Man-object, Interpersonal and Man-machine Relationships

— An Ethical Perspective on Artificial Intelligence

He Huaihong*

Abstract: Traditional social ethics has always been centered on human relationships. In recent years, modern ethics began to systematically reflect the relationships between humans and objects, and the future ethics will need to account for the relationships between humans and intelligent machines. This is mainly because humans may be overtaken by machines in intelligence through which humans gain dominance over all other natural objects. On the ethical thinking of the man-machine relationship, an idea is to be inclined to do subtraction rather than addition. Specifically, we should give priority to and focus on limiting the means and abilities of intelligent machines rather than how to cultivate and set the value judgments of their friendliness. In other words, we should concentrate on how to limit the development of intelligent machines to specialization and miniaturization, especially keeping them within the scope of non-violence.

Keywords: man-object; intelligent machines; non-violence; inequality

Ethics is about judgments regarding living values and behaviors that involve all kinds of human relationships. Briefly, human world is right a world of relationships. By relationships, the paper means interactions between humans, between humans and nature, between humans and themselves, and between humans and transcendence. Now our world is undergoing a momentous change — the rapid development of artificial intelligence has introduced a new existence into the scope of ethical thinking: intelligent robots. Following our own separation from nature as natural objects, there is a real possibility that these intelligent robots could also grow independent from nature. Born as objects but created by humans, robots combine the features of humans and objects. Is it possible that robots might replace humans and become the new "super species?" This is the biggest challenge posed by artificial intelligence and the possibility will be explored otherwise. This paper will discuss

* He Huaihong, Professor and Ph. D. superviser, Department of Philosophy, Peking University.

how humans should approach the man-machine relationship, what we can do for intelligent robots now, and what basic ethical ideas and norms can be put forward.

Man-object relationship

Apart from humans, there are also objects on the earth. As a matter of fact, humans were born as a kind of object, and to this day, they still fall into the category of object or common existence in general terms. But in terms of features, abilities, perceptions and morality, humans stand out from or rival any other objects on the planet. In this way, the relationships between humans and objects came into being, although it was not until the late 20th century that the relationships between humans and natural objects were ethically contextualized as an ethics theory system or ecological philosophy system.

The paper will start with a review of the man-object relationship based on different time spans: the history of the earth, creatures, animals, humans and civilizations, in an ascending order by scope. The first three, of course merely a kind of human prehistory, are offered to give more insights into the natural origins of humankind.

Man apes evolved to walk upright, emancipating their two hands to hold and make instruments with their flexible fingers, and also created fire to cook and preserve food. In particular, the invention of fire also accelerated humankind's cerebral development, and offered them a production tool by, for example, driving away and burning animals with torch. Maybe humans were born social animal, and became even more and more aware of teamwork after they can perceive. 200,000 to 300,000 years ago when it was still in the Stone Age defined by foraging livelihoods, modern homo sapiens already killed many species. They cooperated with each other to drive other animals into the valleys by using torches, shouting, stones and sticks, with most of these animals being left died and only a small proportion being kept behind for food. Notably, large terrestrial animals were humans' top prey. For a long time, then humans were very inexperienced and had no good control of flying birds and swimming fishes, so they were more adept on catching large animals than small ones. According to recent researches, homo sapiens started from the East Africa roughly, and arrived at Asia, Europe, Australia and America. No matter where they got, there were some large animals suffering a sharp increase or even extinction, and even some indigenous human species disappeared.

This may be the first stage of the man-object relationship — humankind stood out from other animals and can leverage instruments and brains to confront and contend against any species or the alliance of all other species. The second stage of the man-object relationship started when humankind entered the agricultural civilization age more than 10,000 years ago. During the period of the history of civilizations, humankind evolved over time into the master of the earth by virtue of abilities of rivaling and manipulating all other species. As a result, humankind no longer lived on foraging, but started to rely on crop farming and animal husbandry to get food and energy. Meanwhile humankind also began using other animals as tools, not to mention lifeless objects such as stones. By using other animals to become more able-bodied, they started transforming the forms of natural objects to serve their purposes rather than being content with what these objects were. Consequently, those crops and animals were completely domesticated, having a far cry from what they used to be.

