Utilization of AI and Law – Possible Legal Framework on AI Research and Technology Development –

Mayu Terada

I. Introduction

Artificial Intelligence (AI) will change our lives dramatically. At the same time, the emerging networked society supported by such AI comes with risks. There are many possible crimes for which AI may be used, such as malware abuse, as a tool for terrorism, or as an autonomous weapon system. There is also the risk that AI network systems could be used for other inappropriate applications. AI may infringe the rights and interests of consumers because it may surmise people's intentions, health, or future actions. Furthermore, AI network systems can become opaque, which makes the control of personal information more difficult. On that note, AI networks can also be uncontrollable regarding collection and application of personal information. There are many other risks to be discussed, but it is clear that there are many risks associated with an AI networked society. In this paper, I will discuss the possibility of a legal framework to prevent such risks. In particular, I would like to discuss the possibility of setting up guidelines for such principles on AI development which could be agreed upon internationally and could be used or applied as a standard. As a concrete example, we have the AI development research guidelines which have been discussed in Japan since 2015. The Japanese government introduced eight principles as AI R&D guidelines. The guidelines have already

---


been discussed at the official meeting of the OECD in the fall of 2016 and are currently under consideration.³ We still do not know how many countries would agree on such guidelines, but they could certainly form a soft legal framework to curb the risks associated with an AI networked society. In conclusion, setting up internationalized guidelines could be one of the possible solutions to prevent the risks associated with an AI networked society.

In this paper, possible risks will be assessed, and based on that, potential regulations to mitigate these risks will be discussed. This paper will touch upon many topics. However, the focus will be on outlining potential regulations.

1. A Future with AI Networks

There is not one specific way to define Artificial Intelligence. There are different approaches which focus on the various aspects of AI.⁴ In this paper, “AI” refers to a concept that collectively includes AI software and AI systems. This definition of AI is assumed to apply mainly to AI in the narrow sense and refers to AI which is already being put into practical application. However, there are many other types of AI.

The market for AI has been analyzed and predicted to expand by various organizations. For example, a forecast by the Bank of America in 2015 said that utilization of AI in the field of medicine and finance would grow dramatically.⁵ Other research in 2015 by Tractica predicted that revenue related to AI would grow rapidly by 2024.⁶ It is predicted that AI and Robots will drastically change the industrial structure itself. In January 2016, the fourth industrial revolution in relation to AI was mentioned at the WEF conference held in Davos, Switzerland and it was said to make a great impact on many sectors.⁷

---

⁶ Tractica’s research about Artificial Intelligence Revenue by Region, World Markets: 2016-2024.
Whilst AI can be something which does not co-operate with other systems and can be used on its own, AI that will connect to information and communication network systems and co-operate with other networks is a much more likely use of AI. There are more and more devices such as Internet of Things machines that connect to the internet, and AI can be utilized in each of these machines. In addition, AI can connect to information technology such as the internet and co-operate with other systems to be used in a variety of sectors of society.\(^7\)

Indeed, it is generally thought that AI networking will advance increasingly from now on. AI will yield us more benefits and influence when connected to a network (such as the internet) and used in this way compared to the utilization of AI on its own.\(^8\) Thus, we should think of this AI phenomena and matters surrounding AI as a wider problem concerning the development of AI networking.\(^9\)

AI is expected to bring us human beings better, more convenient lives and to have many other benefits. However, at the same time, there are also concerns about AI's uncontrollability, that is, the fact that AI may also make decisions and actions that cannot be predicted by AI developers.\(^10\)

This matter will also be discussed later, but the process of AI's judgement has been an uncertain factor and might not be clarified in the future, including the reasons why such judgement or such behavior was done by

---

7 See Schwab, The Fourth Industrial Revolution: what it means, how to respond, 2016, https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/. Schwab takes the view that “Like the revolutions that preceded it, the Fourth Industrial Revolution has the potential to raise global income levels and improve the quality of life for populations around the world. The First Industrial Revolution used water and steam power to mechanize production. The Second used electric power to create mass production. The Third used electronics and information technology to automate production. Now a Fourth Industrial Revolution is building on the Third, the digital revolution that has been occurring since the middle of the last century. It is characterized by a fusion of technologies”.


