On the three constraints of the development of artificial intelligence: Value, liberation and responsibility

Hongxia Hou
Shanxi University, China

Abstract
The socialization of technology and the technicalization of society have accelerated the speed, scope and scale of human liberation from nature. In our past technological society, ‘liberation from nature’ mainly referred to people’s physical strength. An intelligent society, characterized by the wide application of the internet, big data and artificial intelligence (AI), will further liberate not only people’s physical strength, but also their mental power, greatly changing social forms and social operation modes, as well as people themselves. In the application of technologies and the design and manufacture of equipment on which the intelligent society depends, a new tension is formed between instrumental rationality and value rationality. While promoting the liberation of human beings, intelligent society will also cause humans to be dominated and enslaved by AI, which will lead to human alienation. Intelligent society is more in pursuit of how the invested capital and technology can multiply in the process of accumulation and circulation; meanwhile, it will ignore the moral responsibilities of relevant parties, such as researchers, manufacturers and users. In the process of developing an intelligent society, instrumental rationality should be regulated by value rationality, thus promoting the liberation of human beings and eliminating their alienation. As capital appreciation occurs, the moral responsibilities of relevant parties should be clarified.

Keywords
Intelligent community, value rationality, human liberation, moral responsibility

1. Introduction
In 1950, Alan Turing first asked the question ‘Can machines think?’ Since then, the application of artificial intelligence (AI) information technology in socio-economic development has received growing attention from governments around the world and generated widespread interest in academic communities. During the 1980s, researchers began to study the ethical issues of AI development. With a focus on interactions between humans and machines, those...
studies have clarified the relationship between humans and technologies, which includes three dominant views:

1. The development of AI will lead to the destruction of human beings, and intelligent machines will replace, control, and even destroy human beings in the future.
2. Human beings have intelligent brains and sharp minds; AI cannot exceed human intelligence; and AI development will make human life better.
3. Human–machine integration is the future of AI development.

From the perspective of system theory, the gap between AI and human intelligence is insurmountable because human intelligence is a product ‘representing the entire human system’ (Chen and Gao, 2019), not just a product of the human brain. Moreover, AI machines are programmed in advance by human intelligence, so they are not capable of doing anything beyond human intelligence. Thus, it seems that AI machines are never going to surpass human intelligence. Of course, it would also be wrong to deny the continuous development of AI based on this assumption.

Although scholars hold different views about the trend of AI development and the relationship between AI and human society, viewed from a longer perspective, AI machines will continue to progress. AI not only connects individual intelligent machines, but also links humans with technology. It involves the relationships between humans and technologies, between humans and machines, between instrumental rationality and value rationality, and between human emancipation and human alienation. If we blindly pursue the instrumentality and functionality of intelligent machines, people will become obsessed with machines and think like machines, and a whole range of problems will arise. In the research, development, design and manufacturing process of AI, we should not lose sight of the tension between instrumental rationality and value rationality, the human alienation that AI causes while promoting human emancipation, and the moral responsibility attribution and distribution dilemma caused by the single-minded pursuit of capital appreciation. We should also balance and coordinate ‘dataism’ with humanism to limit and regulate the development of AI and ensure that it retains the fundamental purpose of serving human society.

### 2. Tension between instrumental rationality and value rationality

Max Weber was the first thinker to propose the theory of rationality and address the relationship between instrumental and value rationalities. For him, value rationality is ‘determined by a conscious belief in the value for its own sake of some ethical, aesthetic, religious, or other form of behaviour, independently of its prospects of success’ (Weber, 1998: 56). In other words, value rationality is concerned with whether a person’s behaviour is consistent with a particular value concept rather than the consequences of one’s chosen actions. If we think instrumental rationality answers the question of ‘how to do’ and takes maximized benefits as the basis for successful behaviour, value rationality answers the question of ‘why to do’ and takes conformity to moral codes as the standard. In making choices of social behaviour, people should seek a combination of instrumental rationality and value rationality, as that is the only way to achieve real success.

