Journal of Philosophy of Life Vol.13, No.1 (January 2023):44-63

Implications of Automating Science
The Possibility of Artificial Creativity and the Future of Science

Makoto Kureha*

Introduction

Science fiction writer Chiang's (2002) short story titled "The evolution of human science" depicts a future age in which "metahumanity" acquires intelligence that outstrips human intelligence. There all original works of scientific research are generated by metahumans. Most human researchers have quit their jobs, and the rest of them are engaged in "hermeneutics," or the study that aims to interpret metahuman science and make it comprehensible to humans. The narrator of the story considers the *raison d'être* for humanity's science in this age. Although Chiang does not explain in detail what "metahumanity" is, one of its candidates is artificial intelligence (AI)—more specifically, what Searle (1980) calls "strong AI," which has the real capacity for thinking.

In reality, strong AI has not appeared yet, but AI technologies are increasingly applied to scientific research for various purposes. In 2017, *Science* issued a special issue titled "AI Transforms Science," which reported the current status and future prospects of the use of AI in various scientific areas of physical, biological, and social sciences. Moreover, some researchers aim to realize "the automation of science" (King et al. 2004), that is, to make AI systems or robots execute research tasks without human intervention. It is impossible to predict at present whether the future Chiang depicts, where AI takes humanity's place in science, will come true. At any rate, as automation proceeds, modes of scientific research as well as the state of the science community and the relationship between science and society will change dramatically.

This tendency of automating science is remarkable in multiple respects. First,

---

* Associate Professor, Faculty of Global and Science Studies, Yamaguchi University. 1677-1, Yoshida, Yamaguchi City, Yamaguchi 753-8541 Japan.
Email: kureha[a]yamaguchi-u.ac.jp

** This paper is an English translation (with some modification) of my writing part of the following co-authored paper: Kureha, M. & Kukita, M. 2020. 'AI and Scientific Research', in S. Inaba, M. Oya, M. Kukita, S. Narihara, M. Fukuda, & T. Watanabe (eds.), *Artificial Intelligence, Humanity and Society*, pp. 122-169, Tokyo: Keiso Shobo (written in Japanese).

it will radically change the overall state of science and technology, and so it may lead to large-scale innovation and to solving social problems. Second, since science is one of the most distinctive human activities, the idea of automating it prompts us to reconsider the aspects of our humanity itself. One of these aspects is *human creativity*. Renowned physicist Dyson (1988) states, "Science is at its most creative when it can see a world in a grain of sand and a heaven in a wild flower. Heavy hardware and big machines are also a part of the science, but not the most important part" (p. 158). As clearly shown in this passage, the view that human creativity constitutes a central value of science is widespread. However, as I will argue, the automation of science might undermine this creative character.

In this article, I examine the prospect of automating science, focusing on two questions. First, can AI make creative discoveries? Second, what implications may the automation of science have on science and society? It will be shown that the attempt to address these questions leads to a reconsideration of philosophical questions concerning the nature and values of science. I will conclude that the prospect of success in automating creative discovery is not bright at present. Nevertheless, I will also argue, we should anticipate that the automation of science will have many serious implications for science and society. Therefore, we need to specify desirable ways of introducing AI technologies into science and devise measures against the demerits of automating science.

In Section 1, I will introduce the current state and future goals of the automation of science. In Section 2, I will examine whether AI can make creative discoveries. Then, in Section 3, I will consider what implications the automation of science has for the science community and wider society.

1. What Is the Automation of Science?

1.1 The current state of applications of AI in science

Scientific research has been a central target of AI research since its early days. Many studies conducted on this topic constitute a research area known as "machine discovery" or "computational discovery." Two famous examples of their achievements are AM (Lenat 1977) and BACON (Langley et al. 1987). The former is a program that generate programs expressing mathematical concepts; for example, it succeeds in producing Goldbach's Conjecture. The latter is a program that picks up invariants from data; for example, it succeeds in deriving

Kepler's First Law from data concerning planetary orbits. Thus, early studies on machine discovery had aimed at simulating past discoveries made by human scientists. There also are attempts to make computers aid human researchers to discover new knowledge (see Langley 2000), but this approach had not become major until recently.

Now, AI systems are increasingly applied to scientific research more practically. AI systems based on machine learning methods are used for various scientific purposes such as the detection of new particles in physics, classification of celestial bodies from image data in astronomy, prediction of efficient ways of chemical synthesis in chemistry, identification of the genome for a psychiatric disorder in biology, and analysis of the mood of masses from social medias in psychology (see Science News Staff 2017). Moreover, there are attempts to make AI systems survey literature by applying text mining methods (see Stix 2005) as well as attempts to make AI-based robots run experiments (see King et al. 2004; King 2010). Thus, AI is becoming a standard tool in diverse areas of physical, biological, and social sciences. (Although AI is becoming also widely used in applied science areas such as medicine and pharmacy, I focus on its use in basic science areas in this article.)

