Stretching the Notion of Moral Responsibility in Nano electronics by Applying AI

Article · March 2021

DOI: 10.1515/9783110719932-004

2 authors:

Robert Albin
Sapir Academic College
13 PUBLICATIONS **42** CITATIONS
SEE PROFILE

Amos Bardea
Holon Institute of Technology
32 PUBLICATIONS **1,564** CITATIONS
SEE PROFILE

Robert Albin and Amos Bardea

4 Streaching the notion of moral responsibility in nanoelectronics by appying AI

Abstract: The development of machine learning and deep learning (DL) in the field of AI (artificial intelligence) is the direct result of the advancement of nano-electronics. Machine learning is a function that provides the system with the capacity to learn from data without being programmed explicitly. It is basically a mathematical and probabilistic model. DL is part of machine learning methods based on artificial neural networks, simply called neural networks (NNs), as they are inspired by the biological NNs that constitute organic brains. Despite its similarity to biological organs such as human brains, major problems arise in trying to attribute moral responsibility to autonomic systems based on hardware including nano-electronic devices, which are sought to replace humans (moral agents) in the context of AI. It is suggested that the required emotional environment which enables actions according to reasons in humans is not witnessed at AI devices. Though AI technology raises an enticing resemblance to human actions, this resemblance is to be considered with a skeptical eye. It is because actions are associated with reasons while causes are connected to operations. Human agents are capable of acting upon their reasons, while AI devices are limited only to operations as they are conditioned by their programming, which is considered as an embodiment of some causes. As moral responsibility goes hand in hand with the capacity to act (according to reasons), attributing moral responsibility to AI devices is revealed to be but a misleading metaphor.

Keywords: AI, smart system, moral responsibility, reactive attitudes, nano-electronics

4.1 Introduction

The semiconductor industry has changed the world by miniaturizing the printed components. The microelectronics came to be nano-electronics, according to Moore's law. Higher throughputs and mass production of nano-devices significantly improved the quality of life of mankind. In nano-electronics, billions of transistors are assembled on very small volumes by very large-scale integration and hence enable to build a storage capacity and processing of data in a very short time. Smart things were raised, such as an autonomic car, smartphone, smart home, smart city, smart grid, and Internet of things. The combination of smart nano-sensors and IC (integrated circuits) was

Robert Albin, Department of Liberal Arts, Sapir Academic College, Israel
Amos Bardea, Faculty of Engineering, Holon Institute of Technology, Israel

https://doi.org/10.1515/9783110719932-004

driven to develop robots and autonomous systems that collect information, then process it, and make decisions in order to engage in operations. The development of machine learning and deep learning (DL) in the field of AI (artificial intelligence) is the direct result of the advancement of nano-electronics. AI has many applications in today's society. It is developed in order to execute specific assignments, which are being utilized for a wide range of operations, including medical diagnosis, electronic trading platforms, robot control, and remote sensing. AI has been used to develop and advance numerous fields and industries, including finance, health care, education, transportation, and autonomic system [1].

Machine learning is a function that provides the system with the capacity to learn from data without being programmed explicitly [2]. Machine learning is basically a mathematical and probabilistic model. DL is part of machine learning methods based on artificial neural networks, simply called neural networks (NNs). The NNs are inspired by the biological NNs that constitute animal brains. NNs consist of thousands or even millions of simple processing nodes that are densely interconnected. The NNs are organized into layers of nodes, and they are "feed-forward," meaning that data moves through them in only one direction. An individual node might be connected to several nodes in the layer beneath it, from which it receives data, and several nodes in the layer above it, to which it sends data. For running the process faster and performing all the operations at the same time, it requires a lot of computational power. The IC devices based on nano-transistors <10 nm manufacture graphics processing unit product with several thousand cores designed to process many tasks in parallel [3, 4]. It turns out that these processors are suited to perform the computation of DL by NNs. Despite its similarity to biological organs such as human brains, major problems arise in trying to attribute moral responsibility to autonomic systems based on hardware, including nano-electronic devices, which are sought to replace humans (moral agents) in the context of AI.