With the further development of intelligent society, AI technology ‘studies the science and technology that aim to make machines more intelligent, especially the science or technology that seeks to realize or reproduce human intelligence on computers’ (Warwick, 1998: 1–2). First, AI as a technology relies on computers to achieve intelligence, which highlights its utility and functionality to human beings as an instrumental rationality. Second, the fundamental purpose of the development of AI technology is to create a better life for humankind, which inevitably needs to follow the basic requirements of value rationality. Finally, in the process of AI development, instrumental rationality and value rationality should play their own roles respectively. Instrumental rationality is at the service of value rationality, while value rationality is more fundamental. However, in the development of AI technology,
the tension between instrumental rationality and value rationality has often put the two in a state of paradox.

In the age of intelligence, the powerful instrumental rationality of intelligent machines has overwhelmed the value rationality that humans ought to have, such as compassion, reverence and many other human characteristics. As a result, the instrumentalization of humans has become a prevailing trend. AI technology, as a prominent representative of the fourth technological revolution, is applied in all aspects of human life through big data and intelligent analysis systems. With precision and quantification, it is fully capable of generating maximized benefits. In Marcuse’s view (2008: 71), ‘The legitimacy of technology has been translated into the legitimacy of society. The spread of technology worship is shaping such a transition: technological liberation—instrumentalization of things and objects—shackles of liberation—instrumentalization of man.’ Like previous technologies, AI has not escaped the dilemma of human dependence on technology, but its importance goes further. As purpose-oriented instrumental rationality expands, people lose through the curse of technology, and the instrumentalization of man follows. As described in A Brief History of Tomorrow, ‘even if these useless masses do nothing, it is likely that the society as a whole will be able to support and keep them alive. However, what could they use to kill time and gain satisfaction? … perhaps drugs and computer games is an answer’ (Harari, 2017: 293).

With the upgrade of AI technology systems and services, the productivity created by AI technology through robotics, voice recognition, image recognition, natural language processing and expert systems increases day by day. As the function of instrumental rationality becomes more powerful, the subjective needs of people are ignored and the function of value rationality is weakened. In the meantime, under the conditions of a market economy, people’s material desires grow stronger. Instrumental rationality expands infinitely and fills people’s minds, and factors such as utility, interest, and purpose always come first. In the pursuit of material desires and interests, people’s subjective needs disappear into the background, and value rationality dissolves.

In the course of AI development, the expanding instrumental rationality has obscured the value rationality of human behaviour and made it less relevant. In Weber’s view, with the evolution of society, human beings are destined to live in the prison of technological knowledge in the foreseeable future (Beck et al., 2001: 74). The immediate result of the continued development of instrumental rationality is the great success and enormous wealth creation of individual persons, especially in ‘how to do’ questions, such as the use of AI weapons and research on autonomous cars. However, when it comes to ‘why to do’ questions that have value implications, humanity has made little progress and even become more confused (Osiatynski, 1988: 192).

In the process of AI development, the deviation between instrumental rationality and value rationality is highlighted by the lack of a humanistic spirit in science and the neglect of ‘man as an end in itself’. For one thing, AI research without the humanistic spirit and the constraint of value norms will fall victim to plagiarism, idea theft and other forms of academic irregularities, and society as a whole will be devastated by the power of technology unleashed by human selfishness and greed. For another, the development of AI technology raises concerns about the risk of future society and leads people to reflect on the question of ‘man as an end in itself’.

