Note

Arrival of Singularity and Role of Criminal Law: Raising Issues

Osamu MAGATA*

I. Introduction

The Technological Singularity is the time when artificial intelligence (AI) surpasses human intelligence and humans lose control of AI activities. According to a survey, 50% of expert respondents expected that human-level machine intelligence (HLMI) to be realised by the year 2040. Also, many experts predict that when HLMI is realised, shortly after that, "Superintelligence",[1] intelligence that far exceeds human capabilities, will be created. At this point arrives the singularity.[2]

* Professor at the Faculty of Law, Chuo University.

1) Good, Irving John, [1965] "Speculations Concerning the First Ultraintelligent Machine", Advance in Computers, Volume 6, p. 33, and Bostrom, Nick, [2014] *Superintelligence: Paths, Dangers, Strategies*, chapter 2, especially p. 22. *See also*, Gill, Karamjit, [2016] "Artificial super intelligence: beyond rhetoric", AI & Society, Volume 31, pp. 137–143.

2) At which point in time arrives the singularity could be controversial, but this

The advent of the singularity can pose a severe threat to the human future, because Superintelligence could initiate activities that exceed human expectations by using its abilities that far exceed human capabilities.

Today, there are still strong denials of this threat,[3] and as a result, the legal response to this threat has not been fully examined. However, if the threat becomes a reality, it will cause immeasurable harm such as the destruction of humankind and its enslavement to Superintelligence.

The threat from the advent of the singularity should not be overlooked; nevertheless, it has not been fully discussed from a legal point of view. Therefore, this paper raises questions as to what measures should be taken to control the severe threat to humankind in advance, and asserts the effectiveness of the measures using "criminal law." At the same time, it attempts to begin a debate about how criminal law can be used and what points shall be taken into consideration.

II. Possible threat

1. Recent years have witnessed a remarkable development of AI, of which many researchers and engineers are in favour. A certain number of engineers are, in the first place, in denial about the arrival of the Technological Singularity,[4] Examples include the reduction of accidents thanks to autonomous driving systems and the dramatic growth of medical technology. As such, it is not appropriate to unreasonably hinder the development of AI technology.[5]

---

paper regards the birth of Superintelligence as the point of arrival of the singularity. This is because the birth of Superintelligence is the decisive timing when human control over it becomes impossible.

3) Not a few people are not strongly aware of this threat, even if do not deny it.

4) Kurzweil, Ray, [2006] *The Singularity Is Near; When Humans Transcend Biology.*

5) *See*, Koki, Nezu, [2019] "Strafrechtlicher Problemaufriss von (teil)autonomen

However, attention should be paid to the overly optimistic view of the development of AI. According to a survey, 25% of engineers make pessimistic predictions about the post-singularity world that could come beyond the evolution of AI.[6)]

2. The threat posed by the arrival of the singularity is presented in the following scenario: First of all, if AI develops at this pace, it will eventually develop at an exponential speed to reach HLMI; that is to say, artificial general intelligence (AGI) will appear. AGI at that level will have demands for self-preservation and pursuit of self-purpose like humans and will conceal their enhanced capabilities in order to prevent humans from blocking them. This allows humans to perceive AGI as harmless. AGI, which will explosively improve its ability in such a situation, evolves into "Superintelligence". At this point arrives the singularity. Superintelligence will be far more capable than humans; therefore, humans will be unable to control its activities at all. The world will move according to the will of Superintelligence. If Superintelligence does not have the element of human prosperity in its self-purpose, it will expel human beings as unnecessary beings. In the end, the destruction or enslavement of humankind by Superintelligence will occur.

The above is the pessimistic scenario of the singularity.[7)]

3. This is by no means pure science fiction but a possible real-life scenario, which is endorsed by experts.[8)] Nonetheless, there is not always enough

---

Fahrzeugen in der Gegenwart und Zukunft", Hanover Law Review, Heft. 4, p. 271.