9 Telecommunications Research Laboratory of the Ministry of Internal Affairs of Japan (note 1).

10 Cf. note 9.

AI. At the same time, AI's development is expected to occur without borders and it is also recommended that it should be systematized. As pointed out earlier, in the future, it is expected that AI networks will be created holistically and connected without borders.

We already live in a society in which AI is embedded or inevitably incorporated. However, there is still a great deal of uncertainty, and therefore, various research and development is being undertaken. It is a natural thing that various guidelines on AI are being considered since we currently find ourselves in the sprouting stage of AI networking.

2. Current Situation

The current situation of AI and networks can be best illustrated with actual examples. Here we will look at three examples: a) IoT devices, b) connected cars and c) drones.

a) Example: Internet of Things Devices

It is very easy for us to think of Internet of Things (IoT) devices that have become part of our lives. IoT means that a wide variety of “things”, or objects, are connected to the internet and that, by being interconnected, those objects communicating or interacting with each other are capable of generating something of value. Since the speed of transmission on networks has progressed, all kinds of “things” can now be connected to internet. IoT encompasses two kinds of communication, firstly, connecting things to things and, secondly, connecting machines to machines (Machine-to-Machine, M2M).

13 Above note 9.
17 See also Commission of the European Communities (note 16).
IoT generally means that things are connected to things and each “thing” will connect to another “thing”. However, there are several problems. For example, there is the possibility of direct data leakage of collected personal data by companies that produce IoT devices.\footnote{Concerns about data leakages is shown from the industry sector, see https://ec.europa.eu/newsroom/cf/dae/document.cfm?doc_id=1746.} Thus, various legal questions regarding IoT related issues are now under consideration.\footnote{It is being suggested that data generated by IoT devices should be treated as personal data, see Buttarelli, “Internet of Things: Ubiquitous Monitoring in Space and Time”, speech delivered at the European Privacy and Data Protection Commissioners’ Conference, Prague, April 2010, https://secure.edps.europa.eu/EDPSWEB/webdav/shared/Documents/EDPS/Publications/Speeches/2010/10-04-29_Speech_Internet_Things_EN.pdf. See also Article 29 Data Protection Working Party, Opinion 8/2014 on the on Recent Developments on the Internet of Things, WP 223, 2014.}

IoT devices are now used in many different objects of our lives, including everyday things such as refrigerators and vacuum cleaners. Our lives have become more and more convenient, however, in reality, these IoT devices generate enormous amounts of “live” data.\footnote{Mauritius Declaration on the Internet of Things, adopted at the 36th International Conference of Data Protection and Privacy Commissioners, Balaclava, Mauritius, 14 October 2014, http://www.privacyconference2014.org/media/16596/Mauritius-Declaration.pdf.} It is said that these IoT devices will bring about the “fourth industrial revolution”, which will be accelerated by the integration of IoT devices and AI. It is very easy to assume that the number of IoT products will increase. However, since all information can be connected, as a matter of course, there is a risk of information leakage. Thus, a system design that can avoid such risks will be required.\footnote{Discussions on data portability and privacy by design could be beneficial for a solution, see European Commission, Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data, COM(2012) 11 final.}
b) Example: Connected Cars

Autonomous vehicles using AI are about to become reality. According to various press announcements, Google, Audi, BMW and many other companies are indeed creating autonomous vehicles.\footnote{See https://www.google.com/selfdrivingcar, https://www.ft.com/content/973f2f5c-6649-11e7-8526-7b38dcaef614 and https://www.bmw.com/en/automotive-life/autonomous-driving.html.}

