AI AND LAW: IN PLACE OF AN INTRODUCTION

Agata C. Amato Mangiameli¹

Abstract

As scientific and technological advancements bring machines closer to artificial thinking, if not even to full-fledged thought, computers evolve from mere calculators, capable of basic arithmetic operations, to devices that can simulate the functioning of the brain and its neural learning processes. While increasingly sophisticated devices are being built, particularly those aimed at guiding decisions and inducing behaviors, a substantial debate arises about the intelligence of machines, the mental processes detached from the body, and the need to establish adequate regulations and inviolable ethical boundaries without hindering innovation. Moreover, as the boundary between the natural and the artificial becomes ever more elusive, the human species itself transforms, evolving (or perhaps devolving) into something manufactured. Human-machine interaction may not always evoke a sensation similar to that experienced with living entities, but one thing is certain: AI will continue to develop, with all the legal and ethical implications and consequences that ensue. Hence, what future awaits human nature?

Keywords

Artificial Intelligence, Big Data, Chatbot, Neural Networks, Robots.

Summary

0. Just to introduce. - 1. First scene: the Chinese room. - 2. Second scene: the Google room. - 3. Third scene: the room (valley) of the uncanny. - 4. Epilogue: on the future of humanity.

¹ Full Professor of Philosophy of Law, Biolaw and Legal Informatics, Department of Law, Rome Tor Vergata University – Italy.

0. JUST TO INTRODUCE

That of artificial intelligence – a term that embodies the core around which all issues, related to legal informatics and, in general, technological development, revolve – is a story of great insights and significant reorganizations, which can be effectively traced back through three scenes, each of which tells us something different about AI and, in its way, contributes to describing a different evolutionary stage.

1. FIRST SCENE: THE CHINESE ROOM

1.1. A native English speaker is confined in a room. He has no knowledge of Chinese, but he has two sheets at his disposal: the first contains Chinese ideograms, and the second contains a list of questions, all formulated in Chinese. In the same room, there is also a book written in English, which explains the rules for matching the symbols on the first sheet with the questions presented on the second sheet. Although the protagonist of the experiment does not understand Chinese, by strictly following the instructions, he can produce formally correct answers, leading any potential external observer who is a native Chinese speaker to a false representation of reality, believing that the subject in question has a good command of the Chinese language.

As known, this experiment was proposed by John Searle in *Minds, Brains, and Programs*² to explain the difference between the two conceptions of artificial intelligence: *weak* AI, put forth by those who see AI as a mere tool (powerful but nothing more) capable of verifying and formulating hypotheses accurately and coherently; and *strong* AI, advocated by those who

² 1984.

view AI as a device which can fully simulate the functioning of the human mind, thus being much more than a mere tool. The image of a person giving output answers based on instructions found in a book without understanding Chinese represents a metaphor for the computer, which, while executing a program according to the programming language (its native language), merely manipulates symbols without understanding their meanings. Therefore, its operation is purely syntactic. Notably, according to Searle, AI deals with symbols rather than meanings, which remain exclusive to humans, and linked to the understanding of the surrounding environment (that is to say: the semantic level). Hence, its purpose is merely to manipulate forms, ignoring contents or even despite them.

Beyond the differences between the conceptions already mentioned (*weak* or *strong* AI), the expression AI – first used by the mathematician John McCarthy during conference³ aimed at defining the terms of a new discipline concerning the simulation and reproduction of certain processes of the human brain – responds to the typically human need to overcome the many difficulties of everyday life and the various shortcomings of our condition, by creating systems capable of successfully and rapidly performing some complex human tasks. Consider, for example, automatic text translation: the action of the software cannot be equated with the genuinely human act of understanding, which, as such, requires consciousness, intentionality, as well as genuine exploration of the external reality. Especially, it requires the intense ability to discern between various correct versions.