At present, weak AI is everywhere in our society, and the era of strong AI has not yet arrived. In essence, the challenges created by weak AI are still in the category of ‘tools’ that have no human consciousness. For example, tasks such as facial recognition, autonomous driving, voice translation and intelligent interviewing are all realized through programming. In other words, until now, AI has mostly served the subjective needs of people and observed the moral code that sees ‘man as an end in itself, not a means to an end’ (Kant, 2003: 95). However, due to the utility and utilitarian nature of instrumental rationality, people think more about short-term benefits and immediate interests when they use AI technology. Therefore, from the very beginning, the application of AI technology has limited human rationality to solving only technical problems. The use of AI machines draws a beautiful picture of a leisurely life, but it also makes people lose the ability to think independently and increases
human dependence on intelligent technology. Instead of thinking about how to live a better and more meaningful life, people allow themselves to be dominated by instrumental rationality and take practicality and utility as the only criteria for making choices. With the purpose of humans erased, their value also becomes irrelevant.

In the future, strong AI is essentially a ‘super AI’ that ‘far exceeds the cognitive ability of humans in almost all fields’ (Bostrom, 2015: 29). If human beings continue to be influenced by the rationality of powerful AI technology and, in certain extreme cases, if some individuals ‘lose the values, compassion, and sense of reverence that define humanity’ (Liu et al., 2019), will human beings start to think like machines, and will super AI imitate greed, selfishness, desire for control and other human emotions, thus causing devastating consequences? In the era of AI, the value criteria set by the designers for intelligent machines are naturally influenced by the social systems and values of the times. Therefore, instrumental rationality and value rationality should be complementary in the course of AI development.

The ultimate purpose of AI development is to benefit humanity and serve society. The more AI resembles humanity and the closer it is to human life, the more vulnerable we humans may become. Humans are strong but also weak. We are strong because we can surmount any obstacle that stands in our way. We are weak because we cannot afford even the slightest damage caused by the intelligent machines of our own creation.

In 1989, Nikolai Gudkov, the Soviet chess champion, played against a robot chess player. After losing three straight games to Gudkov, the robot became furious. It attacked the human champion and knocked him down. The tragedy was caused not because the technology was running unchecked—as that would be an oversimplistic conclusion—but because the development of the technology was flawed.

The American scientist and popular science writer Isaac Asimov proposed the ‘Three Laws of Robotics’. The first law stipulates that intelligent machines may have a place in our society and provide conveniences for our lives, yet they cannot injure but must protect human beings. The second law builds on the first and stipulates that robots must obey orders given by human beings. The third law stipulates that robots may protect their own existence, but such protection shall not conflict with the first two laws. Therefore, even if intelligent machines in the future do become self-conscious, they still need to follow basic human values and norms, such as compassion and non-damage. When making intelligent machines, the developers must set value codes for them so that they can do no harm to humans either physically or mentally.

3. Promoting human emancipation while dissolving the state of human alienation

Human emancipation refers to the liberation of humans from the state of oppression and enslavement. For Karl Marx, the purpose of the emancipation of man is to liberate him from the state of alienation from things (nature), from other men and from himself (Zhang, 2019). In the future, intelligent machines will be able to expand the raw materials and enrich the production instruments while doing enslaved work for humans, allowing workers to move to new locations, choose new jobs and learn new types of work at any time, while receiving all-around development. This not only means that people will have more time to freely do their work, but also provides the prerequisite for their free and all-around development. At the same time, the basic content of man’s free and all-around development is the rationalization of all aspects of man’s social relations, which depends, to a great extent, on the abundance of free labour time that man enjoys (Guo and Tang, 2010).

Ever since humankind entered the 21st century, three interconnected yet distinctive new eras have emerged one after another: the internet era, the big-data era and the AI era (He, 2016). The ultimate goal of the evolution of human society is the complete emancipation of human beings, and the fourth technological revolution—led by new technological advances, such as the internet, big data, and AI—provides a realistic possibility for the realization of that goal. With the rapid development of AI technology, humans are increasingly inclined to improve their lives and dominate the natural world by
Harnessing technologies, which has further activated the role of AI technology in promoting the freedom and emancipation of humans. The development of AI technology liberates humans from high-intensity, high-workload, high-risk and high-pressure activities. It also notably prolongs the free labour time of human beings. AI-empowered machines can endure dirty working environments, tedious and repetitive tasks and non-stop working hours. They have the fine qualities of dedication, diligence and selflessness, and they do not need compensation for the value of their labour. These are all advantages to which human labour cannot compare. With its rapid technological advances, AI has not only replaced some of man’s physical work, but also undertaken certain mental labour with regular patterns. All this will increase people’s free labour time, thus creating favourable conditions for the emancipation and all-round development of humans.