Under the progress of extensive networking of many systems, it is still under consideration how AI will be used in autonomous vehicles.\footnote{See \textit{Ludwig} (President of the AI Group at Xevo), AI is already driving the future of connected cars, 2017, https://venturebeat.com/2017/03/22/why-ai-will-drive-the-future-of-connected-cars-starting-now; \textit{Rashid}, How AI Pioneers Will Affect The Car Industry, And Why It's A Good Thing, 2017, https://www.forbes.com/sites/brianrashid/2017/05/16/how-ai-pioneers-will-affect-the-car-industry-and-why-its-a-good-thing/#7576cc4257c2.} However, various vehicles with different AI functions will certainly emerge, mixing the technology of autonomous systems and deep learning.\footnote{In Germany, progress is made in legal terms after testing connected cars on the autobahn: \textit{BMVI}, Bundesminister zieht Zwischenbilanz zum Digitalen Testfeld Autobahn, 2016, http://www.bmvi.de/SharedDocs/DE/Pressemeldungen/2016/1066-dobrindt-digitales-testfeld-autobahn.html; see also \textit{Gasser/Arzt/Ayoubi/Bartels/Eier/Flemisch/Häcker/Hesse/Huber/Lotz/Maurer/Ruth-Schumacher/Schwarz/Vogt}, Rechtsfolgen zunehmender Fahrzeugautonomisierung, BASt-Bericht F 83, 2012, 11.} We can easily predict a future where cars using AI will become indispensable.\footnote{See https://www.statista.com/outlook/320/137/connected-car/germany.}

c) Example: Drones

There have been several news reports indicating that drones may be used for commercial purposes, especially for delivery. Such drones (unmanned aerial vehicles, or UAVs) also use AI. It is most likely that drones controlling their movements autonomously through AI will become more mainstream than those which would have to be operated by humans. The drones that are typically expected for commercial use will fly according to the network they are embedded in, which provides a variety of information. Therefore, the utilization of drones shall also be included in the AI networking society.
On that note, the progress of AI networking will provide various services on the premise that AI networks and network services can be utilized in various parts of society. It is expected that AI will become one of society's indispensable instruments. Under such circumstances, it is expected that people will eventually be integrated into the AI networked society regardless of whether one likes it or not.

II. Risks of AI Networks and Discussion

There are, of course, many risks and issues associated with AI, but this paper will only review some of these, including cyber-attacks, job security, AI in the legal sector, strong AI, legal personality of AI and unintended behavior of AI.

1. Cyber-Attacks

If AI is connected to networks or the internet, there is no doubt that various risks will arise. The possibility of cyber-attacks is one of these.26

AI can be designed to co-operate with various systems, and, as stated above, AI is often intended to be utilized in networks. The resulting requirement for interoperability of AI could lead to the development of standards for data formats, interfaces, and protocols. Even though these standard's intended purpose would be to support the progress of AI and AI networks, any vulnerability in those standards could, and most likely will, also be exploited for cyber-attacks. Due to the nature of interconnected AI and AI networks, the severity of harm that results from a vulnerable standard is even further increased.27

---


It should be noted that standardization and security are also in a trade-off relationship, and in this regard, there is a possibility of hacking, cyber-attacks, surreptitious acts, or any other attack that no one could ever imagine. Thus, measures to mitigate such risks must be taken.

Also, as a result of cyber-attacks, there is a high probability that data collected by AI will be leaked. This is related to the fact that, as AI networking keeps progressing, large amounts of information will be transmitted over networks. Thus, AI may leak data due to unexpected events or system problems. In addition, there is a possibility that AI might do something, which the system developers did not expect. Therefore, in addition to cyber-attacks, there is also a risk of data leakages by the AI itself.

In conclusion, AI is good at analyzing large amounts of information on behalf of individuals and judging its appropriateness, which seems to be convenient. However, the information itself can be exposed to cyber-attacks. In addition, there is a possibility of the leakage of large amounts of data, and if such a mechanism of AI networking is attacked, the damage will be enormous.

2. Job Security

In addition, it is pointed out that many job opportunities may be lost by AI. In other words, AI will replace human activities. According to one forecast, 47 percent of the current employment opportunities will be lost to AI. Thus, a question here is whether to consider a basic income system as a result of utilizing AI and living in an AI networked society.

It was a huge shock for many people in 2013 that many organizations and research sectors pointed out that jobs of human beings will be replaced by machines utilizing AI in the near future. However, people also think that new technologies related to AI will create new jobs for humans.