As the human social communities grew larger and larger over time, the economic growth started to

maintain stable and remain on track, and thus a small proportion of the leisure class can be supported to specialize in cultural wellbeing, until cities, characters, metal instruments, nations, and even the spiritual civilization of the Axial Age emerged. Then modern industrial revolution followed, bringing about another take-off. During the period, humans invented steam engine, internal combustion engine and electricity, and designed and churned out a raft of machines by using coal, gasoline and other natural resources, greatly improving human capacity to conquer nature. Compared with the transformed products that still had something reminiscent of what they used to be in the agricultural civilization age, those in the industrial revolution age can hardly find something in common with their former appearances. That somewhat implies that humans have undergone a drastic change and a great improvement in their methods to get food and energy. Therefore, these man-made machines were no longer natural objects, but artificial objects without self-intelligence and abilities of self-learning and self-improvement. Once humankind became the real master of all natural objects across the planet, they can easily conquer any other species or the alliance of these species, and remove the mountains and reclaim the seas, giving the nature a complete facelift.

Then by virtue of what have humans made it? Throughout the process, humans made no progress in physical power, but degenerated in some certain aspects instead. Even till today, humans are still no match for some existing animals in speed, strength, endurance, flexibility, among other aspects, but can totally dominate them. Obviously, it is not physical power but brainpower — or more specific, the dominant violence and might accompanying brainpower — that humans rely on to conquer animals. Hence, disparity has naturally prevailed among the relationship between humans and other animals for a long time. Until recent years, humans start to self-examine and mend their ways with very limited effort, it is, however, impossible to change the unequal nature of the relationship.

With brain power, humans were more empowered physically, to the extent that they can make what they want happen without using their own physical power. Hence no matter how small a notch humans were above other animals, clever humans were bound to make the notch larger and larger. Throughout human evolution process, brain power has played a role — or a dominated role from a modern perspective. By "brain power," the paper obviously doesn't mean whole human perception ability, but focuses on human abilities of perceiving and controlling the material world vis-à-vis the spiritual world, given that humans have also been trying to gain insights into the essence of the world, the meaning of the life, aesthetics, art and other fields. Maybe such human abilities in these spiritual fields even better explain why disparity prevails among the relationship between humans and other animals and why humans are different from other animals by nature. But humans still mainly depend on their ability of controlling the material world to establish superiority and dominance over other animals.

As humankind entered the civilization age featuring the appearance of head worker class, nation and language, brain power saw a drastic and rapid improvement, giving rise to accelerated and even exponential human development: the earth has a history of more than 4 billion years; the creatures more than 3 billion years; the animals about 700 million years; the humans nearly 3 million years; the modern homo sapiens 200,000 years; the civilizations more than 10,000 years; the nations about 5,000 years; the industrial revolutions 300 years; the emerging high-tech civilizations, also called "intelligent revolution," only 50-60 years.

Until the civilization age when humans became well aware of self and "difference between humans and

other animals," a real kind of moral relationship between humans and other animals emerged, but there was still no systematic perception of and intentional adjustment to the relationship which may have to wait on more advances in civilization. Furthermore, even in the age, humans would now and then relapse into the animal-like competition mindset for survival. When that happened, morality hardly made any sense for both parties, and it was also difficult to pass moral judgment on either party, but it was possible to make moral judgment about why humans relapse into such an exceptional state of mind and make every effort in remedy and adjustment.

With regards to adjusting the moral relationship between humans and objects, "moral standing" jumps high on the agenda. When humans conquered other species, especially animals, they didn't, even saved the trouble of having a try to, fully understand how other animals felt and experienced, while they bore them no malice, for humans captured and ate them not out of hate. Unlike humans, animals, creatures, and even all other natural objects have no self-awareness, so they cannot become moral subject. But does that mean humans can treat them arbitrarily? Can they acquire a certain moral standing from humans? And on which is the moral standing based?

Answers to the last question vary according to ecological ethics theory, but most hold that other species or objects are also of intrinsic or innate value. In this connection, albeit they have no self-awareness and cannot be moral subject, they with the intrinsic value should be treated as moral object, or moral patient, and humans should be their moral agent.

Of course, apart from "moral standing," there is also another idea of "moral importance", i.e., other species or objects with moral standing have varying degrees of moral importance. For instance, animals seemingly able of feeling should enjoy more privilege, so abusing animals should be the first to be blacklisted. Living objects and then all other lifeless objects follow. Of course, the whole natural environment can be treated as an eco-system.