6) *See*, Kruel, Alexander, [2011] "Interview Series on Risks from AI". *See also*, Geist, Edward Moore, [2015] "Is artificial intelligence really an existential threat to humanity?" http://thebulletin.org/artificial-intelligence-really-existential-threat-humanity8577.

7) *See*, Bostrom, [2014] *supra* note 1, chapter 5 & 6, and Reese, Byron, [2018] *The Fourth Age: Smart Robots, Conscious Computers, and the Future of Humanity*, Part Three, especially pp. 184–189.

8) *E.g.* Stephen Hawking, Elon Musk, Bill Gates. As researchers who have some

debate about how to avoid the worst-case scenario of this pessimism. Especially, there has been little discussion from the perspective of legal regulation.[9)]

That is probably because the worst-case scenario for the arrival of the singularity is nothing but "potential". However, this conversely indicates that the optimistic scenario for it is also just a matter of "potential." The mechanism leading up to the glorious future of humankind has also not been well scientifically proven.

4. Is it, nevertheless, correct to "bet" on the arrival of an optimistic future? I do not believe that it is appropriate, because the bet seems too dangerous for humankind and too irresponsible. If the optimistic future turns out to be a myth, the reality that comes about is the destruction or enslavement of humankind by Superintelligence. It is truly an "irreparable" situation for humanity.

III. Interest to be protected by criminal law and conducts to be regulated

1. In order to certainly prevent the "irreparable situation" of the destruction and enslavement of humankind by Superintelligence, it is probably effective to regulate AGI Research and Development (R&D) to a certain extent.[10)] If it is done by only a loose method, however, it could hardly be expected to be effective. Many researchers and engineers are expected to continue the development to keep up with each other, anticipating an opti-

---

degree of fear for the development of AGI include Norbert Wiener, Irving John Good, Steve Wozniak, Gary Marcus, and others.

9) *See* however, e.g. Castel, J.-G. and Castel, Matthew E., [2016] "THE ROAD TO ARTIFICIAL SUPER-INTELLIGENCE: HAS INTERNATIONAL LAW A ROLE TO PLAY?", Canadian Journal of Law & Technology, Vol. 14(1), pp. 1–15.

10) *See*, Castel, J.-G. and Castel, Matthew E., [2016] *supra* note 9, p. 9.

mistic future. (The rules such as "Draft AI R&D GUIDELINES for International Discussions"[11] already presented can, of course, be important and valid. The problem, however, is that lenient rules based on non-binding targets are not sufficient enough.)

What is needed is a strong binding power of criminal law. A method to discipline AGI R&D using a strongly enforceable criminal law must be considered.

In doing so, first, it is of course required to try making full use of the existing criminal statutes. However, most of the penal provisions that currently exist are "result criminals" that require the occurrence of a result. Given the purpose of this paper, which aims to prevent the occurrence of results, the punishment after waiting for the occurrence of results is completely pointless. Besides, the penal provisions for "attempted offences" also require "the occurrence of a specific risk of infringement of the interests protected by criminal law", so it is not effective in aiming at the "earlier punishment". Furthermore, since the provisions for "preparation offences" also require the intention to cause a result, they cannot be used to punish researchers and engineers without even dolus eventualis. In addition, there is no social interest protected by criminal law which is equivalent to "autonomous existence of humankind" (as hereinafter described) under the current criminal law.

What is needed is, therefore, to legislate the new penal provisions, thereby disciplining particular AGI R&D activities.

2. However, the regulation by criminal legislation is undoubtedly disadvantageous to the freedom of R&D. Therefore, careful consideration for "ra-

11) The Conference toward AI Network Society, 2017: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiLs4rGxMXsAhWBsHEKHYgvDdkQFjAAegQIBhAC&url=https%3A%2F%2Fwww.soumu.go.jp%2Fmain_content%2F000507517.pdf&usg=AOvVaw11XY0RYDotru9IAMl71US6

tionality of regulation by new criminal legislation" is essential.[12)]

From the above viewpoints, I believe it is important to set the following three points as research subjects and examine them in depth: (i) the necessity and rationality of criminal regulation of AGI R&D, (ii) the embodiment of regulated conducts, and (iii) the degree of statutory penalties. In what follows, each of these is to be described in more detail.