This trend of applying AI to science goes toward the *automation of science*. Researchers enthusiastic about AI aim to develop AI systems that automatically execute the whole tasks of scientific research (such as exploring the literature, designing and running experiments, interpreting data, writing and reviewing papers, and so on). One of the most remarkable achievements in this attempt was done by Robot Scientist "Adam," which King's research group developed (King et al. 2004; King et al. 2009; King 2010). Adam is a robot that generates hypotheses, derives their consequences, and tests them automatically by itself. By doing this, it discovered genes encoding an enzyme required for the growth of yeast in the area of functional genomics. Furthermore, as a future challenge, Hiroaki Kitano, the Director of Sony Computer Science Laboratories, has set a goal "to develop an AI system that can make major scientific discoveries in biomedical sciences and that is worthy of a Nobel Prize and far beyond" (Kitano 2016b, p. 39). To denote this, a new mode of scientific research characterized by automation, some researchers use the term "AI-driven science."[1]

As matters now stand, AI systems are applied to science only to support

[1] To my knowledge, the term "AI-driven science" was coined by Koichi Takahashi, a Japanese biologist working at the Institute of Physical and Chemical Research in Japan.

human researchers. Newspapers sometimes report that *AI has discovered* something, but this is not a precise description of the state of affairs. It is more accurate to say that *humans have discovered something by using AI*, since AI systems at present do not execute tasks *with their own intentions*: they are nothing but non-autonomous tools, as we will see in Section 2. However, there are some ambitious researchers who aim to develop truly autonomous AI systems for scientific research. For example, Kitano mentions "systems that acquire knowledge autonomously and make discoveries continuously" (Kitano 2016a, p. 84, my translation) as the ultimate end of his challenge. Thus, AI-driven science is potentially a long range endeavor.

1.2 Impacts of automating science and features of AI as a scientific technology

Leaders of AI-driven science, such as King and Kitano, expect that the automation of science will bring enormous benefits. They typically mention two types of its merits: practical and intellectual. I will explain them in turn.

As for the practical merits, King and his co-authors say,

> We consider this trend to increased automation of science to be both inevitable and desirable. It is inevitable because it will be required to deal with the challenges of science in the twenty-first century. It is also desirable because it frees scientists to make the high-level creative leaps at which they excel. (King et al. 2004, p. 251)

One of what King et al. call "the challenges of science in the twenty-first century" above is the exponential increase in amount of data to be handled. Situations such as the rapid growth of gene databases and the emergence of petabyte-scale astronomical data through sky surveys make it impossible for researchers to analyze data by themselves. Given this, applying AI can be considered inevitable for science to progress further. Another practical merit King et al. mention, i.e., freeing scientists from non-creative tasks so that they can concentrate on creative ones, is also important, in particular under the present situation in which post-doctoral researchers and graduate students, sometimes called "pipetting slaves," devote long hours in the laboratory to dull and repetitive works such as cleaning pipettes. Furthermore, ambitious leaders of AI-driven science claim that the automation of science will dramatically increase research productivity, and contribute to solving social problems. For example, Kitano mentions the practical merits of his challenge as follows:

> I anticipate that, in the near future, AI systems will make a succession of discoveries that have immediate medical implications, saving millions of lives, and totally changing the fate of the human race. (Kitano 2016b, p. 39)

The other merits of automating science its proponents mention are intellectual: It sheds light on *what science is* (e.g., King 2010; Kitano 2016a). We can see this clearly by referring to what is called the "N = 1 problem" in biology. Biologists' attempts to identify the essence of life are prevented by the fact that all the samples of life available to them belong to a single lineage of terrestrial lives. Even if biologists identify certain features of life common to all the available samples (such as being constituted by cells), they cannot tell whether they are essences of life or merely contingent features of terrestrial lives. Therefore, they put their hope in astrobiology and A-Life to obtain samples of other life forms. The same is true of science: The only sample of science available to us so far has been humanity's science. This fact constitutes an obstacle to discriminate the essential features of science from its contingent ones resulting from constraints by human cognitive and other limitations. Therefore, it is helpful to develop AI capable of scientific research. If AI acquires the intelligence required for scientific research, it will probably develop sciences that are significantly different from humanity's one. The emergence of these alien sciences will contribute to our understanding of what science is.

Although I am somehow sceptic about optimistic discourses by leaders of AI-driven science, I nevertheless agree with them on the prediction that it will radically change the overall state of science and technology. Underlying its potential power to cause such changes are some unique features of AI as a scientific technology. Of course, various technological devices, such as telescopes, microscopes, and computers, have been used in scientific research throughout the history of science. However, AI has some features that are not found in other scientific technologies. I will mention just two of these features.

First, AI systems can be used to carry out research tasks automatically and make human intervention unnecessary. This is the very feature that underlies the practical merits of the automation of science, such as increase in the productivity of research and liberation of people from non-creative research tasks.