Interpersonal relationship

Addressing the relationship between humans, namely interpersonal relationship, is the main focus of ethics, especially traditional ethics. This kind of relationship is easy to be narrowed down to the relation between individuals, and Chinese traditional ethics especially focuses on the relation between relatives. But the interpersonal relationship should include three aspects in the broadest sense of the term: First, the relationship between individuals or between selves and others, like all kinds of relations between an individual and his/her relatives, friends, acquaintances and strangers; second, the relationship between individuals and organizations, large or small, formed on the basis of region, race, culture, religion, politics, interest and most importantly, nation; third, the relationship between human communities, for instance, the relations between religious organizations, even "human generations" and, also most importantly, countries or political communities.

Are the interpersonal relationship and behavior on track towards moral improvement roughly? Back in the foraging civilization age, many small primitive human communities were formed, where equality dominated internally but violence or even atrocity (if this word that smacks of moral judgment is applicable) externally. Till the agricultural civilization age, nation was created. Externally, conflicts between political societies were still quite frequent, but less hostile as those in the primitive age; internally, peace and lenience were further

consolidated. Thanks to the political order, humans felt more secure in their safety, livelihood and education, but the improvement was based on a certain hierarchy system. In the modern society, moral regulation saw a constant expansion in the influence sphere: everyone started to be considered as an equal individual, as the society traversed its long uneventful history from existence equality to personality equality, and to liberty and equality of basic human rights. Decrease in violence was another overall trend. Despite of those highs and lows over the development course, such as World War I and World War II, as well as many civil wars and riots in the first half of the 20th century, violence saw an overall drastic decrease after World War II, especially in the developed and fast-rising countries. Furthermore, the trend can also be seen in families and schools, with bullying and corporal punishment being reduced day by day towards extinction. As the subsistence guarantee system kept improving, people's access to food and medical services were improved, plagues were on the track towards extinction, and human life expectancy was extended on the whole. Amid aforementioned changes, moral regulation was expanding its influence to all creatures and natural objects, although the influence may vary in levels of intensity. Maybe this improvement in the man-object relationship can be considered as an extension of the improvement in the interpersonal relationship.

If we measure human moral progress based on violence and equality respectively[①], we would get utterly different results. Violence wise, the result may come as a roughly smooth line — in the prehistory, or the foraging civilization age, violence was very frequent and atrocious; in the agricultural civilization age, violence was decreased; in the industrial civilization age, despite of many highs and lows, violence was on track towards overall decrease till today, although weapons of mass destruction which may devastate humankind for decades of times still stalk the world. But equality wise, the result turns out to be a tortuous curve: in the foraging civilization age, humans discriminated on the basis of community membership, with community members enjoying high equality; in the agricultural civilization age, inequality dominated basically; in the industrial civilization age, equality was exercised to all social members in all aspects.

Violence and equality are not only two most important standards to observe the interpersonal relationship, but two key standards to examine the man-object and man-machine relationships. Anti-violence and illegal forced conduct which underscores the two basic principles of life and liberty constitute the core of codes of ethics of all cultures and religions, like "Four Don'ts" of the Ten Commandments, Golden Rule of the Christianity, and Loyalty and Forgiveness of the Confucianism. No doubt some kinds of forced violence like nation are still indispensable, but with the "meet-violence-with-violence" aim, nation was born right to address human violence. Nation, as a kind of violence, may also be abused, but if the aim can be adhered to and the necessary pre-implementation procedures can be gone through, it could receive currency among human communities.

So far, humankind have undergone a process from the primitive community age defined by discrimination with community members enjoying high equality while non-community members suffering violence, to the agricultural civilization age featured by inequality and less violence, and to the industrial civilization age characterized by broad-based equality and even less violence. For the constant improvement in the man-object relationship over recent years, the man-object relationship still cannot be put on such an equal footing as the

① The following discourse on historical ages is based on Ian Morris's *Foragers, Farmers, and Fossil Fuels: How Human Values Evolve*, published by CITIC Press Corporation in 2016

interpersonal relationship can. Although emerging ecological ethics theories, especially non-anthropocentrism theories represented by animal rights and liberty theory, try hard to redress the balance in more favor of vulnerable lives, it is physically impossible to achieve true equality between humans and animals. Perhaps it is not only possible but even unnecessary from the perspective of human morality. Some certain holisticism theory under the framework of ecological ethics may make more sense, as we associate it with the universal reason of ancient Stoicism, but it still has to give humans more weight.