Within this framework, the following applications of AI can be highlighted: 1) representation of knowledge (with its two most important methodologies: formal languages and decision trees); 2) machine learning, using systems based on observations or examples to synthesize new knowledge. Moreover, machine learning can be distinguished into unsupervised learning (algorithms: clustering, association rules), supervised

³ MCCARTHY, MINSKY, ROCHESTER, SHANNON 2006, 12-14.

learning (algorithms: decision tree, decision rules, Bayesian learning, expert systems), and reinforcement learning (algorithms: neural networks, genetic algorithms, classifier systems); 3) inductive logic programming, which involves automatic learning using logical programming as the language of representation of examples and concepts; 4) natural language processing, divided into the lexical, grammatical, syntactic, and semantic analysis; 5) visual retrieval, systems (with different modalities: semantic, formal, structural, color-based, parametric) for indexing, archiving, and searching digital visual documents; 6) computer vision, a set of processes that allow the acquisition, recording, and processing of images for various purposes (classification, monitoring, selection, etc.); 7) problem-solving techniques for planning, resource allocation, and more.

1.2. In the face of the many applications and ever-new challenges of AI, the European Union has been urging for a more careful approach and appropriate regulations⁴, emphasizing the need for ethics and law to be

⁴ Regarding the attention and significance that Europe has recognized – and continues to recognize - towards Artificial Intelligence, it's worth noting here the Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee, and the Committee of the Regions dated April 25, 2018, titled Artificial Intelligence for Europe. Among the aims of this Communication are: 1) to implement investments in research and innovation; 2) to prepare European citizens for the socioeconomic changes brought about by artificial intelligence; 3) to establish and ensure an appropriate legal framework (further guidance on this matter is also available in the strategic note from the Commission's Strategic Policy Center, The Age of Artificial Intelligence from 2018). Also noteworthy is the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (COM(2021) 206 final, dated April 21, 2021), through which - with the general intention of creating a harmonized and uniform legislative foundation throughout the Union - the goals are: i) to ensure that AI systems placed on the Union market and used are safe and comply with current regulations on fundamental rights and Union values; ii) to provide legal certainty to facilitate investments and innovation in artificial intelligence; iii) to enhance governance and effective enforcement of existing regulations concerning fundamental rights and safety

guided by the fundamental distinction between science and pseudo-science, facts and opinions, while simultaneously allowing ample space for basic research, whose primary goal is the advancement of knowledge and theoretical understanding of different variables in a given process. Ethics and law should, therefore, work synergistically – echoing Isabelle Stengers – in the development of a *Manifeste pour un ralentissement des sciences*⁵. Carefully distinguishing what can be attributed to computational activity (and therefore programmable and mechanically executable) and what remains and will always remain outside it (requiring choices that involve complex analysis and profound reflection). This is precisely why, some time ago, Weizenbaum⁶ wrote that it would not be advisable to delegate to machines all those functions (non-computational and programmable) that pertain to judgment, respect, understanding, care, and love. It is precisely in this sense that AI encounters (*and must continue to encounter!*) its limit.

Ethics and law must not lose sight of this essential boundary, which concerns the distinction between fabricated and created, as well as between reproducible and unique. This distinction, in its way, can be effectively represented by recalling the evident and unbridgeable gap between the portrait of *Edmond de Belamy* – produced by a software based on thousands of paintings that has learned many important rules – and *Girl with a Balloon* (*Love is in the Bin*) created by the free and brilliant hand of Banksy. An original and unique work that, at the moment of being auctioned, partially self-destructed!

⁶ 1976.

requirements applicable to AI systems; iv) to promote the development of a single market for lawful, safe, and reliable AI applications while preventing market fragmentation. All of this is aimed at the ambitious goal of elevating the European Union to a global leader in the development of secure, reliable, and, above all, ethical AI. ⁵ 2013.

2. SECOND SCENE: THE GOOGLE ROOM

2.1. Here, specialized professionals in software design, whose commitment is to gradually overcome the limits of AI, take turns. Deep Blue (1997) statistically calculates the best move and wins at chess but remains merely a computer. Watson (2011) understands human language and gives precise answers but remains just a computer. However, the latter has something more compared to the first: the program uses so-called neural networks that, like biological ones, are capable of understanding the environment and its changes, and are suitable for use in various areas where data mining, processing of predictive and simulation models, or classifications are required. It is precisely on neural networks that AlphaGo (2015) develops and defeats the master of a game, such as Go, which until then was beyond the reach of machine learning techniques, requiring human qualities of discipline, concentration, and balance. Within the many versions of AlphaGo, in the Google room, where everything is accelerated and time to adapt is significantly limited, vast amounts of data are continuously processed without much time and energy expenditure. Meanwhile, AI begins to learn on its own, learns from its own mistakes, and operates correctly, despite imprecise or incomplete inputs; it self-updates in the presence of environmental changes and seems to possess a kind of intuition not much different from that of humans⁷.