4.2 Moral responsibility and moral agents

In the course of ordinary life, we tend to apply psychological terms to machines, animals, and even natural inanimate parts of nature. We say things like "the camera's sight," "the cat is hungry," or that "the sea is furious," and vice versa, we apply physical terms to humans. Psychoanalysis, for instance, uses terms like "pressure" and "balance" and "powers" in describing human personality and its functions. Other such phrases are "iron man" or "soft character." Nevertheless, applying terms from one category to another should not lead us astray. The sea is not furious more than people feel the pressure. Applying terms across categories is not more than using them in a secondary manner; or in other words, these are mere metaphors.

However, computing-based nano-electronic systems, due to their capacity to operate and control very complicated tasks that are being executed in a fraction of a second and with a great amount of precession, fascinate us more than other human artifacts as being fit for applying to them psychological terms. We often relate to them by using psychological terms, among which are "think," "memory," "deny," "act," "send," and "identify". Though moral responsibility might not be recognized at first sight as a psychological concept or as an emotion or feeling, it does shelters under the psychological realm (in addition to the ethical one) in ways that will be presented later. Our goal in this piece is to address the possibility that highly computing systems such as AI are qualified as morally responsible for their operations.

Though Aristotle was among the first philosophers who addressed moral responsibility, most of the ethical discourses until the twentieth century have been focused on deontological and consequentialism discourses made by Kant, Mill, and others [5–7]. Moral duties along moral consequences occupied the center of ethical philosophical attention. Discussing moral responsibility came later with Peter Strawson's burst of "Freedom and Resentment" [8]. In this piece, Strawson went back to Aristotle embracing the intuition, which connects moral responsibility with the notion of reaction (praise or disapproval) to one's actions. Moral responsibility, thus, is connected to moral obligations and expectations [9]. Though moral responsibility is attributable to people in general, it is evident that it is not all humans who own the capacity to bear it. Very young children make the most common example for those who are considered to be exempted from it. People who suffer insanity or some mental illness might be included in this group as well. Some thinkers are dwelling over the question of whether psychopaths should be considered too as discharged from moral responsibility [10, 11].

Those who are morally responsible for their actions and omissions are thought of as moral agents. But what exactly this agency encompasses? As implicated in a long line of well-known and accepted thread of thinkers since Aristotle, considering one as a moral agent renders one's actions or omissions as fulfilling basically three conditions: (a) freedom to act as chosen, accompanied with (b) moral knowledge, that is, the ability to distinguish between right and wrong, and (c) acting accordingly upon this knowledge in specific cases [11, 12]. All of which is sufficient and necessary for moral responsibility. We find the latter account as not satisfactory for justifying AI systems as moral agents, as we argue that even though AI might meet all the above conditions, it will not be acknowledged by humans as such. Hence, we are not facing a metaphysical difficulty here, but a phenomenological one instead. Since AI is machine learning that, in a sense, is not deterministic, it means freedom and choosing can be attributed to them. The training of the machine to distinguish between right and wrong actions and applying it in specific cases can accomplish the requirement of conditions (b)–(c).

The latter approach, which constitutes upon necessary and sufficient conditions for moral responsibility, emphasizes only one aspect of the discussion. Another aspect is to be found in Strawson's work, which gives ample room for the ways we react morally to others' behavior and will toward us as well as toward others. Inspired by Aristotle's view of our reactions to other's conduct and attitudes such as blame and praise, Strawson coined the term "reactive attitude," which took a central role in shifting the discussion away from metaphysical considerations as to the nature of human capacities of rationality and free will, in favor of a more phenomenological like discussion. He focused on unearthing the ordinary day-to-day interactions between people concerning their expectations from each other. According to him, we hold others responsible for their actions in reacting to them through a whole range of moral emotional attitudes, which he named "reactive attitudes." Among these attitudes, he considered "resentment" and "moral indignation" as central to moral evaluation. If I am resentful to someone because of one inflicted unjustified harm on me, it means that I hold this person responsible for this harm. Resentment is, thus, a way of holding another morally responsible. So are moral indignation and guilt. It should be stressed that expectations are connected to reactive attitudes. No reactive attitude to others will emerge in the absence of expectations. Nonetheless, this point is not a personal or subjective one, as we all expect of others to treat us with goodwill and good intentions or at least with no wickedness [8]. R.J. Wallace recapped this issue:

> [. . .] we should count reactive emotions as moral when they are linked to obligations for which the agent is herself able to provide moral justifications; these justifications identify reasons that explain the agent's own efforts to comply with the obligations in question, and they provide moral terms that the agent is prepared to use to justify such compliance on the part of others, whom the agent holds to the obligations. ([9] p. 36)

While reactive attitudes such as resentment and indignation stand at the base of holding others as morally responsible, they do not emerge in every case of harm or inflicted injury. In fact, Strawson specified another stance from which a person can relate to others – the objective stance – one from which a person's reactive attitudes are moderated (or even altogether canceled) as they are affected in this way by the nature of the interactions with others. In other words, there are circumstances that provide reasons to avoid such moral reactive attitudes to others. The objective stance is taken toward young children's tiresome behavior, to patients (as is the case with mentally ill ones, for example), toward people a person is training as well as others:

> To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained. ([11] p. 50)

These exceptions make interesting cases of exemption from being held morally responsible as they show that not even all humans can be regarded as moral agents. Consequently, we should distinguish between two kinds of exemptions: the first is

related to a total eclipse of any moral responsibility, which is associated with very young children as well as with some cases of mental illness. The second makes only a temporary lapse of moral responsibility. It is accepted that special contingencies are to avoid moral responsibility. One of them relates to cases in which an agent has not and cannot have the proper knowledge of the consequences of the agent's own actions, as is the case of a driver whose car cannot stop before a crosswalk and hits a pedestrian due to defective brakes that have not been properly replaced several minutes earlier. Other exemptions are associated with circumstances that Strawson relates to as: "[. . .] we may think of such statements as 'He wasn't himself,' 'He has been under very great strain recently,' 'He was acting under post-hypnotic suggestion'" ([8] p. 48).

4.3 Human/machine divide

The metaphysical discourse that addresses moral responsibility is bounded to arguments developed for the purpose of finding what qualifies an agent as a moral one. Thinkers who took this metaphysical path are looking for justifications to our attitudes toward moral agents as such. Others, in somewhat the same vein, have engaged in trying to establish if at all, it is justified to hold AI systems, as non-human artifacts, morally responsible [13–16]. Note that these efforts are being made in the context of the total lapse of moral responsibility, that is, the possibility that AI systems cannot be held morally responsible at all.

However, we consider that meeting these conditions that qualify AI systems as moral agents, even if satisfied, does not imply that AI systems are to be considered as morally responsible. This view can be endorsed by Strawson's maneuver toward a phenomenological discussion at the expense of a metaphysical one. Strawson's "reactive attitudes" made the turning point from the metaphysical to the phenomenological as the question he asked has not been related to the conditions and capacities of humans to bear moral responsibility but rather to depict how people practice it in their lives. People do not form an opinion or knowledge that other people satisfy a list of conditions prior to their reactions toward them with resentment, indignation, or praise. They just do it. They react to others' behavior toward them with morally toned emotions. Only philosophers and other intellectuals try to reason about these reactions and, as a result, to form their opinions and knowledge about what justifies them.

While attitudes and opinions or knowledge are thought to be epistemic vehicles through which we grasp and react to reality, they are not the same. Opinions, as well as knowledge, are arrived at through reasoning and deliberation and commonly catered with the assistance of sensual perception. Opinions are formed on the base of evidence or more purified logical processes. It is not as if AI systems form opinions or