Second, AI systems may generate knowledge beyond human understanding. Since older devices presuppose manipulation by humans, they are designed so that their behaviors remain within the range of human understanding. In contrast, AI systems do not necessarily have this limitation. Indeed, they may even have the potential to bring forth kinds of "alien science," a system of knowledge that understands the world in certain ways different from ours. Kevin Kelly, the founding editor of *Wired* magazine, makes this point by saying,

> AI could just as well stand for "alien intelligence." (…) An AI will think about science like an alien, vastly different than any human scientist, thereby provoking us humans to think about science differently. (Kelly 2016, p. 48).

However, the automation of science may bring significant problems and undesirable effects as well as the benefits described above. As an example of these problems, it blurs *who* does the research at the price of making human intervention unnecessary. This will pose difficult problems concerning credit and responsibility for scientific research. As another example, researchers who use AI in their research face a "black box" problem: What they discover by using AI systems might go beyond their understanding. This can pose difficult questions as to whether such findings qualify as "knowledge" (or "scientific knowledge") and how much epistemic value they have.

Further discussion of these issues is beyond the scope of this article. Instead, I concentrate henceforth on the issues concerning *creativity*.

2. The Possibility of Artificial Creativity

Leaders of AI-driven science focus especially on scientific *discovery*. Since discovery is the process of yielding new knowledge, and since it is thought to be one of the most creative phases of scientific research, its automation will have great impacts both practically and intellectually. However, because scientific discovery requires creativity, it can be difficult for AI systems (and robots) to execute it automatically. This section examines whether AI can make creative discoveries, and, if possible, how they can be automated.

2.1 What is creativity?

Let us clarify what "creativity" means by referring to Boden's analysis of the concept. Boden (2004) defines creativity as the capacity to create ideas or artifacts that are *new*, *surprising*, and *valuable*.[2] Although she defines it as a kind of capacity, activities of exercising it and products created by such activities can be called creative, too. Typically cited examples of such activities are artistic creation, technological invention, and scientific discovery.

In addition to a definition, Boden proposes two taxonomies of creativity. The first distinguishes two sorts of creativity in terms of the kinds of "newness": One is "*psychological creativity*," which concerns a creative product that is new to the individual who produces it; and the other is "*historical creativity*," which concerns a creative product that is new to humanity (or to a certain community). Of course, scientific creativity is classified as the latter. Hence, for scientific AI to be creative, it must avail itself of some method to generate ideas that go beyond past knowledge of the overall science community.

Boden's second taxonomy distinguishes the following three types of creativity in terms of the ways surprising ideas are generated: (1) "*combinatorial creativity*," which is achieved by combining familiar ideas in some unfamiliar way; (2) "*exploratory creativity*," which is achieved by exploring a "conceptual space" following some style; and (3) "*transformational creativity*," which is achieved by transforming a pre-existing conceptual space. Here, Boden uses the term "conceptual space" to refer to a space that consists of all possible ideas about some topic. In the context of scientific discovery, all possible hypotheses to a question constitute the conceptual space concerning the question, for example. Section 2.3 will examine approaches to artificial creativity by referring to these three kinds of creativity.

2.2 Can artificial creativity be realized?

There has been a lively debate concerning whether AI can be creative. A common argument against artificial creativity is called *Lady Lovelace's objection*.

[2] Though the condition of surprise is often dropped from the requirements for creativity in psychological literature, I think it captures an important aspect of creative activities such as scientific research. For example, Dyson says, "For science to be great it must involve *surprises*, it must bring discoveries of things nobody had expected or imagined." (Dyson 1988, p. 165, emphasis added.)

It is so called since its most classical version is found in Ada Lovelace's comment on Charles Babbage's analytic engine. According to this objection, machines (such as computers and robots controlled by them) cannot be creative because they can do only what they are programed to do. If a machine creates something, the relevant creativity should be attributed to its programmer rather than the machine.

However, also famous is Turing's (1950) reply to this objection:

> This may be parried for a moment with the saw, "There is nothing new under the sun." Who can be certain that "original work" that he has done was not simply the growth of the seed planted in him by teaching, or the effect of following well-known general principles. (p. 450)

Here, Turing points out that even human creativity is not *creation ex nihilo*, and argues that being programmed is on a par with being taught or informed. The moral we can learn from his reply is that to think artificial creativity cannot be realized in principle might be tantamount to *mystifying* human creativity.

Opponents of artificial creativity would not be convinced, yet. Rather, they might claim that Lady Lovelace's objection can be refined by focusing on *autonomy* (see Boden 2014). The revised argument goes as follows: Machines cannot be creative, since creativity requires autonomy, and machines that can do only what they are programed to do necessarily lack it. In this context the term "autonomy" is used in the philosopher's sense in which it denotes the ability to set one's own goals freely (for example, choosing the tasks). It should be distinguished from mere "automaticity" (or "autonomy" in the roboticists' sense), which denotes the ability to operate without being controlled by other agents. It is important to note here that agents can be automatic without being autonomous: There can be agents that execute tasks set by others without being controlled by others.