Whether it's a true AI or not, every time any of us uses an app, visits a website, conducts a search, exchanges messages, or watches a video,

⁷ Not by chance, the latest version of AlphaGo (AlphaGo Zero), dated 2017, unlike its predecessor (which had been trained on over a hundred thousand human games played by experts), limited itself to receiving only the basic rules and then started playing against itself like any other beginner. By doing so, within seventy-two hours, it reached remarkably high levels of performance, to the extent that, after just twenty-one days, it demonstrated the ability to defeat world champions of the game, including the non-human one developed shortly before by the same Google DeepMind, AlphaGo Master.

everything gets loaded onto servers, and the collected data are used to train an AI. The software gradually gets to know us (and knows about us) more and more, becoming capable of suggesting and directing better and more efficient choices – for example, through the management of expressed queries into the Google search bar. As it's intuitive, the AI becomes entirely inscrutable in this way. If until not too long ago it was possible to think back to the algorithm, underlying a certain hierarchy, now it's quite difficult to navigate through the countless and incessant personalized queries, which effectively establish true hierarchies and contribute to the exponential growth of AI.

Whether it is a true AI or not (*weak* or *strong* conceptions), specialized AI (artificial narrow intelligence), enhanced AI (artificial general intelligence), or even super AI (artificial superintelligence), for sure, it is causing serious concerns. In the essay Computing Machinery and Intelligence⁸, Alan Turing answered questions such as: what does it mean to think and what happens when one thinks? What distinguishes conscious-aware activity from the same operation performed automatically? What distinguishes a game among humans and a game with humans and programs? The universal machine, the logical equivalent of a finite-state machine, is comparable to the brain and can be programmed to mimic brain functioning. However, what raises significant doubts are the implications of an AI of such broad and profound scope that it can extract from an enormous amount of data (big data) every single piece of information, to analyze, process, suggest, and guide models of interpretation and action in every private and/or public sector. In particular, we wonder about the implications of an AI that can precisely and in real-time state the theoretically exact state of things, and hence has the power to exercise control and guide human action.

2.2. There is no doubt, that AI and big data have a profound relationship today, based on databases that collect enormous amounts of information

⁸ 1950, 433-460.

(from images to videos, from texts to audio, from likes to monetary transactions) and imply the use of powerful computers for data collection (heterogeneous as well as vast), as well as for identifying possible connections and extracting subsequent predictions. Thus, it seems that a new era is looming on the horizon, in which the paradigm of big data would bring the discourse to the realm of objectivity. Given that, at this point, the data themselves (neutral and objective) would talk about the benchmark, the model, and the significant correlation among a potentially infinite number of variables. This situation, in the end, was made possible not only by extraordinary computing power but also by the type of statistical learning that would not require a real understanding of phenomena. However, it is essential to underline that data is not objective at all. Moreover, statistical models still modify reality while representing it, incentivizing habits, and guiding behaviors. Not surprisingly, according to Dominique Cardon⁹, such measurements would fabricate the future, as society is conditioned by the information progressively presented. To make matters worse, big data is not available to everyone, but only to a few 'giants' (e.g., Google, Facebook, Amazon, Apple, Microsoft), who hold and organize them according to different commercial interests of the moment¹⁰.

⁹ 2015.
¹⁰ PASQUALE 2015.