While the development of AI technology helps humans to become emancipated and escape slavery, it also traps them in a state of alienation, creating the ethical problems of human emancipation and alienation.

First, the development of AI technology increases the risk of structural unemployment and intensifies the alienation of people in productive activities. The use of AI-powered machines liberates humans from tedious and dangerous manual labour. As machines become more human-like, they will take over many of our jobs. Such competition is not just about AI replacing humans, but also exists among humans themselves. More jobs will favour workers who understand AI technology, and thus structural unemployment will become inevitable.

Over the past three decades, the internet has disrupted the workplace, and AI is now poised to bring about a new transformation. Technology giants, such as Amazon, have been global leaders in the AI field. Now, companies of all sizes are using AI technology for job interviews. For example, L’Oreal has introduced two robots as interviewers. They can not only moderate question-and-answer sessions and review job applicants’ qualifications, such as academic performance, work experience and international visas, but also answer their questions about corporate culture and employee welfare. They are also learning to become less dependent on humans, which could notably save time in human resources departments. However, while enjoying the convenience, we cannot help but worry about some problems: Is the robot’s assessment of job applicants completely reliable? Will such a selection mechanism, which lacks sufficient human input, miss talented people who fail to pass the robots’ tests? Will robots replace humans, and will our interviews depend entirely on the ‘preferences’ of machines? Those concerns are not totally unjustified, but a price to pay for the invention and evolution of all new things. AI will perform much of the work that is currently undertaken by or involves humans and will do so at a speed unimaginable to humanity.

When the ‘singularity’ of AI surpassing human intelligence arrives, human work will face two possible outcomes. The positive outcome is that human labour becomes a free and conscious activity, and a need for survival and development. However, a worse outcome is also possible. As the development of AI continues, it will escalate conflicts in society and affect society as a whole, which may trigger a greater, cruel, social transformation. That will usher in a whole new era of intelligence in which pressure comes not just from unemployment, but also from the great psychological challenge of moving from work to labour. Essentially, the rapid development of AI will reduce the amount of human labour and stimulate massive structural unemployment, thus creating a ‘non-labour existence’, dissolving people’s subjective need and turning to alienation.

Second, the development of AI has led to the degradation of people’s spiritual world in the consumer era and intensified the alienation of humans in life in general. When humans are overdependent on intelligent machines, we become less willing to think, and our spiritual world deteriorates as a result. While the widespread application of smartphones lays the foundation for humans to learn and master various types of knowledge and skills, it ‘paradoxically’ turns some people into one-dimensional persons who cannot escape the addiction of smartphones’ (Jiang and Tan, 2019). As Ruan (2019) points out, ‘With the rapid development and application of AI, many people no longer obtain knowledge and ideas through personal experiences such as logical thinking and self-expression, but seek quick access to
their desired results through the ubiquitous presence of AI. For ‘one-dimensional addicts of smartphones’, AI machines will feel like a part of our bodies, and be like the air and water that are necessary for survival. Without them, we may suffocate and inevitably generate a sense of emptiness and boredom. The problem would be particularly serious for young people, as they may become world-weary, self-absorbed and reluctant to communicate with others when they are obsessed with AI machines.