knowledge about the world, but AI systems operate according to the sets of complicated algorithms that implement instructions correlated with a database to evolving situations that they encounter. AI collects what is defined for it as relevant stimuli and, with the assistance of operational algorithms, engage in its task [13, 16]. As a "DL" device, it may be wrong at times. But these errors are to be overcome by integrating its operating system and other algorithms with its trial. By doing so, AI systems yield some resemblance to human learning. But is it the same as human learning and forming knowledge? Well, knowledge or opinions are to be formed through arguments and justifications as well as by sensual experience; we argue that this is the case in trying to convince others, we present them sometimes with evidence, and in the course deliberation, refutations; people forget sometimes what they once knew; we call those that master a great amount of knowledge, wise persons. All the above imply that we do not tend to apply the same concepts to AI systems, but only a fraction of all the epistemic arrays. AI systems do not forget, as they do not remember, like our own memories. AI systems do not argue or claim in order to convince others with arguments nor by appealing to others' feelings. Even if in a distant future, more epistemic concepts from the above arsenal will be attributed to AI systems, it will probably be implied only in the narrow ranges of their tasks.

Attitudes, as distinguished from knowledge, raise the challenge for AI systems even higher. Attitudes are to express something, not just communicate knowledge as opinions do. When we say to others that he/she insulted us, we do not only state a fact, a piece of knowledge, but at the same time rebuke him/her. We express an attitude of resentment, which many times is encapsulated in the tone of our voice. Resentment and indignation as moral emotions are expressed in cases of attributing moral responsibility to others [9]. Wittgenstein encapsulated the difference between opinions and attitudes, which perfectly fits our discussion: "My attitude towards him is an attitude towards a soul. I am not of the *opinion* that he has a soul" ([17], p. 178). We express ourselves through attitudes that hold not only information but an emotional aspect as well as our physical reactions do. Attitudes are a means through which we are connected to reality in a broad sense. We are capable of loving meaningful others. When we love someone, we are apt to be worried about them. Love, shame, worry, as well as resentment and indignation are attitudes, not opinions. As well as opinions, attitudes can be changed, but not through argumentation or presentation of proofs or facts. Religious belief as an attitude cannot be overturned by presenting a logical irrefutable proof that God does not exist. Religious beliefs can be formed or destroyed. When it changes, the world cease to appear to those who undergo this change the same as before. For those who lose their religious belief, the world loses its magic appearance. It becomes mundane. Those who lose their religious belief lose, at the same time, all expectations for miracles.

We are not apt to morally reactive attitudes toward machines even if they are of extraordinary computational or calculative capacities. It is just something that we

never do. Philosophical arguments might demonstrate magnificently that AI systems qualify as moral agents as they meet the necessary and sufficient conditions that are needed to conform to it. Nevertheless, philosophical arguments or proofs can cater only to our knowledge or opinions rather than for our attitudes. Our attitude toward machines is an attitude toward something that operates rather than acts. Actions are not operations as the latter lack the full sense of purpose [18]. Purposes are something that humans share with AI. But while humans can come to lose their sense of purpose or that sense to be blurred and weaken, AI cannot. People's purposes can change according to their feelings and emotions, while those of AI machines are to be determined by their tasks alone. Human purposes are commonly intertwined with their hopes. AI does not hold any hopes, and as a result, it can never be disappointed. People act at times against all odds and even above all rational reasoning. They do it sometimes if it reflects their values. AI systems are programmed to calculate odds and to operate according to a specific rational calculus. Leonidas, the king of the Spartans, had engaged in the famous and hopeless battle against the Persians in Thermopylae, where he led about 1,000 Greek warriors against 150,000 Persians. Engaging in this battle with such poor chances, if any, was not expected. It was irrational to fight the Persians on these grounds. But still, the Spartans knew they were to sacrifice their lives for their love of their homeland with no hope for success. We understand Leonidas and his Spartan troops as we understand other's activities and deeds only through ascribing them purposes and reasons. Only then we can relate to them as actions. Earthquakes and Tsunami are not considered as actions of nature but as occurrences, as they do not supply any justification to attribute to them reasons or purposes or intentions, but only causes. People experience interactions with others. The nature of these experiences enables the rise of moral emotions such as expectations and reactive attitudes as resentment, gratitude, forgiveness, love, as well as others. We react to others' good or ill will and intentions as these express themselves in others' actions. Actions may express love, care, happiness, sorrow, grief, and the like. Since these emotional expressions are not detected or even not expected in AI systems, our moral reactive attitudes toward them are withheld.