Maybe the fact that humans and natural objects belong to different species fundamentally explains their unequal relationship. Natural objects have no power of reason and self-awareness, yet all objects that have consciousness or have only feeling and life are instinctively inclined to self-preservation. Even under the holistic scenario, all lives shall live together and all kinds of existence shall have a symbiosis, but each life, either involuntarily or voluntarily, tends to put self-preservation first, not other species' existence, which is actually understandable. In this connection, humans can be morally required to spare as most attention as possible to the existence of other species, rather than take better care of the existence of other species than that of themselves. If not, humans would lose their human nature, even their object nature.

Then can we also examine the man-machine relationship from the two aforementioned standards in an effort to establish a certain kind of relationship making for a sharpest decrease as possible in violence?

Man-machine relationship

With the successive invention and rapid development of state-of-art technologies like computer, network, robot, biotechnology and nanotechnology, humankind is in the middle of a new round of industrial revolution, whose key may be called "digital revolution," "algorithm revolution," or "intelligent revolution" in a more comprehensive term. If previous industrial revolutions can be defined as strong efficiency enhancers to solve physical impossibilities for humans, today's intelligent revolution then is speed accelerator to solve mental impossibilities for humans. In retrospect, brain power has played a certain role throughout human evolution process, and as far as the paper is concerned, it seems to continue to play a role, even a dominated role in the modern era.

Driven by these technological revolutions and innovations, humankind is stepping in a new age defined by emerging intelligent revolution which may help human civilizations get rid of such labels like "industrial revolution" or "industrial civilization." Currently, breakthroughs in advanced technologies are still put under the broad category of "industrial civilization," yet the new innovations driven by the intelligent revolution could be classified independently in the future, thus defining a new civilization age. Meanwhile the "industrial civilization age" may still be considered as a human civilization period defined by utilizing and transforming natural objects, while the future AI-centered civilization age may represent a completely new period characterized by innovative artificial creations. Then a new problem about human ethics would arise therefrom: how do humans deal with the man-machine relationship?

As a matter of fact, humans started to think of the relationships between humans and objects or humankind and nature from the perspective of survival and development strategies and technologies from the beginning, and humans also started such thinking quite early from the perspective of spiritual culture. For instance, the earliest Greek philosophers already attempted to learn about the essence, composition,

elements of the nature, and the commonality and difference between humankind and nature and the right relationship between humankind and nature. For another instance, Chinese ancient thinkers also put forward such propositions as "Tao follows the way of nature" and "man is an integral part of nature." As shown above, ancient humans also ever put forward and introduced many rules to protect nature and ecology, and recent decades have witnessed the appearance of a systematic philosophy on environmental ethics. But it is hard so far to say that there is a systematic ethics theory about the man-machine relationship. Why? The most direct reason, of course, is because: AI is posing challenges to human ethics in just the past few years.

However, based on the grounds supporting the man-object relationship which is also quite new to human systematic ethics, we can expound on the question: Why humans used to attach little importance to the man-machine relationship from the perspective of ethics? Building on a survey of ethic discourses calling for attention to the man-object relationship put forward by the environmental philosophy over past years, the arguments in favor of taking good care of other living things and the whole ecology roughly include the following aspects:

First, feeling, mainly applicable to animals. Like humans, animals can also feel pain, largely physically but also psychologically, covering both the suffering animal and its companions. For example, if a goose is shot down, its companions will also feel pain, hovering around while whining again and again, not to mention the wounded goose. Second, lives, including plants. If you pick a flower, it will wither soon; if you uproot a tree, it will lose its exuberance forever. Third, integrity, covering all natural objects, especially those on the earth. All things, either alive or lifeless, constitute the ecological integrity where humankind lives. Therefore, all these things are nearly interdependent holistically. Fourthly is naturality. As nature was born before humans, natural objects can exist without humans, yet the same applies in reverse by no means. Lastly are human feelings towards nature perhaps. Nature often makes humans feel beautiful and even can generate a sense of grandness, abstruseness, solemnity, and awe in humans' hearts. Building on the aforementioned reasons, the paper argues that humans should not only be kind to animals and other lives, but lifeless objects. Though objects are lifeless, humans should also try to protect their originality and naturality by conserving some original wilderness, wetlands, snow peaks and others, if they want to maintain the balance of the whole ecology system.