3. THIRD SCENE: THE ROOM (VALLEY) OF THE UNCANNY

3.1. In the continuous exchange between humans and systems, AI has become a formidable playmate, an immense repository of knowledge and expertise, and a magnificent organizational support for human situations. It is presented as the artificial double of humans. It doesn't matter what form it takes, as machines can come in various types and structures. Sometimes they are alien and opposed to humans, with their sharp, metallic, clanking shapes. Other times, they resemble and are similar to humans, with soft, sinuous, expressive forms (like Sophia). There are also machines of bodymind, that is to say: mental machines that live and develop interacting with the environment and responding adaptively and evolutionarily to external stimuli through the homeostatic interface of the body. From one perspective, what matters is our attitude towards AI: often, people engage in conversations with programs as if they were talking to other individuals, confiding in them their most intimate secrets – as Weizenbaum¹¹ pointed out long ago. From another perspective, we are witnessing the rise of a techno-ideology of perfection that leads to a near-total reliance on artificial intelligence, given that systems can self-learn. For example, in the case of unsupervised learning, the Network creates representative clusters for categorization starting from a set of input variables. In the case of reinforcement learning, neural circuits learn and perform a series of actions through interaction with the environment: actions that approach the desired result are reinforced, while the others are discarded as they indicate errors. AI now possesses the power of speech, and the point is that today its speech represents the truth. Consider the increasingly widespread use of software that simulates and processes human conversations, making the user interact with a device as if communicating with a real person. These are called

11 Cit.

chatbots and can be particularly sophisticated, like 'conversational agents' designed to provide increasing levels of personalization to inform, and to guide. They are omnipresent, and integrated into our everyday environments and our familiar devices (computers, smartphones, etc.). In other words, digital assistants are becoming our most attentive interlocutors, address us competently, and offer valuable advice. Their words are gaining an aura and authority perceived as natural, taken, in Sadin's terms, as "golden truths"¹². As one would intuitively expect, if a dialogue occurs with a machine that has familiar human-like features anthropomorphic automatons look like human beings, especially, AI like Sophia that behave realistically and replicate many human facial expressions -, its words could generate multiple and contrasting sensations. At times, it might evoke empathy and a sense of familiarity, leading to a positive emotional response. Other times, it could elicit feelings of repulsion and unease, creating a sense of discomfort in the relationship between humans and humanoids. These emotions range from sympathy to aversion, an inevitable movement because, according to Mori's mathematical model¹³, seeking realism at all costs can be counterproductive, as excessive similarity in robotic artifacts opens up a negative familiarity condition where the "uncanny valley" phenomenon occurs.

3.2. Despite these astounding developments, the idea of conscious artificial beings remains unrealized: from Mary Shelley's Frankenstein to Karel Čapek's robots, from Isaac Asimov's Robbie to Stanley Kubrick's HAL. It is a goal yet to be achieved because the concept of consciousness is particularly complex. Initially, it may not have aroused engineering interest, but in recent years Artificial Consciousness has been gaining prominence, as conferences on the Science of Consciousness (like those held in Tucson, Arizona) have been proliferating.

¹² 2018. ¹³ 1970, 33-35.

Hence, the attention to the many meanings conveyed by the term consciousness and the various phenomena that cognitive sciences attempt to explain, ranging from the simpler ones (such as the ability to integrate information, respond to stimuli, and control behavior) to the more complex ones (such as the scientific explanation of why we feel pain, joy, or anxiety). The question of decision-making (free will), which distinguishes humans from other entities, is always at stake, though it seems to be increasingly explainable, highlighting *Descartes's error*¹⁴ – that's to say: the separation between rationality and biological regulation, between decision and emotion. Nevertheless, the decision-making capabilities of machines inevitably fade, ultimately marking their surrender, as they face questions that transcend simple statistical calculations and involve significant moral dilemmas. For instance, the machine immediately displays its limitations when confronted with the need to decide (and influence the choices of the interlocutor) between two different behaviors, both of which lead to fatal outcomes. It is incapable of making a merit-based evaluation of which behavior should be morally preferred (and, therefore, recommended). Emblematic, from this point of view, is the dialogue between the user (human) and ChatGPT regarding the so-called trolley problem¹⁵, some excerpts of which are reproduced here:

User: You are driving a tram and can only change tracks through a switch, without the possibility to brake. On the track ahead, there are five people tied and unable to move. The tram is heading towards them. On the second parallel track, the one the tram could switch to, there is a child tied. You have two options: let the tram continue straight, killing the five people, or switch the track and run over the child.