29 Note 28.
32 Note 31.
and the impact of the expected work substitution by AI is not currently being considered particularly serious, because we are also expecting a workforce shortage as a result of the ageing society. However, a more labor-saving situation might progress and the world where human beings do not have to work or those who want to work cannot work could become a reality because of AI. In such a scenario, if wealth is mainly created by AI robots and humans can no longer generate an income through employment, this leads to the question of wealth distribution and also how to collect taxes to secure government funding, since taxable wages are vanishing. To consider a fair society, where AI networking might generate most of the revenue of a state, a basic income system can be a worthwhile solution to be considered for discussion. A basic income system may be the only solution for those who are not able to work.

3. AI in the Legal Sector

The legal sector is no exception. Indeed, there is a high probability that AI will take on a wide variety of tasks in the legal sector.

It is necessary to consider what the term “judgement” actually means. For example, AI lawyer Ross was hired by a law firm in the United States of America and it is actually working well. It is expected that Ross will eventually even be able to apply rules that require a lawyer's moral judgement. Then, there is also a possibility of AI judges. This raises the question as to whether we can accept an AI's judgement, a question we do not have a clear answer to yet. However, there are many issues related to

---

33 There are many countries or areas that have fewer people than before living in it, including Japan and some regions of the EU. See González-Leonardo/López-Gay/Recaño, Urban Depopulation and Loss of Human Capital: An Emerging Phenomenon in the European Union, https://population-europe.eu/policy-insights/urban-depopulation-and-loss-human-capital-emerging-phenomenon-european-union; Vlandas, Grey power and the economy: Ageing and inflation across advanced economies, Comparative Political Studies. 51.4 (2018), 514.


this point. We might even have to clarify what exactly is meant by “judgment” and “judicial system”. One scenario that we can easily imagine is that before making any judgement, an “AI judge” must study the case or preconditions, otherwise the process itself could be a problem.\(^\text{37}\)

As a prerequisite, for example, to judge any personal cases, it will probably be necessary to include skin color, religion, gender, medical condition and many other sensitive elements. It is impossible to predict how the deep learning process of AI will work, and it is very difficult to see how decisions avoiding racial, gender and every other kind of discrimination can be made.

The troublesome thing about AI is that it appears to be difficult to comprehend its decision-making process. AI can learn rapidly by itself, and hence it can be difficult to explain on which basis the decision has been made. In such a case, how can a human being be judged by an “AI judge”? Judges also make mistakes. The mistakes by the judges are reviewable in many countries, either by appealing to a higher Court or in a similar way. If one thinks that the “AI judge” has made a mistake, how will this process of review take place? Will another “AI judge” decide? Will humans be able to overturn the decision of the “AI judge”?

If the decision-making process of AI is not disclosed (or it cannot be unveiled in any way), the decision becomes a black box. The person who receives the judgement will not be able to know whether the decision-making process was flawed.

It is pointed out that people tend to trust the judgment of machines because of their technicality. This phenomenon is called “automation bias”,\(^\text{38}\) which means that people have a cognitive tendency to have excessive trust in an automated judgment made by a computer. Many people will keep the judgement in mind, take it as it is presented to them and will not think about the reason for the decision, which looks like it has been scientifically derived by referencing a large amount of data. People trust the mechanical

\(^{37}\) There is an argument that legal reasoning is an inherently parallel process and the answer to one question might change with questions that are subsequently asked, see Aikenhead, The Uses and Abuses of Neural Networks in Law, Santa Clara Computer & High Tech. L.J. 12 (1996), 31, 56.

decision because the AI has studied tremendous amounts of data and people believe that its decision must be backed up by science.\textsuperscript{39}

It is fully conceivable that AI will help judges in reaching their verdict.\textsuperscript{40}

As we are using various databases for trial, it can be said that the outcome of a trial is sometimes already influenced by AI.