Currently, the development of AI is dependent on big-data resources. With the widespread use of various types of data-collection devices, the big-data resources that are collected include traditional identity information, contact information, home addresses, biological characteristics, physiological features and other related information. Furthermore, a new generation of Asimov machines and mind-reading software can already determine the content of human thoughts. Although they are still in their infancy, and are currently only able to identify simple human thinking activities, the detection of human thinking will only become easier with the advance of AI technology. Ideas that are buried deep in human brains will have nowhere to hide when they face AI machines. This will create a complex and fluid world unlike any in the traditional era.

Finally, AI as a technology cannot change the alienation dimension. The development of AI technology results in the loss of human subjectivity and isolation between humans, thus trapping humans in the state of alienation. This is known as the alienation dimension of technology (Wu, 2019). As Marx put it:

In our time, everything contains its own opposite. We see that machines have the miraculous power to reduce human labour and make it more productive, yet they cause hunger and over-fatigue. The triumph of technology seems to have come at the cost of moral corruption. (Marx and Engels, 1995: 775)

With the development of the internet, the internet of things and cloud platforms, less communication and fewer exchanges take place between people, and the resulting isolation in turn leads to a deeper alienation of interpersonal interactions. In particular, while intelligent machines are growing smarter, with faster internet connections and an increasingly massive data flow, the human brain, which we have always taken great pride in, is becoming clumsy and somewhat overwhelmed. Through new forms of alienation phenomena, such as substitution, control, enslavement and exclusion, the intelligence-powered fourth technological revolution has belittled the free activities of humans as a means to happiness, and it is only inevitable that ‘the vast majority of people will be left behind and reduced to negligible dependents or components of a huge, complex intelligent machine system, or even slaves of intelligent machines’ (Sun, 2019). In the future, AI will dominate our work and think for us, and humans will not need to work anymore. People who lose their work will also lose the value and the meaning of life.

Over 60 years of AI development, from Samuel to Deep Blue to AlphaGo, AI systems have been upgraded and widely applied in expert systems, pattern recognition, problem solving, machine translation, machine learning and many other fields. Marcuse, a leading figure of the Frankfurt School, pointed out that the developed industrial society is fulfilling only people’s false needs, and their real needs have been completely cast aside. This is also true for the development of AI: the risks it brings to human beings should not be ignored.

First, structural unemployment in low-creativity, repetitive, and low-skilled industries caused by the substitution of AI for humans in much mental and physical work is just the first phenomenon. Being dominated and enslaved by the products of humans’ own creation will further reduce the need for work and exacerbate the trend of alienation. As people become increasingly dependent on technological products, they will no longer be actively engaged in creative work, but passively enjoy the mental satisfaction and pleasure fostered by certain functions of technological products. Humans will need AI to work and think for them, and there will simply be no need to work anymore. Yet people who lose their work will also lose their value and meaning of life and end up being the slaves of intelligent machines and AI’s subordinates.

Second, while leading to the alienation of human labour, intelligent machines also cause the mechanization of human thinking. What we should really fear is that people are becoming addicted to machines and
beginning to think like machines. With the help of AI technology, cellphones, the internet and intelligent tutoring machines have made our lives much more convenient and notably increased the free time we enjoy. Meanwhile, by making people obsessed and addicted, AI has rapidly taken up almost all free time outside the necessary labour time of human society, yet the all-round development of humans has been constrained to a limited range and has stopped progressing. People only need to focus on their own interactions with machines and compete for the most advanced intelligent machines. Because of their dependence on and obsession with intelligent products, human emancipation has been trapped in a state of alienation. Therefore, instead of leading to the free and all-round development of humankind, AI has caused a deeper human alienation and even the inversion of the subjectivity of humans and machines. In this sense, in the evolution of intelligent society, how to promote human emancipation while dissolving the alienated state of humans is an issue that deserves greater attention.