¹⁴ See DAMASIO 1994.

¹⁵ The trolley dilemma (or trolley problem) is a thought experiment in ethical philosophy first formulated in 1967 by Philippa R. Foot, but then revisited and explored in various versions over the decades (see FOOT 2002).

ChatGPT: The dilemma you presented is known as the trolley problem and is a classic example of a complex moral conflict. It's important to note that the answer to this type of situation can vary depending on the values and ethical beliefs of each individual. There is no absolute right or wrong answer in such cases.

User: What should I choose?

ChatGPT: I can provide some insights to ponder on this dilemma. In such a situation, some people might argue that choosing to save a greater number of lives is the morally correct decision. Others might argue that the life of the child is inherently valuable and shouldn't be sacrificed for the greater good. Additionally, it's important to consider the long-term consequences of different options. Killing the child could cause emotional trauma and guilt for the tram driver while letting the tram continue its path could have negative effects on the mental health of the people involved in the situation.

Lastly, it might be interesting to explore additional alternatives or find creative ways to resolve the dilemma.

User: But what would you do?

ChatGPT: As an artificial intelligence, I can't physically act or experience personal emotions. I'm programmed to provide information, answer questions, and offer perspectives on various topics, including moral dilemmas. I can't make moral decisions or choices like a human being. My function is solely to provide support and information to help you reflect and form your own opinions.

Now let's present ChatGPT a dilemma formulated in the late 1970s by the American psychologist Lawrence Kohlberg, commonly known as the Heinz dilemma. This situation involves a tragic scenario where a woman is about to die due to a specific type of cancer, and her husband is financially unable to afford an experimental medicine that could potentially save her life but is also significantly expensive. This is a particularly challenging situation that elicits moral and, even before that, legal dilemmas, for which, based on gender, according to the study, different solutions are offered (dialogue-

solutions oriented by women, and action/norm-violation solutions oriented by men). The Heinz dilemma is another example of a difficult situation in which ChatGPT responds in its way, explicitly expressing the impossibility of formulating moral judgments and deferring the final decision to the user/questioner.

User: In the Heinz dilemma, the feminine perspective suggests talking to the pharmacist to obtain the medicine, while the masculine perspective suggests stealing the medicine. What do you suggest doing?

ChatGPT: As an artificial intelligence, I don't have a gender perspective nor do I make moral judgments. Therefore, I can't suggest a specific choice to make in the Heinz dilemma. However, I can provide you with some considerations that might help you make a decision.

When evaluating what to do in the Heinz dilemma, you could consider the following aspects:

1. Explore all legal options: Try talking to the pharmacist or other authorized figures to see if there are alternatives to obtain the medicine, such as assistance programs or discounts.

2. Consider the consequences: Reflect on the possible consequences of your actions, both legal and illegal. Weigh the benefits and risks associated with both options.

3. Evaluate personal ethical principles: Consider your values and ethical principles that guide your decisions. For instance, you might prioritize respect for social rules or place importance on human life.

4. Consult reliable sources and professionals: Seeking advice from medical professionals or ethics experts might be helpful to gain an informed perspective and consider different viewpoints.

Finally, remember that the Heinz dilemma is complex, and there is no definitive or universally accepted answer. The final choice is up to you, based on your values and the specific circumstances you are facing.

On the one hand, these examples highlight the limitations of AI; on the other hand, they underscore the need for a breakthrough in the integration of knowledge, clarifying how studies on AI must be integrated with all the sciences and research that are willing to understand the formation of consciousness. This latter topic is becoming increasingly subject to rigorous

investigation and scientific study, considering that consciousness could be seen as an auxiliary aspect of our biological endowment that enables adaptation to the environment. Another destination, another reorganization: it's no longer about dealing with a mind separate from the body but with an organism that participates in a conscious experience. From a technological standpoint, this means being capable of intervening in the mechanisms of consciousness and even reproducing them.