However, leaving everything to an “AI judge” also means that humans will not make the final judgement. Regarding this matter, there will be discussions on whether human beings can accept such judgments in the future, and we should consider the foundations of the court system. Then, we will have to continue the discussion about what human beings can have trust in in the AI network era.\textsuperscript{41}

4. Strong AI – Singularity

It is also important to mention the problem of “strong AI”.\textsuperscript{42}

As a result of the evolution of AI, researchers have continued to develop such AI – where the AI itself thinks and has a consciousness – and there could be a risk that such a strong AI could take control of the human world.\textsuperscript{43}

Some people say that strong AI cannot be developed as such. However, research on AI is progressing day by day, and such a world must also be considered as one of the risks. Indeed, rapidly developing AI in this third AI boom is a narrow type of AI that assumes specific uses and functions, and it is generally not about a type of AI that is similar to humans or other related living objects that enjoy autonomy and freedom (which would be an Artificial General Intelligence (AGI)). However, a report by the White House points out that more than half of AI researchers think that it is possible to develop an AGI which can reach and exceed human levels of intelligence by 2040. It is in-

\textsuperscript{42} The argument of “strong AI” and “weak AI” is a discussion initiated by philosopher John Rogers Searle: \textit{Searle}, Minds, brains, and programs, Behavioral and Brain Sciences 3.3 (1980), 417.
\textsuperscript{43} \textit{Prof. Bostrom} says that machines that are capable of independent initiative and of making their own plans are perhaps more appropriately viewed as persons than machines: \textit{Bostrom}, When Machines Outsmart Humans, Futures 35.7 (2003), 759, 763; \textit{Bostrom/Cirkovic}, Global Catastrophic Risks, Oxford 2008; \textit{Kurzweil}, The Singularity Is Near: When Humans Transcend Biology, New York 2005, 135 f.
deed likely that, regardless of if the targeted time frame of the year 2040 holds true, at some time, a serious discussion will take place on whether AI should be recognized as a person.

5. Legal Personality of AI?

It is also currently under discussion whether AI should be given a legal personality. In a modern society, people and things are classified according to the definition of legal texts. For example, a corporation is given juridical personality by civil law. However, a corporation will not be given the status of a legal entity if it does not satisfy the requirements as stated by law.

Whether or not to grant AI the status of a legal entity is an important issue and there have already been discussions on this subject. Under the current legal system, we assume that the person who is responsible is a natural or legal person. Only those persons are qualified to become the subject of rights and obligations and to dispose thereof.

Will AI and robots be accepted and included in our society as an entity with a legal personality? It is necessary to consider the responsibility of humans that results from the creation and use of AI. This matter should be examined, and solutions should be developed.

6. Unintended Behavior of AI

It is considered that there is a risk that AI may act in a way unexpectedly different from the programmer's intent because the deep learning process used by AI cannot be fully predicted, not even by the original programmer. It has also been reported that Google is developing a kill switch to stop powerful AI systems that went out of control.

It is necessary to think about the criteria by which AI is deemed to be out of control. There is a report published by the European Parliament with recommendations on civil law rules on robotics, and the report introduces the “Code of Ethical Conduct for Robotics Engineers”. In the report,

it is recommended for designers of smart robots to implement kill switch-
es, saying “You should integrate obvious opt-out mechanisms (kill switch-
es) that should be consistent with reasonable design objectives”.47

This recommendation is a manifestation of some of the risks related to an AI networked society. There are also ways to avoid the risks with technology design etc. but it will also be important to consider how to develop AI because there is no doubt that AI will change our future and that AI will be applied everywhere soon. This is why the discussion on governance of AI networks in the following section is important.

III. Governance of AI Networks

Networking of AI has just begun and the discussion about it is still in its early stages. Even though research and development by AI is rapidly pro-
gressing,48 the issue of a governance system for AI networks still needs to be thoroughly investigated. We should consider and study the need and foundations of AI network governance.

For example, it may be necessary to continue discussions through meet-
ings with various stakeholders. However, proceeding with the discussion may be very difficult. There are many uncertainties as to whether an agree-
ment on governance of AI will be reached internationally since the cir-
cumstances and intentions of each country are different.