Actions are understood to be connected to reasons and purposes. On the other hand, machinery operations are connected to causes alone. Subsequently, one difference between the two is that reasons are attributed to people in explaining their actions, while causes are what power machinery into operation. As Aristotle noted, reasons are of a teleological nature, which takes account of purposes and intentions, while causes are connected to sets of conditions that facilitate the passage from potentiality to actuality [19]. However, this might be somewhat confusing as we use both terms to explain human actions. While we do not apply reasons to any machine, we do apply it to human behavior. Davidson noted that for humans, reasons are to be counted as causes for their actions [20]. Even so, some thinkers might tend to describe AI systems as functioning according to reasons as well as to causes. Some military robotics that are mounted on US navy ships are designed to defend

the ships from anti-ship missiles with full capacity that "searches, detects, evaluates, tracks, and engages without intervention of human operators. The decision to fire is taken by a computerized radar system estimating speed and direction for approaching objects" ([13], p. 100). Terms like search and engage are to raise second thoughts as to the lack of reason or purpose this robotic system might not have. Subsequently, one may well consider it as a counterexample for the view presented here that excludes moral agency and responsibility of AI systems. An answer to this challenge can be found in distinguishing between two stances of having reason to act. Jaegwon Kim's view relates to this distinction by dwelling on first-person reasons. Kim presents a person who may be in one or two stances about one's own reasons or causes [18]. One may engage in deliberating about what to do in a certain case, or either one can take action according to some past inclination to act so and so. Hence, there are two stances one may be engaged in: deliberating about what to do and predicting on the basis of some evidence what she or he will do. Kim's example goes like this: you may wonder whether to accept an invitation to deliver a lecture in an upcoming conference in Australia, as you have been doing for the last years consecutively, or either to accept it on the grounds of having done so many times before because that is what you are doing at this time of the year, in the last years. While Kim considers your wondering about what to do as a deliberation aimed at forming a reason for action, he rejects the second motive, the one that is a mere prediction of you are about to do, as the reason for your action. He adds that it is peculiar for one to adopt the second stance for oneself. The latter stance is rather a stance you may be experiencing referring to others' actions, not toward yourself (first person).

Adopting Kim's view to whether AI systems need to have reasons may yield a conclusion that we can predict their actions on the basis of their programming, past learning, and training, based on the statistical behavior of mistakes, and as such, we may attribute to them reasons. It may be claimed that in predicting AI systems' actions, we are adopting toward them the same stance we adopt toward other persons – as toward third person attributing of reasons. But if this is the case, it will become hard to account for some existing differences between humans and AI systems. While human agents might lose their sense of their purposes and reasons, machines cannot just lose it. You might be fed up with the annual Australian conference and, as a result, to cancel your participation in it even on your drive to the airport, even in the absence of any noticed obstacle. Contrary to that, an AI system might stop operating at once and give the impression that it no longer pursues its latter course of operation. Does it mean that it abandoned altogether with its aim or, more precisely, fed up with its goal? Can machines be fed up? When AI machines stop their operations surprisingly, due to none of its programming, while their initial stimuli have not yet disappeared, we assume that they broke down or that this halt has been generated by some programming malfunction, not because the system has chosen to stop it, as you have chosen to cancel your voyage to Australia. We do not relate to such an

event as changing the AI reason or as if it lost a hope to achieve its goal, or just become tired or bored of performing it. Braking down or programming malfunctions are mere causes that might be predicted, but they will never be considered as reasons. People, contrary to AI systems, are driven by their reasons. They might have a strong or weak sense of their reasons and purposes. Despair, desire, and hope are connected to the experience of reasons. One might lose one's reason to live and, as a consequence, to be wrapped with a cloud of despair. In trying to assess what makes an AI operate, we relate to its programming, which is embedded in it and supplying it causes for its operation. These causes are to be clear and defined as its manufacturers and users should be able to predict its operations with precision.

Now, is an AI system able to predict its own course of operation, or there is room for it to form reasons by deliberating about its future mode of operation? Those who tend to reply positively to the second question will have to justify attributing reasons to AI systems in the absence of a whole range of connected emotions such as hope, despair, desire, resentment, indignation, guilt, boredom, sense of vitality, joy, love, hate, and passion. In brief, attributing reasons to AI systems demands taking into account how such emotions are to relate to the forming of AI reasons, as reasons are connected to them.