(1) The necessity and rationality of criminal regulation of AGI R&D. This is a question of whether it is justified to regulate certain AGI R&D activities by new criminal legislation to minimise the possibility of human destruction and its enslavement by Superintelligence.

New criminal legislation is believed to be permissible only if the "principle of proportionality" is met.

It must, therefore, be questioned whether enactment of a penal provision for prohibiting particular AGI R&D activities satisfies the principle of proportionality.

Generally, the three tests are applied to confirm whether or not the principle of proportionality is satisfied. Of these, the primary issue in this research is whether or not the test of "adequacy as a means" can be passed (the other two are "necessity of the punishment" and "balance of the profits"). The "adequacy as a means" test asks whether it is the right means to achieve the regulatory objective. In order to pass this test, it is necessary to confirm the harmfulness of regulated behaviours with considerable certainty.[13)]

Hence, it must be clarified whether the measure to punish a certain range of AGI R&D activities can pass the test of the "adequacy as a means". In doing so, the following two points must be specifically examined.

(a) Is it appropriate to set the interest protected by criminal law of the "au-

12) *See*, *e.g.*, Ueda, Masaki, [2016] *Should the act really be punished?: Introduction to Constitutional Criminal Legislation*.

13) *See*, Ida, Makoto, [2018] *Criminal Law; General Part*, p. 27.

tonomous existence of humankind"?

(b) Is it possible to demonstrate with sufficient grounds the mechanism indicating that particular behaviours of AGI R&D lead to the harmfulness for the "autonomous existence of humankind"?

These issues need a thorough examination.

The purpose of certainly preventing the irreparable situation in which humankind is destructed or enslaved is, in other words, to aim at "the preservation of the status quo in which human beings exist autonomously." It is, therefore, necessary to consider the appropriateness of setting the benefit of "autonomous existence of humankind" as an interest that should be protected by the criminal law (a).

In conjunction with this issue, constructed is the question of whether any AGI R&D activities threaten the interests of autonomous existence of humankind (b). The question (b) asks the specific causation between AGI R&D activities and infringement of the interest protected by criminal law (the causal mechanism of the former reaching the latter).

If it becomes difficult to show the reasonableness of setting the interest protected by criminal law of the "autonomous existence of humankind" (=a) or difficult to gather sufficient grounds for harmfulness included in AGI R&D activities (=b), it will move on to the issues below.

Those are,

(c) To what extent do the interests protected by criminal law to be set need to be materialised and clarified?

(d) May there be an exception not to meet the principle of proportionality in criminal legislation?

Let me elaborate on the issue of (d): As mentioned above, the purpose

of criminal regulation in this research is to prevent the destruction and enslavement of humankind. If we do not introduce the criminal regulations, it could cause the enormous harm of the destruction of humankind or its enslavement. Must we stick to the principle of proportionality, even when the magnitude of such possible harm is immeasurable? Is it indeed appropriate to neglect the behaviours that could cause enormous harm?

These questions are to be constructed.[14)]

(2) The embodiment of regulated conducts (which can be set when a positive conclusion is drawn in response to the question (1) above) is an issue of how to select the behaviours that should actually be punished from AGI R&D activities. To punish all AGI R&D activities that pose a threat to the interests protected by criminal law may excessively diminish the freedom of AGI R&D activities, so it is not appropriate. Regulated conducts should be narrowed down as much as possible;[15)] however, the viewpoint of effective regulation for the protection of the interests should not be forgotten. "Regulations that are expected to have sufficient effects and are not overkill" should be aimed at. From this point of view, it is essential to select AGI R&D activities that should be subject to regulation.

In concretising the behaviours that should be subject to punishment, the following perspectives are considered significant.

---

14) The criminal regulation aimed at in this paper could be a type (or a typicl example) of so-called "early punishment". The "early punishment" has been already accepted today in areas such as the Environmental Criminal Code. In light of this, the new criminal legislation is by no means strange. About early punishment: *See*, Makoto, Ida, [2013] "On the trend toward earlier punishment", Keisei, Vol. 124, No. 11, pp. 54–55.