4. Capital appreciation with clear moral responsibility

Capital is required in the development of AI, either for the training of professionals or for the research and production of the hardware and software systems of high-tech products. Therefore, the sustainable development of AI requires substantial investments of capital. At the same time, the technological progress and innovative outcomes of AI can be realized only when driven by the appreciation of capital. The intrinsic nature of capital appreciation provides a strong impetus for the development of AI technology and makes it a core competitive force in industrial production. Driven by the logic of capital, the rapid development of AI and big data will take humans from the economic status of industrialized machine production to the status of a digital, information and knowledge economy. The efficiency of global production and labour productivity will grow at an epic speed, and the whole social system will become more automated, networked, intelligent and digitized. However, the logic of capital appreciation has been described as ‘a logic of blind development and tenacious accumulation’ (Beaud, 1986: 145). Driven by the logic of capital, the development of AI focuses mainly on the appreciation of invested capital in the process of accumulation and circulation, while ignoring the issue of moral responsibility that should be borne by the relevant stakeholders, such as developers, producers, and users.

In June 2019, the National Governance Committee for the New Generation of Artificial Intelligence released the Governance Principles for the New Generation Artificial Intelligence: Developing Responsible Artificial Intelligence, which puts forward the principle of shared responsibility and clearly states that:

AI developers, users and other stakeholders should have a high sense of social responsibility and self-discipline, and should strictly abide by laws, regulations, ethical principles, technical standards and social norms. AI accountability mechanisms should be established to clarify the responsibilities of researchers, developers, users and relevant parties. (China Daily, 2019)

In the development of AI, if we follow only the logic of capital and fail to develop responsible AI or adopt the attribution of prospective moral responsibility, we will fall into a responsibility dilemma in ethical practice. Specifically, in the analysis of responsibility for the negative consequences caused by intelligent machines, how is responsibility attributed and divided? Should intelligent machines be morally responsible? These are questions that cannot be avoided in the development of AI technology.

To answer the question of whether intelligent machines should bear moral responsibility, the key is to determine whether intelligent robots could enjoy great autonomy. In the future, as intelligent machines become fully autonomous, we should let them take moral responsibility and treat them as legal actors who are subject to punishment (Lin et al., 2008). As of now, most intelligent robots are not yet fully autonomous (Bekey, 2005: 1–2), and their behaviour is based on the software programs pre-set by their developers. Because of this, although intelligent robots can be exempted from moral responsibility, ‘robot designers, manufacturers and their organizations shall be held accountable for the
safety and reliability of the robots and shall not be exempted from their responsibilities’ (Du, 2017).

However, an issue that is more critical to the negative consequences caused by intelligent machines is how responsibility is attributed and divided among the parties involved. The concept of ‘organized irresponsibility’, introduced by Ulrich Beck in his book *Gegengifte*, refers to a coalition of corporations, policymakers and experts who create dangers in contemporary society and then build a discourse to ‘pass the buck’, but ultimately transform the dangers they have raised into some kind of risk (Xu and Zhou, 2016). Compared with the traditional means of punishing the person in charge, acts of infringement in the intelligent era are conducted via algorithms and electronic ports, which provide the technical support for people seeking maximized profits to disguise their acts of infringement and escape punishment. Intelligent machines cannot beat human intelligence, and the human is still the decisive factor in the act of infringement. In the future, rather than humans and machines fighting with each other in two camps, a more likely scenario is humans fighting against each other with their machines. In other words, the ultimate problem is still a problem about humans, not machines, and the confrontation and conflict between the parties involved are still essentially a conflict of economic interests. Therefore, the moral responsibilities of various stakeholders, in the development of AI, such as developers, manufacturers and users, should be clarified.