However, all these reasons mentioned above seem not to apply to artificial machines and robots. Made from silicon-based materials, or with metal alloys added, they have no body sensitivity shared by humans and animals; without the ability of self-growing, self-evolving and self-producing, they seem to be lifeless, and they are also not a part of nature, since they never exist in the nature before, and are just artificial objects made from some natural materials; they cannot make humans feel beautiful by nature — or in other words, whether they are beautiful or not all depends on human aesthetics and design — and cannot also generate a sense of grandness and awe in humans' hearts. Therefore, machines fared even worse than natural objects, for humans used to treat them more arbitrarily out of necessity: dismantling, scrapping and disposing them. Few people had the thought of being as kind to machines as they were to animals, though there were some people willing to spend time on machine maintenance, but they did that only for enjoying better and longer services. Doubtlessly, no one really hates machines. As a matter of fact, even Luddites in the past destroyed machines mainly for giving vent to their anger towards humans.

But why is it time for us now to ethically deliberate on the man-machine relationship? When can this

change be dated back to? What factors have made it necessary for us to think over the machine ethics?

Back in the early years of the industrial civilization age, machines were still products made and completely controlled by humans, so there was no any ethics-related controversy. The situation started to change perhaps when machines began to self-learn and self-improve, or more clearly, when automatic machines emerged. Then till the invention of intelligent robots, machines started to possess some abilities unique to humans, namely material control and instrumental rationality. Notably, it is the two abilities that humans have relied on to conquer other animals. When machines became more and more intelligent, they started to possess technical rationality or instrumental rationality, the two abilities which are exactly most highly praised and most widely popularized by modern people. In terms of this aspect, machines are somewhat humanoid, but by perception abilities, like feeling, will, and whole self-awareness, they are still subhuman. But now that they are partly humanoid, is it possible that machines could evolve into a quasi-human product that can feel and has its own will, even self-awareness one day? If this is the case in the future, we should start to consider the ethic relationship between humans and machines, though they may be only partly humanoid, shouldn't we? Of course, perhaps humans' fear about whether machines would continue to evolve into an intelligent winner overtaking even replacing humankind most strongly motivates humans to take this question seriously.

Because many science and technology experts may bend their mind to research and development, so we should thank writers and artists for their perseverance in raising every possible man-machine ethic problem in their works like many science fictions, movies and televisions.[①] Czech writer Karel Čapek introduced the word "robot" and raised the man-machine relationship problem in his 1921 science fiction play Rossum's Universal Robots, which makes the writer quite ahead of his time. This play lays bare a broad spectrum of robot creators and makers with disparate intentions like money, science, or even humane cause. Take Harry Domin — by creating robots, this group leader wants to liberate humans from heavy labor, making humans become dignified leisure "nobles." To realize the dream, his company churned out a large number of robots and forced them to replace humans all over the world to toil long hours, yet the president's daughter paid a visit and required the company's humane treatment of robots. A decade later, robots all over the world started to rebel against humans. They organized the International Robots Association, killed the manager of the factory, and replaced humans to become the world ruler. Unsurprisingly, they soon found themselves also in the grip of how to self-produce and self-duplicate.

American science fiction writer Isaac Asimov put forward the Three Laws of Robotics in his work Runaround for the first time: 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 the orders given 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 Laws. The three laws are arranged in priority order, i.e. the preceding laws enjoy higher priority, and the laws preceded cannot violate them. Specifically, a robot even shall not obey the orders given by a human being to injure another human being (for instance, orders given by a master to ask his/her robots to kill himself/herself); a robot must protect its own existence as long as such protection does not injure human beings and go against human beings' orders. If this is the case that a robot is going to

① The paper wants to highlight the role of literature and art, and hopes that politicians and research fellows can see more related literature works and films where all kinds of possibilities, especially potential risks, are approached in a more imaginative, open and far-sighted way.

injure a human being, or otherwise, a robot must kill itself as long as human beings give the order of suicide. Obviously, the three laws are anthropocentric.

If these laws are put into place, robots naturally have to take the trouble of judgment. Or, to put it more clearly, how do robots discriminate between a human being to be harmed and a human being giving the order? Does the human being here refer to individual humans or the humankind given possible contradictions between them? On top of that, they still have to judge in which cases human beings are subject to harm, and which human being or human beings they should choose to injure in case of inevitable harm; which order they should obey in case of more than one order given by different human beings; and so on. Asimov also depicted some of the aforementioned contradictions and dilemmas in his works. He not only pondered over the man-machine relationship — this relationship is doubtlessly unequal — but also attempted to home in on the explicit appearance of ethical laws to regulate the relationship, which gives a very meaningful start.