The claim that creativity requires autonomy seems plausible to a certain degree (though not to the full extent, since we do not understand even how autonomy is realized in humans). More controversial is the claim that machines *necessarily* lack autonomy. To defend this claim, some theorists argue that autonomy requires life. However, life is a complex phenomenon that consists of many features of living systems (such as reproduction, metabolism, self-organization, evolution, development, etc.), and it is not clear which feature is relevant to autonomy and why it is relevant. Moreover, it is not obvious that AI systems necessarily lack the relevant feature: Studies in A-Life have shown that many interesting features of living systems can be realized by artificial systems. Anyway, it is without controversy that no AI systems applied to scientific research today lack autonomy, and there seems to be little prospect that truly autonomous AI scientists will appear in the near future. Thus, the revised version of Lady Lovelace's objection seems reasonable for the time being.

Nevertheless, we should be careful to identify what this consideration means. According to the idea, even when an AI system devoid of autonomy produces something comparable to products of human creative activities, it does not qualify as exercising true creativity: it turns out instead that the AI system only simulates creativity. This view, though, does not exclude the possibility that even such a non-creative AI system can exhibit as good performance as truly creative agents do. Moreover, at present, proponents of AI-driven science do not necessarily aim to develop truly autonomous, strong AI scientists, but rather to devise AI systems as mere useful tools for scientific research. So, if AI systems can perform research tasks as well as human researchers, it can be argued that whether they are truly creative or not in themselves is not important for the immediate goals of AI-driven science. Furthermore, regardless of whether such systems are deemed truly creative or not, we can still wonder how to develop such high-performance AI systems. Therefore, let us put aside the issue of artificial autonomy, and consider next how creative discoveries can be achieved through the use of AI. (Henceforth, for simplification, I use the term "creativity" to refer not only to the capacity exercised by truly autonomous agents, but also to the capacity exhibited by non-autonomous systems that generate ideas in as high-performance as them.)

2.3 How can artificial creativity be realized?

In this section we examine three ways to achieve creative discoveries through the use of AI, which correspond to three kinds of creativity Boden identifies (i.e., explorative, transformative, and combinatory creativity).

The combinatorial approach

A major view of creative activity (including scientific discovery) states that it consists of an *unfamiliar combination of familiar ideas*. This view has been put forth by various theorists (e.g., Poincaré 1908; Asimov 1959) and supported in the psychology of science (e.g., Simonton 2004). Let us call it the "combinatorial view." In the domain of science, a famous example of discoveries accomplished in this manner is Darwin's (and Wallace's) discovery of evolution by natural selection. It is said that Darwin came to his theory of evolution by combining multiple ideas such as the idea of "overpopulation and weeding out" which he drew from Malthus, the idea of selective breeding of animals and plants, and the idea concerning how species diverge, which he confirmed during the voyage of the Beagle (see Asimov 1959; Bowler 1983).

Can we develop AI systems that make creative discoveries by adopting the combinatorial view? Combination can be accomplished by AI and robots. However, as Boden (2004) states, most of the resulting products are not valuable. Therefore, a certain method of selecting valuable ideas is required. Taking this point into account, once Poincaré (1908) denied that machines can make mathematical discoveries, since the rules concerning selecting valuable ideas are too subtle for machines to apply. Though his argument seems to be question-begging, it is persuasive that the combinatorial approach faces a dilemma: On the one hand, to conceive of a *surprising* idea, unexpectedness of the combination is important; on the other hand, as Boden (2016) suggests, some kind of relevance of combined ideas is important for conceiving of a *valuable* idea. Unexpected combinations of relevant ideas are rarely found. Especially, it seems quite difficult to conceive of combinations of relevant ideas each of which belongs to a different knowledge domain, as we find in Darwin's theory of evolution.

What capacities or mechanisms do the trick in the case of humans? An often-cited candidate is *analogy*, or the kind of inference that derives knowledge of unfamiliar problems or situations from knowledge of familiar problems or situations. According to cognitive scientists Holyoak and Thagard (1995), it involves the mental act that associates knowledge of some familiar problem or situation (a "base") and knowledge of unfamiliar one (a "target"), and finds out some structural similarity between them. Holyoak and Thagard also point out that analogy plays an important role as a "mental mechanism for combining and recombining ideas in novel ways" (ibid., p. 13) in creative thinking including scientific discovery, and they mention many examples to demonstrate it. For example, acts of analogy that find similarities between the struggle for life in humans and that in animals and plants and between selective breeding (artificial selection) and natural selection enabled Darwin to build his theory of evolution.