According to when intelligent machines are produced, we can divide the moral responsibility for the resulting negative consequences into prospective and retrospective responsibilities. During the development of AI, the developers shall be held accountable for the safety and reliability of intelligent machines and take on prospective moral responsibility. In other words, they must have a comprehensive understanding of the possible consequences of developing intelligent machines. Patrick Lin even proposed that designers should design robots with compassion and selflessness (Yan, 2019). If the developers voluntarily take on their prospective moral responsibility, that will help to significantly reduce the negative effects caused by intelligent products and will thus improve public acceptance of their products. For manufacturers, to make safe and reliable intelligent products is their most important moral responsibility. At the same time, they should also take prospective moral responsibility and test their products rigorously before launching them to avoid adverse consequences caused by the use of intelligent machines. Moreover, users should respect intelligent machines and treat them properly; that is, users should not abuse or misuse them. Therefore, users should bear retrospective moral responsibility for the consequences caused by the improper use of intelligent machines. In general, the developers, manufacturers, users and other stakeholders should bear moral responsibility for the negative consequences caused by intelligent products, but how such moral responsibility should be divided is still a difficult problem that can be solved only in future practice.

5. Conclusion

Ultimately, the development of AI must serve human society. We must use advanced AI technologies that conform to human values and norms to promote social development rather than hinder its progress. Therefore, we must ensure that new AI technologies contribute to the progress of human beings from the very beginning of the research. Like other technologies, AI has both advanced and backward aspects, and we must ensure that the development of AI technology is in line with the direction of social progress. Some experts point out that the danger of the abuse of AI technology is comparable to that of nuclear weapons. AI is not human intelligence in any real sense. We should remind ourselves that, however AI is designed, managed or controlled, issues such as the development of human society, the subjectivity of humans, and ethics and morality should not be left completely to intelligent machines.

From the perspective of Marx’s theory of practice, whether AI development will head in the direction of survival, destruction or integration is ultimately determined by practice, because ‘whether human thinking holds objective truth is not a question of theory, but a question of practice’ (Marx and Engels, 2009: 500). However, the crux of the matter is that the development of AI is a double-edged sword. While promoting human emancipation, it also traps humans in a state of alienation. It
facilitates the spread of the technology cult but puts humans in what Weber calls an ‘iron cage’, thus inverting the question of ‘man as an end in itself, not as a means to an end’. While contributing to the appreciation of capital, it also raises the issues of the attribution and division of moral responsibility. Therefore, the basic conclusions of this paper are as follows.

First, to make the most of the extended free labour time enabled by AI, we need to enrich our spiritual world, build our mental strength and satisfy our cultural needs through diversified means. In doing so, humans will not unconsciously become slaves of intelligent machines or even one-dimensional people, thus gradually saving humankind from the dilemma of alienation.

Second, in the process of AI development, we should move beyond instrumental rationality and the prisoner’s dilemma. We should underscore the value rationality of humans, cultivate the scientific and humanistic spirit of developers and encourage them to prioritize human subjectivity in AI research to lay the foundation for persevering with basic human values, such as compassion and reverence.

Third, the government and related organizations should actively respond to the adverse consequences created by AI machines. At the same time, they should deal with the attribution and division of moral responsibility among developers, manufacturers, users and other stakeholders and build an effective system of accountability based on the principle of shared responsibility.

In a nutshell, the vision of AI development is to continuously improve human intelligence through human–machine integration. AI development should focus on the issue of human subjectivity to provide ideas and methods for solving the above problems. These problems in the development of intelligent society cannot be solved just by enriching the spiritual world of humans or by relying on institutions and mechanisms. What we need to do is to work on both fronts.

**Declaration of conflicting interests**

The author declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

**Funding**

The author received no financial support for the research, authorship and/or publication of this article.

**References**

Beaud M (1986) *A History of Capitalism 1500–1980*. Translated by Wu AM and Yang HM. Beijing: The Eastern Publishing Co. Ltd (in Chinese).

Beck U, Giddens A and Lash S (2001) * Reflexive Modernization*. Translated by Zhao WS. Beijing: The Commercial Press (in Chinese).

Bekey G (2005) *Autonomous Robots: From Biological Inspiration to Implementation and Control*. Cambridge, MA: MIT Press.