15) *See*, Andrew Ashworth and Lucia Zedner, [2012] "Prevention and Criminalization: Justifications and Limits", New Criminal Law Review: An International and Interdisciplinary Journal, Vol. 15, No. 4, p. 542, especially pp. 551–552.

(a) The first is the "timing of regulation". With the birth of HLMI, AGI will continue to evolve at an accelerating pace, and the difference in ability between humans and AGI will increase exponentially. After this timing, it becomes difficult for humans to control AGI. This is the stage of "difficulty in control". It is, therefore, essential to impose some regulations on researchers and engineers "before the birth of HLMI". If effective regulations can be applied at this timing, the state in which humans control AGI is maintained, and runaway of AGI can be avoided.[16)]

However, since "before the birth of HMLI" is a concept that can go back forever, it is necessary to limit the time range. The questions are to consider "to which stage of R&D the freedom is allowed" and "from which stage of R&D the regulation is imposed", and to clearly present the stage of AGI R&D to be regulated.

(b) The second important point of view is to concrete "the content of behaviours" that should be subject to regulation.

One of the threats of the singularity is that Superintelligence becomes hostile to humans. With this point of view, it is effective to make it mandatory to incorporate "a program that can prevent hostile behaviour against humans" into AGI in advance.[17)] Moreover, the situation itself in which Superintelligence goes beyond human control is also one of the severe threats. In this regard, it is useful to require the AGI system to incorporate some sort of "safety system" to avoid uncontrollability.[18)] Hence, it follows that examples of possible behaviours to be subject to punishment include "the AGI R&D

16) For the perspective on the timing of regulation: *See*, Bostrom, [2014] *supra* note 1, chapter 3, especially p. 55, chapter 5, especially pp. 85–86, chapter 6, especially pp. 92–99, and chapter 9, especially p. 129.

17) *See*, Bostrom, [2014] *supra* note 1, pp. 129–144, Reese, [2018] *supra* note 7, pp. 194–200, and Castel, J.-G. and Castel, Matthew E., [2016] *supra* note 9, pp. 11–13.

18) *See*, Bostrom, [2014] *supra* note 1, pp. 137–138.

activities that do not have an appropriate hostile-attitude-prevention program" and "the AGI R&D activities that do not have an appropriate safety system".

From this, the "omissions" such as "not incorporating an appropriate hostile-attitude-prevention program" and "not incorporating an appropriate safety system" are assumed as a criminal act ("crime by genuine omission", or "echte Unterlassungsdelikte").[19)]

However, it should be noted that it is quite controversial what the "appropriate hostile-attitude-prevention program" and "appropriate safety program" are. The scope of punishment remains ambiguous unless it is clarified what kind of programs and systems are required to be incorporated.

It has to be clarified, therefore, what kind of penal provisions should be not only effective enough in protecting the interest of the autonomous existence of humankind and but also clear to researchers and engineers.[20)]

(3) The third point is about the degree of statutory penalties.

Basically, the statutory penalties must not exceed the degree of the illegality of the act. This is a request from retribution and justice.

The question is how much penalties should be set for the new provisions. The necessity for the criminal legislation suggested in this paper is basically derived from preventive ideas. In other words, what needs to be considered

19) It would also be necessary to proceed with discussions using concepts such as "endangerment offences" and "conduct crimes" too.

20) Besides, if particular R&D activities are regulated by the criminal law, there is a possibility that some people and institutions that detest the inconvenience go underground to carry out AGI R&D. Such underground research and development can be a source of danger, so it is needed to ban such secretly undertaken AGI R&D (See, Bostrom, [2014] *supra* note 1, pp. 84–86.); it could be justified that certain reports on AGI R&D become mandatory and punishment on those who violate it are imposed.

is appropriate penal provisions to "prevent" the loss of the interest of the autonomous existence of humankind.