However, there is a finding suggesting that analogy plays only a *limited* role in scientific discovery. Psychologist Dunbar (1997) shows, on the basis of fieldwork in several labs, that analogy that associates pieces of knowledge from different science areas (to take an example from molecular biology, one that invokes knowledge of something other than organisms rather than knowledge of organs in the same organism or knowledge of different organisms) is rarely used in research practice, and that analogy is less frequently used for generation of hypotheses than for explanation. Given these points, it might be better to regard analogy as a means to understanding rather than to discovery.[3] At any rate, it is sure that we do not have enough knowledge of the role analogy plays in scientific discovery or of the mechanisms underlying analogy.

Another consideration is concerned with *aesthetic judgement*. Poincaré claimed that the selection of useful combinations of ideas is enabled by aesthetic sensitivity. Though he also argued that machines cannot make aesthetic judgement (and therefore he denied the possibility of artificial creativity), there is room for doubt in this regard. Thus, an interesting orientation of future research lies in examining how we could develop AI systems capable of making aesthetic judgement, as well as trying to understand the role of aesthetic judgement and the nature of aesthetic values in scientific discovery.

The exploratory approach

Another view of creativity, which is especially popular in cognitive science, states that scientific discovery consists in the *exploration of possible ideas*. Let us call this view the "exploratory view." This view is a natural consequence of fundamental assumptions of cognitive science, namely, that scientific discovery is problem solving and that problem solving is searching the problem space (i.e., the space that consists of all possible solutions) for the solution by means of heuristics (see Simon 1996). For example, according to cognitive scientist Anzai (1985), there is evidence in the literature that Watson and Click's discovery of the double helix structure of DNA and Faraday's discovery of the law of induction were accomplished in this manner.

Then, should we adopt the exploratory view to achieve creative discoveries by AI? Surely exploration can be accomplished by AI systems even much more

[3] Actually, some complication must be added to this consideration, since it seems that we cannot discover what we cannot understand. I do not pursue this point further in the present article.

effectively than by humans. Nevertheless, there is a good reason to think that producing something merely by following a mechanical procedure of exploration does not qualify as a creative act. Philosopher Novitz (1999) shows this clearly by referring to the example of Goodyear's invention of vulcanized rubber. To make rubber products heat- and cold-resistant, Goodyear had combined rubber with various substances at hand haphazardly, and it took many years to finally produce sulfur. Though no doubt Goodyear made an important discovery as a result of an admirable effort, it is difficult to say that his discovery is a creative one comparable to Newton's and Darwin's ones.

For this reason, some theorists (e.g., Gaut 2003) claim that creativity requires a certain non-mechanical factor such as "flair." However, such a claim does not qualify as satisfying unless what kinds of cognitive capacities and mechanisms constitute the "flair" is clarified. Moreover, proponents of the exploratory view would reply that, even given that the good-old fashioned "generation-test method" Goodyear adopted does not suffice to yield creative achievements, we can achieve them by adopting certain more sophisticated procedures. For example, Dawkins (1986) says, "Effective searching procedures become, when the search-space is sufficiently large, indistinguishable from true creativity" (p. 66).

Nevertheless, it seems reasonable to suspect that a discovery accomplished by exploration is not deemed radically creative as long as it is made according to some pre-existing style. What matters is the nature of the surprise evoked. According to Boden (2004), any idea resulting from exploring a pre-existing conceptual space can be described and yielded by applying pre-existing generative rules. Though it sometimes causes a surprise that something unexpected happens, we can understand that it comes under familiar patterns once it happened. In contrast, radically creative ideas are ones that could not be yielded merely by applying pre-existing generative rules: They evoke huge surprises by making possible something that had not been thought to be possible.

The transformational approach

Then, how can such a radically creative discovery be achieved? Boden claims that *transforming the conceptual space* is a means to such achievement.[4]

[4] Novitz denies Boden's identification of radical creativity with transformational creativity, arguing that radically creative ideas can be produced even when any relevant conceptual space does not exist. I concede that the transformational approach is not the *only* means to radical creativity. What matters for

Although the kinds of discoveries made in this manner are rare, Kepler's discovery of his First Law is one example. The pre-modern astronomy Ptolemy established was governed by the thought that the orbits of planets are round. Kepler departed from this way of thinking after examining Tycho Brahe's observation data concerning planetary orbits and found that planets orbit elliptically. As is well known, this discovery led to the scientific revolution in the 17th century.

The transformational approach seems to grasp better the creative character of creative discoveries than the exploratory approach does. Then, is it possible to allow AI to transform the conceptual space automatically? According to Boden (2004), transformation of a conceptual space can be achieved by deviating or modifying the constraints that shape it. For example, non-Euclidian geometrics was established by removing the fifth postulate ("parallel postulate") of Euclidian geometrics. Once, as is in this case, we specify the constraints that shape a conceptual space, we might succeed in yielding transformation of the conceptual space automatically by formulating the task of deviation or modification of it as a search problem. This is an interesting possibility. However, there is a difficulty to the approach: Most scientific problems are ill-defined, and therefore it is hard in many cases to specify the constraints in any formal procedure. Philosophers of science Bechtel and Richardson (1993) point out this feature of scientific problems by saying, "the constraints defining an adequate solution are not sharply delineated, and even the structure of the problem space itself is unclear" (p. 15). Unless we can specify the constraints, we can neither deviate nor alter them. Of course, this difficulty also confronts human researchers attempting to make scientific discoveries. Indeed, it might be the very reason that discoveries depend heavily on accidents or serendipity. As long as AI does not overcome this difficulty, any attempt to automate the process of discovery will not achieve great success.