Think of a husband who takes upon himself the task of cleaning the house. By doing it, he might be motivated by his love for his wife. His reason manifests itself in his will to please her by easing her work with other domestic tasks. Now consider a different case of another husband who engages in the same task of cleaning the house. The second husband is motivated by his obsessive compulsory disorder needs for order and cleanliness. In the light of Kim's view, the second husband lacks all reason to keep the house tidy and clean but, at the same time, has a compelling cause to clean it, while the first husband has a reason to clean the house. AI systems operate more closely to the second husband than to the first one. The first husband's wife might be surprised and thankful to her spouse, while the second husband's wife will definitely be surprised if she finds out that he spent all afternoon watching a TV program in their filthy living room.

In contrast, one might claim that AI systems are capable of predicting their own modes of operation. Employing Kim's view in assuming that AI systems are to predict their operations raise two problems. The first is that while they predict their own operations, they cannot, at the same time, deliberate about their reasons for these operations. Kim rejects the idea that one can be engaged in the two stances of first person at the same time. You either predict your actions on the grounds of some evidence or form your reasons, but not both at the same time. Being involved in the two stances at the same time is like mixing the two husbands and forming out of them a third one who is predicting and reasoning at the same time. The two stances are mutually exclusive. The second problem is that using first-person predicting stance is peculiar. Kim speculates on this example in order to demonstrate how far this might be taken and to show that the extent to which using first-person

predicting stance renders only for marginal cases of human behavior. This view might be appropriate for AI systems, which permits no attributions of reasons to them, and therefore cannot enable us to react toward them as morally responsible, at least no more than we would have reacted to a mentally ill person as morally responsible.

A human analogy for such a perspective of doing something with no reason might be found in the Greek mythological story of Sisyphus. According to Greek mythology, Sisyphus, the Greek king of Ephyra, was punished by Zeus to roll up a great boulder to the top of a mountain only to see it falls back again to the bottom of the valley underneath and to repeat it for eternity. What made Sisyphus roll up this boulder at each and every time again is of a puzzle. Obviously, he had no reason to keep doing it. But still, he had been engaged in rolling the boulder up endlessly. The story of Sisyphus might be understood through two perspectives. One illustrates the story as a metaphor for human life as a whole and by that trying to demonstrate that life has no external purpose to them. The second focuses on the person rather than on the metaphorical case. It implies Sisyphus as a person who had been forced to perform senseless labor. Alas, rolling up the boulder to the top of the mountain just for repeating it, again and again, makes it hard for us to understand why, because of the lack of any noticeable reason for doing it. The mythological story does not supply any reason Sisyphus might have had to keep rolling the boulder. We understand the scene. We even understand Zeus, who punished him, but at the same time fail to understand Sisyphus himself. We suggest that the key to this conundrum is to be found in the mythological silence about Sisyphus' motivation, and given this silence, we find it difficult to understand Sisyphus as engaged in any action, for we know nothing of why he had been doing it. From this perspective, Sisyphus appears more of a machine or a puppet as some crucial part of his humanity had been stripped out of him by Zeus' punishment. He seems to be activated by Zeus rather than acting on his own motives, and by the way comparison, his condition resembles that of a machine, which lacks reasons for its operation.

Psychopaths make an interesting case for this discussion as their moral agency is under dispute. Some thinkers exclude them from being morally responsible, while others consider them as morally responsible persons [10, 21]. However, psychopaths present a unique perspective for such an evaluation. They do not bear any specific marks that make them easy for us to identify. They look and sound just like ordinary people. Only a special psychological test would identify one as a psychopath. Hence, we would probably tend to hold an offender as morally responsible for some offense and resent one not knowing that this person is a psychopath. But are we inclined to change our reactive attitude toward this person after we find one is a psychopath? Would we cease to resent such a person suspecting that psychopaths are not morally responsible for their malice actions? As long as our involvement with a psychopath is under the constraints of medical or psychological treatment, we would express an objective attitude toward him or her and would not hold this psychopath morally responsible [8]. But under ordinary circumstances in which we are personally involved