Later generations have made plenty of effort to modify and complement Asimov's Three Laws of Robotics, but most of their effort tends to result in more or stricter requirements, with Asimov's supplementary law providing an example. Perhaps out of concern that robots may be abused as villains' bodyguards, Asimov himself added a more preferential zero law: "A robot may not injure the humankind". But constant addition would make robots' judgment burden rise, and would also activate the loopholes of misjudgment or imposture and making use of "humankind interests."

Thinking about regulating man-machine relationship

The paper will put forward another idea which is both different from that of pioneer Asimov and those of today's machine ethics experts.

Briefly, current in-use robots can be divided into two categories: civil robots and national robots. At the moment, humans are directing their efforts in ethical regulation of the man-machine relationship perhaps towards: First, stipulating and guiding machines' values, by, for example, designing and cultivating intelligent machines' obeying of orders of "being kind to humankind," and teaching machines the moral values of putting human first; second, standardizing machines' behaviors and means, for example, one of Asimov's Three Laws of Robotics, "A robot may not injure a human being"; third, limiting machines' abilities, especially by preventing the development of generic super intelligence.

The paper holds that the first direction is neither necessary nor possible, and even highly risky. Besides, it also conflicts with the other two directions. In this context, the other two directions may be the focus of our effort, but there could be some differences between being applied to civil robots and being used to national robots, as shown below:

Some scholars suggest pre-designing motivation and values of "being kind to humankind" or "working to maximize humankind's interests" in machines. Yet humans have to, if they do so, have machines develop their own generic and comprehensive abilities, and even acquire a kind of self-awareness. If not, machines could not bear the heavy burden of judgment: for instance, what do humans mean by "being kind to humankind," or what are "maximum interests" or "whole interests" of humankind. That's because such judgment entails a kind of generic, comprehensive ability, and even a holistic approach not only to humankind's material interests, but to humankind's diverse spiritual, cultural and feeling needs. In this way, machines would have to possess the

self-awareness same with or similar to humans' so as to gain all-around insights into them. However, it seems impossible. Being neither carbon-based lives nor primates, machines have no body sensitivity, nor ability to comprehend spiritual culture exclusive to lives, for this kind of comprehension ability is more than inputting and memorizing all literature about humankind, but is based on personal experiences accumulated by myriad lives that can feel and be inspired in history. Furthermore, if machines could really make comprehensive judgment and take actions, they don't necessarily depend on the self-awareness same with that of humans who have body sensitivity and death foundation, and they may rely on a kind completely unknown to humans. In this way, humans could have no empathy with machines as they do with other humans.

But humans had better keep the value judgment ability exclusive to themselves as a way to maintain their independence from and even their dominance over machines, for humans cannot remain completely dependent on intelligent machines and outsource all things to them. If we had better not let happen the scenario that the majority depends on the minority in artificial intelligence field, we must not also let happen this scenario that humans have to depend on machines in artificial intelligence, spirit, and value judgment. Maybe humankind would have to meet their doom when they completely rely on machines in value judgment. Therefore, we had better keep machines being "objects" as they are. Perhaps what humans can do or only do is limiting their means and abilities, not establishing a set of anthropocentric value system for them. If a machine really develops the abilities of value judgment and self-building, the author is afraid that it could establish its own value system quite soon. That will be a value system unfathomable for humans or a "goal system" exclusive to itself — just like even experts fail to figure out the specific computing process about how robots win intelligence tests and defeat go masters. In fact, there still remains a lot of "black box operation," and it would be all the more so if the machine has a "heart" — the "black box" could be much larger and even could be a whole one.

Given this, the paper is considering another approach to Asimov's Three Laws of Robotics when it comes to, at least, civil robots, namely doing subtraction rather than addition. And by doing subtraction, the paper means to subtract Asimov's laws to the greatest extent, with only one law left: A robot may not injure the humankind, the first half of the First Law of Asimov's Three Laws of Robotics. To put it more clearly, this simple law means that a robot must not resort to violence towards humans. Notably, violence here includes not using compulsory means to limit humans' freedom, for example, compulsory imprisonment, compulsory detainment or locking up humans like the robot in *Ex Machina*. We can consider making "nonviolence" an unshakable principle and initial unchangeable bar password for all machines that all secondary application and manufacturing machines cannot change. In this way, humans actually may have to compromise on plenty of convenience and expectation brought by machines, for instance, humans can no longer make and use "robot bodyguards," because of a question: if good humans can use the violence of these machines, can bad humans abuse it in a more arbitrarily way?