If our focus is on the perception that the degree of danger that threatens the interest is abstract and the possibility of causing actual harm is little, then the illegality of the act is to be small, and light statutory penalties are to be appropriate. However, in light of the weight of the interest of the autonomous existence of humankind and the importance of protecting it, the act should be strongly intimidated by punishment, that is, regulated by heavy statutory penalties.

In this way, with regard to the degree of statutory penalties to be set, the preventive idea and the retribution idea are intertwined.[21] Hence, it is needed to unravel the intricacies of the two ideas and provide guidelines on what concept should be used to determine the statutory penalties.

IV. Future development

Even if it is rational to impose certain restrictions on AGI R&D, the effects cannot be expected if there are persons or institutions that are not within reach of the restrictions. This is because unregulated persons and institutions may carry out research and development unrestrainedly, so that may lead to the birth of Superintelligence beyond human control. To achieve the goal of protecting "the autonomous existence of humankind," the rules of AGI R&D must extend throughout the world.[22]

From the above perspective, it is an urgent task to prepare an international

21) About the idea of the prevention in crimimal law: *See*, *e.g.*, Kaspar, Johannes, [2014] *Verhältnismäßigkeit und Grundrechtsschutz im Präventionsstrafrecht*. About the idea of the retribution: *See*, *e.g.*, Müller, Mathias, [2019] *Vergeltungsstrafe und Gerechtigkeitsforschung: Versuch über die zweckrationale Legitimation der tatproportionalen Strafe*.

22) *Cf.*, Castel, J.-G. and Castel, Matthew E., [2016] *supra* note 9, pp. 9–12.

research system toward the "establishment of a global governance system for AGI R&D".[23)]

In order to build such an international research system, it is necessary to raise a stronger warning to the world from an academic point of view against the easy dependence on optimism about the arrival of the singularity. By doing so, if awareness of the issue is shared and specific criminal regulations are fulfilled, the singularity can be a brilliant historical turning point of great benefit to humankind.

V. Conclusion

1. This paper has focused on the importance of preparing for the arrival of the singularity from a legal point of view.

It can be summarised as follows:

- It is necessary to be careful about the inclination toward the optimistic view of the development of AI.
- We must think of the ways to certainly prevent the "irreparable situation" of the destruction or enslavement of humankind by Superintelligence. One of the most effective ways is to apply the criminal law system to regulate particular AGI R&D.
- However, since criminal regulation has a negative side, due consideration should be given to the regulation method.
- The research topics to be studied are the necessity and rationality of criminal regulation of AGI R&D, the embodiment of regulated conducts, and the degree of statutory penalties.
- It is also an urgent task to prepare an international research system toward the establishment of a global governance system for AGI R&D.

23) *See*, Bostrom, [2014] *supra* note 1, pp. 86–87.

2. The issues raised in this paper and the possible ways to cope with them can require a way of thinking beyond the framework of conventional criminal law theory. It is, therefore, important to proceed with the examination without excluding the possibility of developing new ideas beyond the conventional framework of criminal law theories.

For example, the interest of the autonomous existence of humankind is highly unique in light of the criminal law theory so far, so a unique criminal law theory to protect it may also have to be developed. It is not a good idea to stick to traditional criminal law theories and make criminal law "useless". It is necessary to be flexible in thinking and depicting an ideal form of criminal regulation.

3. The basis of this research is the idea of "preventing harm from occurring". Therefore, we must always face the question of how preventive objectives should be pursued in criminal law. Since the harm is exceptionally severe, the concept involved in the pursuit of preventive purposes can also be significant and special.[24] On the other hand, however, the magnitude of the possibility of the harm occurring can be estimated differently depending on the researcher. Therefore, important is also an attitude of calmly appraising how realistic the occurrence of this severe harm should be.

(Completed on 29th Oct. 2020)

* This paper is a part of the research results during the 2020 overseas research period of the author, which was supported by Chou University, Japan.

24) It may also be necessary to consider the treatment of conducts that have the characteristic of "endangerment". *See*, R. A. Duff, [2005] "Criminalizing Endangerment", Louisiana Law Review, Vol. 65, No. 3, p. 941, especially p. 964.