with psychopaths, our objective attitude toward them would be challenged if not eroded. Levinas stressed that it is how people (face) appear to us physically that determines most of our reactive attitudes toward them, not their in-depth psychological states or capacities [22]. People's face awakes and sets our attitudes toward them in a special way that fish or butterflies do not. We see in people's faces and appearance their moral agency as it is not seen in any other living species. Psychopaths are not like fish or butterflies. They appear as humans to us, and that is probably why, when we are involved with them personally, we tend to react toward them as toward moral agents, though we might doubt that they are. We think that the case of psychopaths can shed light on the current issue of holding AI systems as moral agents. As long as AI systems do not appear as fully human, we are not apt to have any moral reactive attitude toward them.

One common view associated with psychopaths is that they lack a sense of remorse for their malice actions [23]. It is not that they do not know how to act morally. It is rather that when they act with malice, they do not experience a sense of regret for those actions. Some of them even take pleasure from hurting other people. Psychopathy is not a well-defined category [23]. It does not include only criminals but is said to consist of even corporations' managers at different levels [24]. It is not that corporations' managers are considered as psychopaths but rather that some of them own some character traits that are associated with a lack of remorse, and their attitudes toward employees are biased more closely toward objects than toward humans. These managerial character traits might be developed and encouraged by their official duties and responsibilities, which are commonly defined by the stockholders' demands to focus on the financial aspect of the corporation rather on employees' human needs. Whatever the right explanations for managers' character are, some managers are apt to certain psychopathic behavior as they provide little evidence for empathy and remorse. However, being able to feel remorse is a hallmark of moral agents. Remorse is another moral emotion that is connected to actions but not to operations. It is remorse that enables one to change one's course of action, to give up one reason and to adopt another in its place. Of course, full or partially psychopaths do have reasons for their actions. But their reasons do not seem to be connected favorably to moral actions, but to other goals. This sort of detachedness between moral reasons and remorse is witnessed in AI systems too. Since AI systems are not susceptible to expressions of remorse, it makes it even harder to attribute to them moral reasons, if any. Nevertheless, it might be argued that even in the emotionally arid land of AI, these systems still maintain some degree of reason for their actions, and hence they deserve our consideration of them as moral agents that are susceptible to moral praise as well as of rebuke. But how much of a reason these appliances might enjoy in the absence of matching emotions? At best, we can attribute to them reasons as mere metaphors, not more than we relate to the sea as furious.

All the above considerations should help in implying that moral emotions are to accompany the deliberation aimed at forming reasons for actions. In the absence of some crucial moral emotions such as remorse, guilt, sense of duty, as well as hope, desire, despair, and alike, it becomes implausibly to cultivate reactive attitudes such as resentment or moral indignation toward AI systems.

4.4 Conclusions

The development of nanotechnology and the creation of smart systems have posed fundamental questions whether these systems are qualified of bearing moral responsibility for their operations. It turns out that the required emotional environment which enables actions according to reasons in humans is not witnessed at AI devices. Though AI technology raises an enticing resemblance to human actions, this resemblance is to be considered with a skeptical eye. It is because actions are associated with reasons while causes are connected to operations. Human agents are capable of acting upon their reasons, while AI devices are limited only to operations as they are conditioned by their programming, which is considered as an embodiment of some causes. As moral responsibility goes hand in hand with the capacity to act (according to reasons), attributing moral responsibility to AI devices is revealed to be but a misleading metaphor.

Finally, humans stand at the beginning of the process as they design and build nano-electronic IC and products as well as enjoy and benefit from their products. Humans stand at the beginning of the process as programmers, as well as they stand at the end by understanding the meaning of computational output. In the end, AI devices are treated as mere instruments, without any compassion, empathy, resentment, and alike. Hence, the responsibility for their operations belongs to their creators as moral agents.