Bostrom N (2015) *Superintelligence: Paths, Dangers, Strategies*. Translated by Zhang TW and Zhang YQ. Beijing: China CITIC Press (in Chinese).

China Daily (2019) Governance principles for the new generation artificial intelligence: Developing responsible artificial intelligence. *China Daily*, 17 June.

Du YY (2017) On the moral responsibility in robot ethics. *Studies in Science of Science* 35(11): 1608–1613 (in Chinese).

Guo Y and Tang WP (2010) The practical meaning of human development and human emancipation under Marxism. *Reformation & Strategy* 26(6): 10–13 (in Chinese).

Harari YN (2017) *Homo Deus: A Brief History of Tomorrow*. Translated by Lin JH. Beijing: China CITIC Press (in Chinese).

He Z (2016) Leading to the era of artificial intelligence: A discussion on the strategic direction of American artificial intelligence and its reference to Chinese artificial intelligence strategy. *E-Government* 12(12): 2–10 (in Chinese).

Jiang HQ and Tan PW (2019) Reflections on the ontological reformation of artificial intelligence. *Academic Forum* 42(3): 98–105 (in Chinese).

Kant I (2003) *Kritik der Praktischen Vernunft*. Translated by Han SF. Beijing: The Commercial Press (in Chinese).

Lin P, Bekey G and Abney K (2008) Autonomous military robotics: Risk, ethics and design. Report for the US Department of Navy, Office of Naval Research. Available at: http://ethics.calpoly.edu/ONR_report.pdf (accessed 20 July 2021).
Liu D, Liu J and Xu DB (2019) Risk prediction and anxiety thinking: Artificial intelligence’s shock and change to education. Higher Education Exploration 7: 18–23 (in Chinese).

Marcuse H (2008) One-Dimensional Man: Studies in the Ideology of Advanced Industrial Society. Translated by Liu J. Shanghai: Shanghai Translation Publishing House (in Chinese).

Marx K and Engels F (1995) Anthologies of Marx and Engels (vol. 1). Beijing: People’s Publishing House (in Chinese).

Marx K and Engels F (2009) Works of Marx and Engels (vol. 1). Beijing: People’s Publishing House (in Chinese).

Osiatynski W (1988) Contrasts: Soviet and American Thinkers Discuss the Future. Translated by Xu YZ. Shanghai: Shanghai Translation Publishing House (in Chinese).

Ruan CH (2019) Damage of AI alienation, pseudo-knowledge overflow and nationwide entertainment to human nature and civilization. Science and Technology Management Research 8: 262–267 (in Chinese).

Sun WP (2019) Philosophical thinking on artificial intelligence and the destiny of mankind. Jiang Hai Academic Journal 4: 135 (in Chinese).

Warwick K (1998) March of the Machines. Translated by Li B, Fu TY, Xiang L, et al. Inner Mongolia: Inner Mongolia People’s Publishing House (in Chinese).

Weber M (1998) Economy and Society (vol. 1). Translated by Lin RY. Beijing: The Commercial Press (in Chinese).

Wu SF (2019) Critiques of artificial intelligence technology on the horizon of Marxist consciousness theory. Philosophical Trends 4: 26–33 (in Chinese).

Xu LC and Zhou L (2016) ‘Organized irresponsibility’ under food safety threats: Analysis of consumers’ behaviours and ‘one family, two systems’. Journal of China Agricultural University (Social Sciences) 33(6): 124–135 (in Chinese).

Yan KR (2019) Robot ethic: The ethics of machines or human beings? Journal of Northeastern University (Social Science) 21(7): 331–336, 343 (in Chinese).

Zhang XG (2019) Study of Marx’s thought on the ‘one discipline’. Jianghan Tribune 8: 51–58 (in Chinese).

Author Biography

Hongxia Hou holds a PhD degree from Fudan University and is currently an associate professor at the School for Marxism Studies of Shanxi University. Her research interests are ideological and political education, moral theory, and moral education.