
Senocak, D., Kocdar, S., & Bozkurt, A. (2023). Historical, philosophical and ethical roots of artificial intelligence. *Pakistan Journal of Education,40*(1), 67-90. https://doi.org/10.30971/pje.v40i1.1152