Robots can also be used as a strong tool to save humans. When humans are subject to injury, robots have also many choices to save humans without using violence. Specifically, a robot can help humans escape and can also serve as a highly sensitive and responsive monitoring and alarm system as ways to make criminals unable to escape and pay the price. That's how a robot helps humans. In this way, we humans still have a powerful safety assistant, but we still cannot allow machines to use violence. Violence should continue to be controlled by humans, and it is also a responsibility humans should shoulder. In other words, a machine may

not involve in any violence, remaining isolated from and even completely uninformed about any violence. It could be an "animal" of whole peace.

When it comes to national robots, this law may not directly applicable, for nation defined by violence just cannot dispense with violence, which makes repulsion of violence in national robots impossible. Despite of the impossibility, we can still differentiate between the two kinds of application of national robots: domestic application and international application. The paper suggests that focus should be on banning any violent killer machines for domestic application, while on advancing specialization and miniaturization for international application. A few years ago, 56 countries allegedly already engaged themselves in developing killer robots, and some intelligent killer machines like unmanned aerial vehicles and killer bees have succeeded and come on stream. Against the backdrop, if all violence should be banned on civil robots, the fact is that the ban could be something akin to a physical impossibility on national robots, because nation is always violent in this way or otherwise. But the violence can be weakened to specialized and miniaturized weapons at least, instead going so far as to be intensified to weapons of mass destruction.

A quite great number of humans have been calling for a complete ban on the research and development of killer machines, but as long as a major country rejects, other countries seem not to give up. In this context, we can at least consider some prior bans and gradual restrictions. For instance, we can prohibit the development of killer machines towards weapons of mass destruction by only allowing the temporary existence of some existing specialized, well-targeted and miniaturized killer robots. Of course, the likes of demining and defusing robots are naturally allowed. Nations, especially major nations, can consider signing some multilateral protocols to ban, for example, nuclear proliferation and biochemical weapons. After all, there are some precedents for this practice, like the case of poison gas. This lethal weapon was created and used during World War I, but was banned later, even during the cruel World War II. Although it proves impossible to impose a complete ban on national violence for nation by nature means the monopoly of violence across a certain region, a responsible nation, especially a major responsible one, should prevent irresponsible nations or terroristic organizations from developing and abusing killer robots.

No doubt it is merely an idea, even a naive one. That's because it is human nature to have an infinite variety of motivations, as capitalists are eager for profits, blocs and nations for private interests, and scientists for knowledge out of curiosity. Each motivation may nullify restrictions on robots' abilities and means.

But tougher the picture gets, less complicated the regulation should be. With the philosophy in mind, the paper tries to put forward the aforementioned stipulation. Albeit it is very simple, it may be the scenario that to be simple is to be reliable. Stipulation must come simple and explicit, preferably as a ban — a ban is more feasible than a positive mandate — in an effort to reduce and even exempt robots' judgment burden, which may be followed by a limitation to their abilities. Controlling, implanting and cultivating value motivation is more difficult than controlling behavior. With regards to judgment, machines may be adept at quantifying profits and probability, yet parameters like humans' feelings and emotions are hard to quantify and compute. Therefore, it may be not advisable to give up or fail to give priority to this idea. Or rather, it is advisable to consider putting in place the law of "forbidding any machines from using violence" first at least for all civil robots. Furthermore, perhaps we should also make artificial intelligence "less clever," namely deliberately limit the development of intelligent machines to specialization and miniaturization as a way to restrict their self-awareness and comprehensive abilities, thus preventing their development towards super generic intelligence.

Interpersonal relationship as the key to addressing the man-machine relationship

The paper provides a review and some discussion of interpersonal and man-object relationships in previous parts, completely because of challenges posed by the man-machine relationship.

On comparison of the man-object and man-machine relationships, the paper finds some similarities between the two. Specifically, both objects and machines have no self-awareness, and both are grossly overmatched by humans, while there are also some differences between them. By taking the key one for example, intelligent machines combine the attributes of humans and objects: they without self-awareness are still artificial objects, and they with some attributes and abilities exclusive to humans can surpass humans especially in computing. In the future, they may overtake humans in all fields.