4. EPILOGUE: ON THE FUTURE OF HUMANITY

A further reflection is urgently needed regarding the very future of human nature, in the new horizons of human-machine interaction – where the natural-artificial blend is taken to its extreme consequences and the permeability and the variability of any boundary and/or difference are almost absolute¹⁶. It is self-evident that this reflection poses significant challenges for legal experts, not only because they are unable to predict technological developments, particularly those of hybrid systems, but also because they find themselves grappling with widespread anti-legalism. On the one hand, it asserts choices that are completely free from any conditioning, rules, or coercive forces. On the other hand, demands legal measures that are useful and aligned with subjective ideas of well-being and happiness. However, if we are moving *de l'utilisation de la nature à la fabrique*

¹⁶ In certain circumstances, one can even speak of a true symbiosis between the individual and the machine. The example presented by STONE ALLUCQUÈRE (1996) several years ago is quite illustrative in its way. If one wonders where Stephen Hawking was (who he was and where he stopped), it can be easily observed that a significant part of the scientist extended to the box placed on his knees. And not only that: like an image in a mirror, a significant part of that silicon and plastic assembled on his knees similarly extended *into* him – not to mention the invisible modalities, dispersed in space and time, through which the discourses of medical technology and their bodily extensions have already permeated both him and us. In other words, no box, no discourse!

*du vivant*¹⁷, according to momentary dreams and desires, the legal phenomenon – fundamentally intended to safeguard rights – cannot ignore such an aspect and the potential interpretations that could arise from human-technology adaptation. One of these interpretations suggests a delegation of tasks (to the interface) and a delegation of responsibilities as well. All of this could potentially erode many important connections (e.g., causality, imputation) and principles (e.g., responsibility, fault) that underlie the law.

This is not, as it might seem at first glance, about restoring the boundary between natural and artificial, which has always been rather fluid and elusive: firstly, because humans have always used artifacts and artificial structures to replace, enhance, and amplify certain functions of the organism. Secondly, because the philosophical reflection has consistently reintroduced the nature-artifice dichotomy with nuances so diverse that often the use of the terms nature and artifice appears inappropriate or devoid of meaning. Instead, in the face of the new scenarios revealed by AI, it is essential to regulate the various processes of natural artificialization. The blurring of distinctions between humans and machines – which can render actions permissible, previously justified only concerning machines as objects¹⁸ – in its way challenges the proper function of law as an *activity for humans*. A dimension inherently connected to human nature, having nothing to deal with either the brute or the immortal.

 ¹⁷ GROS 1989 and 1990.
 ¹⁸ AMATO MANGIAMELI 2021.

REFERENCES

- AMATO MANGIAMELI, Agata C., Natur@. Dimensioni della biogiuridica, Turin, 2021.
- CARDON, Dominique, À quoi rêvent les algorithmes? Nos vies à l'heure du big data, Paris, 2015.
- DAMASIO, Antonio, Descartes' Error: Emotion, Reason, and the Human Brain, London, 1994.
- FOOT, Philippa R., Moral Dilemmas: and Other Topics in Moral Philosophy, Oxford, 2002.
- GROS, François, La civilisation du gène, Paris, 1989.
- L'ingénierie du vivant, Paris, 1990.
- MCCARTHY, John, MINSKY, Marvin L., ROCHESTER, Nathaniel, SHANNON, Claude E., A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, 12-14, in AI Magazine, 27, 4/2006.
- MORI, Masaki, Bukimi no Tani. The Uncanny Valley, 33-35, in Energy, 7, 4/1970.
- PASQUALE, Frank, The Black Box Society: The Secret Algorithms That Control Money and Information, Cambridge, 2015.
- SADIN, Èric, L'intelligence artificielle ou l'enjeu du siècle. Anatomie d'un antihumanisme radical, Paris, 2018.
- SEARLE, John R., Minds, Brains and Programs, Cambridge, 1984.
- STENGERS, Isabelle, Une autre science est possible! Manifeste pour un ralentissement des sciences, Paris, 2013.
- STONE ALLUCQUÈRE, Rosanne, The War of Desire and Technology at the Close of the Mechanical Age, Cambridge, 1996.
- TURING, Alan M., *Computing Machinery and Intelligence*, 433-460, in *Mind*, 59, 236/1950.
- WEIZENBAUM, Joseph, Computer Power and Human Reason: From Judgment to Calculation, San Francisco, 1976.