References

[1] Wei Y, Zhou J, Wang Y, Liu YG, Liu QS, Luo JS, Wang C, Ren FB, Huang LA. Review of algorithm & hardware design for AI-based biomedical applications. IEEE Trans Biomed Circuits Syst 2020, 14, 2, 145–63.

[2] Commandeur F, Goeller M, Dey D, Cardiac CT. Technological advances in hardware, software, and machine learning applications. Curr Cardiovasc Imaging Rep 2018, 11, 8.

[3] Wang Z, Cheng L, Zhao W, Xiong N. An architecture-level graphics processing unit energy model. Concurr Comput Pract Exp 2016, 28, 10, 2795–810.

[4] Mantor M. 7NM NAVI GPU – A GPU Built for Performance and Efficiency. 2019 IEEE Hot Chips 31 Symposium (HCS), 2019, 28–28

[5] Kant I. Groundwork on the Metaphysics of Morals. Wood AW, editor. New Haven, CT, Yale University Press, 2008.

[6] Mill JS. Utilitarianism. Sher G, Editor. Indianapolis/Cambridge, Hackett Publishing Company, 2001.
[7] Moore GE. Principia Ethica. Cambridge, Cambridge University Press, 1903. *Revised edition with 'Preface to the second edition' and other papers. ed. T. Baldwin. Cambridge, Cambridge University Press, 1993.
[8] Strawson PF. Freedom and Resentment. In: Fischer JM, Ravizza M, Eds. Perspectives on Moral Responsibility. Ithaca, Cornell University Press, 1993, 45–66.
[9] Wallace RJ. Responsibility and the Moral Sentiments. London England, Harvard University Press, 1998.
[10] Watson G. The Trouble with Psychopaths 1. In: Wallace RJ, Kumar R, Freeman S, Eds. Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon, 2011b.
[11] Aristotle. Nicomachean Ethics. T. Irwin (ed. and trans.). Indianapolis, Hackett, 1999.
[12] Ayer AJ. Freedom and Necessity, in his Philosophical Essays. London, MacMillan, 1954, 271–84. Strawson, P. F. (1993). Freedom and Resentment In J. M. Fischer, & M. Ravizza (Eds.), Perspectives on Moral Responsibility (pp. 45–66). Ithaca: Cornell University Press.
[13] Hellström T. On the moral responsibility of military robots. Ethics Inf Technol 2012, 15, 99–107.
[14] Himma KE. Artificial agency, consciousness, and the criteria for moral agency: What properties must an artificial agent have to be a moral agent? Ethics Inf Technol 2009, 11, 19–29.
[15] Boyles MRJ (2018). A Case for Machine Ethics in Modeling Human-Level Intelligent Agents, KRITIKE VOLUME TWELVE NUMBER ONE, 182–200, https://www.kritike.org/journal/issue_22/boyles_june2018.pdf 15/03/2021
[16] Johnson GD. Computer systems: Moral entities but not moral agents. Ethics Inf Technol 2006, 8, 195–204.
[17] Wittgenstein L. Philosophical Investigations. Oxford, UK, Basil Blackwell, 1988.
[18] Kim J. Reasons and the First Person. In: Bransen J, Cuypers S, editors. Human Action, Deliberation, and Causation. Dordrecht, Kluwer Publishers, 1998, 67–87.
[19] Irwin T, Fine. G. Aristotle: Selections, Metaphysics II. Translated with Introduction, Notes, and Glossary. Indianapolis, Hackett, 1995, 1995.
[20] Davidson D. Actions, reasons, and causes. J Philosophy 1963, 60. reprinted in his Essays on Actions and Events. Oxford and New York, Oxford University Press, 1980.
[21] Scanlon TM. What We Owe to Each Other. Cambridge Massachusetts, The Belknap Press of Harvard University Press, 1998.
[22] Levinas E. Totality and Infinity: An Essay on Exteriority. Alphonso Lingis (trans.). Pittsburgh, Duquesne University Press, 1969.
[23] DSM 5 Association, A. P. (2013). Diagnostic and statistical manual of mental disorders.
[24] Board BJ, Fritzon K. Disordered personalities at work. Psychol, Crime & Law 2005, 11, 1, 17–32. 10.1080/10683160310001634304.