Domestication is how humans to achieve direct manipulation of animals. With the past several centuries of domestication, animals have undergone a change in their temperament, and humans can manipulate them by giving simple oral and gesture orders. Meanwhile this method also applies to beasts. Even if they are out of control, humans would not face a catastrophe, while in the case of machines which humans manipulate by programs and commands, humans may suffer a complete failure. As some scholars warn, humans have only one chance, for a slightest slip may make intelligent machines humankind's "last invention."

The ethical focus of the man-object relationship is on: How should humans be kind to animals and other objects as the dominator of this relationship? While that of the man-machine relationship is on: Although humans are the dominator of this relationship, the future may see a reversal of mutual positions. Then the paper, building on a hypothesis that how machines would treat humans, homes in on: What should humans do at the moment? What can humans do at the moment? But a great predicament arises therefrom: the hypothesis that humans may treat machines now in a way in which machines may treat humans in future involves a point which is completely unfathomable or even unpredictable for humans.

Although humans are working on how to treat and regulate intelligent machines, interpersonal relationship doubtlessly remains the paramount focus of our effort, for humans have to put forward all potential problems and their countermeasures, persuade and discuss with each other, create a social culture that pays attention to every aspects of artificial intelligence, and take the whole humankind's interests into full consideration.

Yet there could be a "critical minority" playing a significant role at a critical juncture. The critical minority includes: scientists and technicians who work on the front lines of AI research and development; entrepreneurs and capitalists who can always influence the research orientation as investors of AI development projects; officials and heads of governments who decide or manage AI policies and laws, and sometimes also make related decisions; intellectuals including writers and artists who keep inquiring into AI's nature and its potential effects and implications on humans. When it comes to AI, most humans should or can only share its development fruits, but cannot participate in its decision-making. The same is true of the potential "Sword of Damocles" (an allusion to the imminent and ever-present peril faced by those in positions of power). For instance, developing and using nuclear weapons were not voted through by the majority years ago.

Maybe humans can finally establish a set of safe and reliable value system for intelligent machines, but they should treat lightly before finding a real sound solution. Specifically, humans had better keep machines less clever, sophisticated and independent, and limit their abilities to simple computing or algorithm fields as

a tool and device. If machines have self-awareness and feeling, they may feel unequal and unfair, but this kind of inequality is after all inevitable in the interest of humankind's survival.

Humans should have learned to control themselves. If time could flow backwards, humankind should perhaps have given more attention to their spiritual culture with a focus on accelerating development of their ability of controlling themselves while slowing development of their ability of controlling objects. We are already amazed at the development power and speed of modern civilization. Back in the foraging culture age, humankind ever experienced a slow development pace which has objectively lengthened humankind history. Till the agricultural civilization age, although the traditional society then enjoyed a quite fast development pace, it mainly depended on cycles of time and space as a way to draw out humankind history. This time-space cycle theory is both theoretically and conceptually well-founded, and can be expressed in concrete terms as below: time wise, the cycle is reflected by successive establishment of dynasties on the same region, and space wise, the cycle is embodied by constant appearance of civilization empires on different regions. But in the industrial civilization age defined by the theory of evolution, the egalitarianism and the globalization, there was no longer the objective control of development speed and power. Therefore, limiting the development of intelligent machines to specialization and miniaturization which is especially within the scope of non-violence as far as possible may be the currently best way forward for humankind.

I, Robot, a 2004 American science fiction action film, is a really apt wake-up call for the potential generalization and violence-oriented tendency of intelligent machines. In the film, the latest-generation robots, following their acquisition of super abilities and then self-awareness, begin to interpret Asimov's Three Laws of Robotics from their own part. Thinking that they make a better judgment about humans' interests than humans themselves, they give the order to ask new-generation machines to kill their predecessors, imprison humans forcibly and kill rebels. A police head in the film sighs with heavy irony, "Well, then I guess we're gonna miss the good old days when people were killed by other people."

REFERENCES

James Barrat: *Our Final Invention: Artificial Intelligence and the End of the Human Era*. Beijing: Publishing House of Electronics Industry, 2016.

Nick Bostrom: *Superintelligence: Paths, Dangers, Strategies*. Beijing: CITIC Press Corporation, 2015.

(Translator: Zhuang Qiuyue; Editor: Yan Yuting)

This paper has been translated and reprinted with the permission of *Exploration and Free Views*, No. 7, 2018.