As the discussion above demonstrates, all three approaches to scientific discovery by AI have some difficulty or uncertainty in the present situation and cannot be seen as a decisive way to make creative discoveries yet.

---

radical creativity is the effect that a novel style of thought and a novel conceptual space emerge, rather than the means to it. At the same time, I agree with Boden that the transformation of pre-existing conceptual space is, though not *essential*, an *effective* means to that effect.

2.4 Possibility of the brute force approach to artificial creativity

In this subsection, let us examine a currently proposed approach to AI-driven scientific discovery. The approach in question is the one proposed by Kitano. He puts forward "a brute force approach in which AI systems generate and verify as many hypotheses as possible" (Kitano 2016b, p. 46). This can be regarded as an extreme version of the exploratory approach. As Kitano realizes, this approach differs substantially from ones in which human researchers achieve discoveries. He states that the current state of scientific discovery, which depends on unreliable intuition and serendipity, is at the level of "cottage industry" (ibid., p. 41). Thus, he declares, "AI scientific discovery systems have the potential to drive a new revolution that leads to new frontiers of civilization" (ibid., p. 48).

Can Kitano's brute force approach accomplish his purpose to bring about innovation in the mode of research? It may be possible, and his proposal surely is worth pursuing. However, when it comes to creativity, there is a worry, namely, that the science resulting from it does not seem to exhibit creativity. This is because, as Goodyear's case suggests, a discovery accomplished by merely exploring some pre-existing conceptual space yields no huge surprise, no matter how exhaustive the exploration is. Proponents of the brute force approach would reply that, though the approach in question may not achieve something like human creativity, it will achieve a different kind of creativity. It is true that creativity can take other forms than ours; to think otherwise is to commit a sort of anthropocentric chauvinism. However, why would activities that are considered non-creative if they are carried out by humans be deemed creative when they are carried out by machines? If we are to reject anthropocentrism, we should think that whether activities are creative or not does not depend on *who* carries out them. This thought leads to exclusion of the brute force approach from the means to creative discovery. Moreover, if the brute force approach will nonetheless become prevalent, the value of science as a creative activity might be compromised.

Let us summarize this section. While there is no sound ground for the claim that AI cannot make creative discoveries in principle, we do not have enough knowledge to automate such discoveries. Indeed, we do not fully understood how humans achieve them yet. Therefore, in order to develop AI systems that make them, we should accumulate empirical knowledge concerning human creativity at the first onset. If we avoid such a steady effort and harry to automate science in ways that resort to brute force, it could lead to a state of affairs in which some non-creative mode of scientific research would prevail.

There is no doubt that, as the use of AI becomes common in science, modes of scientific discovery will change. On this occasion, it seems fruitful, at least for the time being, to establish organizations in which AI systems and human researchers can cooperate in an appropriate way, rather than leaving the whole process of discovery to AI systems. As long as AI is introduced into science in such a manner, it will become a powerful tool for discovery due to its great capacity for exploration.

3. The Implications of Automating Science

In this section, let us consider what implications the AI-driven science may have on the scientific community and wider society. The automation of science will not only benefit scientists and other people, but also bring about undesirable effects. Although there are many kinds of worries (some of which I mentioned briefly in the end of Section 2.2), here I will examine just two of them: technological unemployment and undermining the value of science.

One of the worries about the automation of science is the threat of technological unemployment. Will AI take human researchers' jobs, as Chiang depicts? At present, many researchers would answer "no" to this question. For example, Frey and Osborne's famous report "The Future of Employment" (2013) states that jobs requiring creativity (such as those of artists and scientists) cannot easily be automated. Indeed, as we saw in Section 1, leaders of AI-driven science (e.g., King et al. 2004) often claim that the automation of science will liberate researchers from dull tasks and enable them to concentrate on creative works. We have, though, some reasons to cast doubts on such optimistic expectations, as explained below.

The prediction that truly creative AI will not appear seems plausible in the short term. However, this does not mean that human researchers' position is secure. The reason is that the place of creative tasks in the whole system of science may not remain constant. Rather, even given that tasks requiring creativity are difficult to automate, the weight of these tasks in the whole scientific research might decrease. Norman (2007) makes this point with respect to the general context of the effect of automation:

> In general, whenever any task is automated, the impact is felt far beyond the one task. Rather, the application of automation is a system issue, changing the way work is done, restructuring jobs, shifting the required tasks from one portion of the population to another, and, in many cases, eliminating the need for some functions and adding the need for others. (p. 117)

Norman's general observation applies especially well to science. In contemporary society, scientific research is regarded as an important means of innovation and solution to social problems, and so a huge amount of public resources is spent on it. Therefore, if it turns out that AI-driven science is much more productive in generating useful knowledge and applications than traditional modes of research, it is possible that governments will spend resources on the former rather than the latter and, consequently, many researchers will lose their jobs.

Also, the prediction that AI will liberate researchers seems too optimistic. It is often said that past attempts to automate labors have not liberated humans from unattractive works, but instead gave birth to populations engaged in mechanical and inhuman works. Norman says, "Even successful automation always comes at a price" (ibid.) and mentions drawbacks of automation such as the need for maintenance. To apply these general lessons, we should not expect that the automation of science will eliminate dull tasks in research. Indeed, the job of scientific researcher itself might be unattractive even if it would not be replaced.

Another worry about the automation of science is concerned with its more far-reaching effect: it may undermine a central value of science. Science constitutes an important part of human culture. However, the automation of science may undermine its cultural value by decreasing the room for human ingenuity in scientific research. To specify the threat concretely and to find a way to deal with it, it is useful to refer to discussions held in the past when new technologies were introduced into science. In this spirit, let us reflect on discourses made when "big science" emerged.

Big science is a family of scientific programs that are funded large budgets, carried out by large teams of researchers, and exploit large devices (e.g., giant telescopes, spacecrafts, particle accelerators, nuclear fusion reactor, etc.). Starting with the Manhattan Project, they emerged during and after the Second World War. Since then they have caused controversies due to the impacts on science and society. For example, as we saw above, Dyson viewed human creativity as the basis of a central value of science and claimed that large devices are not important for scientific research. Likewise, Weinberg, a nuclear physicist who coined the term “big science,” was worried about the consequences of big science’s growth (Weinberg 1961). One of his concerns was directed at scientists’ tendency of “spending money instead of thought” (ibid., p. 162), which, Weinberg states, may ruin science. These discourses suggest that scientific research has an aspect as a drama, protagonists of which are humans who try to understand the world by exercising their creativity. This is, however, the very aspect that the automation of science may endanger.

Therefore, in introducing AI into science, we must specify its implications and devise some measures against its demerits beforehand. To set a guiding principle in this attempt, we should refer to Weinberg’s following comment:

> Big Science is an inevitable stage in the development of science and, for better or for worse, it is here to stay. What we must do is learn to live with Big Science. We must make Big Science flourish without, at the same time, allowing it to trample Little Science (ibid., p. 162).

The first sentence of this passage corresponds to King’s diagnosis that science will inevitably be automated (see Section 2.2). Thus, punning on Weinberg’s passage, we should say, “What we must do is learn to live with AI-driven Science. We must make AI-driven Science flourish without, at the same time, allowing it to trample Humanity’s Science.”

Thus, we must discuss potential demerits of automating science and measures against them so that we find desirable ways of introducing AI into science. Some of the matters on the agenda are concerned with science policy: for example, resource allocation between AI-driven science programs and traditional ones, measures to ensure employment of researchers, and so on. Some are concerned with science education, such as alteration of science curriculum in universities.[5] To address these issues, it is important to take opinions from various stakeholders such as researchers, practitioners of science policy, science education and science communication, and broader citizens.

[5] The automation of science also raises issues concerning institutional systems of scientific research such as the authorship system and the referee system, although they are not discussed in the present article due to the limitation of space.

Conclusion

This article examined the implications of the automation of science, focusing on creativity. My view is that, although the attempt to automate science faces difficult challenges in realizing artificial creativity, it nevertheless will have significant impacts, both desirable and undesirable, on science and society.

In conclusion, I stress that the automation of science raises many issues that require transdisciplinary research and discussions. On the one hand, to address issues concerning its effect on science and society, it is essential to hold a discussion whose participants include not only AI researchers and scientific researchers who use AI systems, but also researchers and practitioners of science policy, science education, and science communication, as well as sociologists and philosophers of science. On the other hand, issues concerning the possibility of automating scientific discovery provide opportunities for reconsidering philosophical questions such as "What is scientific discovery" and "What kinds of values does science have?" from a new perspective. Although these topics are deeply philosophical, they also require contributions from researchers of empirical sciences such as cognitive scientists, psychologists, and sociologists. By coping with these issues in transdisciplinary collaboration, we will gain a better understanding of the nature and values of science, and this will be one of the most important benefits of AI-driven science.

References

Anzai, Y. 1985. *The Psychology of Problem Solving*, Tokyo: Chuokoron-shinsha. (Written in Japanese)

Asimov, I. 1959/2014. 'How do people get new ideas?', reprinted in *MIT Technology Review*, October 20, 2014. [Retrieved February 4, 2020 from https://www.technologyreview.com/s/531911/isaac-asimov-asks-how-do-people-get-new-ideas/]

Boden, M. A. 2004. *The Creative Mind: Myths and Mechanisms (Second Edition)*, London: Routledge.

——— 2014. 'Creativity and artificial intelligence: a contradiction in terms?', in E. S. Paul & S. B. Kaufman (eds.), *The Philosophy of Creativity: New Essays*, pp. 224-244, Oxford: Oxford University Press.

Bowler, P. J. 1983. *Evolution: The History of an Idea*, Berkeley: University of California press.
Chiang, T. 2002. 'The evolution of human science', in his *Stories of Your Life and Others*, pp. 241-244, New York: Tor Books (originally published in 2000 as 'Catching crumbs from the table', *Nature* 405 (6786): 517).
Dawkins, R. 1986. *The Blind Watchmaker*, Harlow: Longman Scientific & Technical.
Dunbar, K. 1997. 'How scientist think: On-line creativity and conceptual change in science', in T. B. Ward, S. M. Smith & J. Vaid (eds.), *Creative thought: An investigation of conceptual structures and processes*, pp. 461-493, Washington, DC: American Psychological Association.
Dyson, F. J. 1988. 'Science and space', in his *Infinite in All Directions*, pp.158-179, New York: Harper & Row.
Frey, C. B. & Osborne, M. A. 2013. *The Future of Employment: How Susceptible Are Jobs to Computerisation?* [Retrieved February 4, 2020 from https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf]
Gaut, B. 2003. 'Creativity and imagination', in B. Gaut, & P. Livingston (eds.), *The Creation of Art: New Essays in Philosophical Aesthetics*, pp. 148-173, Cambridge: Cambridge University Press.
——— 2010. 'The philosophy of creativity', *Philosophy Compass* 5 (12): 1034-1046.
Holyoak, K. J. & Thagard, P. 1995. *Mental Leaps: Analogy in Creative Thought*, Cambridge, MA: MIT Press.
Kelly, K. 2016. *The Inevitable: Understanding the 12 Technological Forces That Will Shape Our Future*, New York: Viking Press.
King, R. D., Rowland, J., Oliver, S.G., Young, M., Aubrey, W., Byrne, E., Liakata, M., Markham, M., Pir, P., Soldatova, L.N., Sparkes, A., Whelan, K.E., & Clare, A. 2009. 'The automation of science', *Science* 324 (5923): 85-89.
King, R. D., Whelan, K. E., Jones, F. M., Reiser, P. G. K., Bryant, C. H., Muggleton, S. H., Kell D. B. & Oliver S. G. 2004. 'Functional genomic hypothesis generation and experimentation by a robot scientist', *Nature* 427 (6971): 247-252.
King, R. D. 2010. 'Rise of the robo scientists', *Scientific American* 304 (1): 72-77.
Kitano, H. 2016a. 'The day AI win the Nobel prize and the future of humanity:

An ultimate grand challenge in AI and scientific discovery', *Jinkou Tinou* 31 (2): 275-286. (Written in Japanese)
——— 2016b. 'Artificial intelligence to win the Nobel prize and beyond: Creating the engine for scientific discovery', *AI Magazine* 37 (1): 39-49.
Langley, P. 2000. 'The computational support of scientific discovery', *International Journal of Human-Computer Studies* 53 (3): 393-410.
Langley, P., Simon, H. A., Bradshaw, G. L. & Zytkow, J. M. 1987. *Scientific Discovery: Computational Explorations of the Creative Processes*, Cambridge, MA: MIT Press.
Lenat, D. B. 1977. 'The ubiquity of discovery', *Artificial Intelligence* 9 (3): 257-285.
Norman, D. 2007. *The Design of Future Things*, New York: Basic Books.
Novitz, D. 1999. 'Creativity and constraint', *Australasian Journal of Philosophy* 77 (1): 67-82.
Poincaré, H. 1908/2001. *Science and Method*, reprinted in his *The Value of Science: Essential Writings of Henri Poincaré*, pp.355-567, translated by F. Maitland, New York: Modern Library.
Science News Staff 2017. 'AI is changing how we do science. Get a glimpse', *Science* Jul. 5, 2017. [Retrieved February 4, 2020 from http://www.sciencemag.org/news/2017/07/ai-changing-how-we-do-science-get-glimpse]
Searle, J. 1980. 'Minds, brains, and programs', *Behavioral and Brain Sciences* 3 (3): 417-424.
Simon, H. 1996. *The Sciences of the Artificial (3rd edition)*, Cambridge, MA: MIT Press.
Simonton, D. K. 2004. *Creativity in Science: Chance, Logic Genius, and Zeitgeist*, Cambridge: Cambridge University Press.
Stix, G. 2005. 'Molecular treasure hunt', *Scientific American* 292 (5): 88-91.
Turing, A. 1950. 'Computing machinery and intelligence', *Mind* 59 (236): 433-460.
Weinberg, A. 1961. 'Impact of large-scale science on the United States', *Science* 134 (3473): 161-164.