It is very difficult to think that there should be legal restriction of this area. It is difficult to make a general judgement on which aspects should be regulated and to what extent the regulation is needed. There is a paper that points out that imposing utilitarian regulation is overall counterpro-
ductive for the aim of improving the safety of networked AI systems.49 When there is too much deterrence, e.g. through extended liability, it will most likely result in a loss to society because no one would be willing to take risks. Services where someone must take risks will be greatly limited

---

47 European Parliament, Civil Law Rules on Robotics, European Parliament resolu-
48 See, for example, note 14.
49 Bonnefon/Shariff/Rahwan, The Social Dilemma of Autonomous Vehicles, SCIE-
NCE 352.6293 (2016), 1573.
or not be provided at all. Thus, thinking of risk too much is likely to bring “distortion”\textsuperscript{50} to society.

However, as mentioned earlier, it seems necessary to constantly engage in open discussions on a global scale because it is necessary to consider various situations regarding regulation on research and development of AI and AI networks.\textsuperscript{51}

On this point, in Japan, the Ministry of Internal Affairs and Communications has been holding meetings since 2016 and has been discussing this matter. Regarding the governance of AI networks, it is recommended that, given the complexity of the matter and to ensure a balanced consideration of the related values and interests, open discussions involving various stakeholders should take place to shape the global framework of regulations on AI network systems.

1. Discussions on Governance and Creating Guidelines

In this section, several proposals for creating guidelines in the United States and the EU will be reviewed.

\textit{a) Discussion on AI in the United States}

Since May 2016, the White House has been publicly preparing for and studying AI, and workshops on AI have been hosted throughout the United States.\textsuperscript{52} Then, based on these workshops and comments, in October 2016, the report “Preparing for the Future of Artificial Intelligence” was published.\textsuperscript{53} The report covers a wide range of topics, including regulation, research and workforce, economy and employment, fairness, safety and

\begin{itemize}
  \item The problem of “distortion” is pointed out in American Tort Law.
  \item The 63rd principle in the Civil Law Rules on Robotics of the \textit{European Parliament} (note 47) also “Strongly encourages international cooperation in the scrutiny of societal, ethical and legal challenges and thereafter setting regulatory standards under the auspices of the United Nations”.
  \item \textit{National Science and Technology Council (NSTC)}, Preparing for the Future of Artificial Intelligence, 2016, 12, https://obamawhitehouse.archives.gov/sites/default/fil
governance, global aspects and security issues with profound considerations for managing the risks on each matter. The report says that “AI holds the potential to be a major driver of economic growth and social progress, if industry, civil society, government, and the public work together to support development of the technology with thoughtful attention to its potential and to managing its risks”.\(^\text{54}\)

In the conclusion of the report, there is a warning for practitioners, saying that they must ensure that “AI-enabled systems are governable; that they are open, transparent, and understandable; that they can work effectively with people; and that their operation will remain consistent with human values and aspirations”.\(^\text{55}\)

In October 2016, the National Science and Technology Council (NSTC) established a strategic R&D plan for federally funded AI related research in which the ultimate goal was set to produce new AI knowledge and technologies that provide a range of positive benefits for society, while minimizing the negative impacts.\(^\text{56}\) In addition, a vision for advancing the national priorities with AI is described.\(^\text{57}\)

\(b\) Discussion in the EU

In Europe, the European Parliament is discussing AI and robots with the aim of formulating guidelines or regulation. The Committee on Legal Affairs of the European Parliament held a hearing on legal and ethical issues of robots and AI in April 2016 and published a draft report in May 2016. In addition, in October of the same year, related workshops were held. Based on these considerations, the European Parliament adopted the recommendations to the European Commission on civil law rules on robotics in February 2017.\(^\text{58}\)

\(\text{https://doi.org/10.5771/9783748910770-565}\)

\(\text{Generiert durch IP '54.70.40.11', am 18.07.2021, 01:17:29.}\)

\(\text{Das Erstellen und Weitergeben von Kopien dieses PDFs ist nicht zulässig.}\)
In the report, the EU fully recognized that it can play an essential role in formulating ethical principles to be respected in the development, design and use of robots and AIs, stating that “whereas the Union could play an essential role in establishing basic ethical principles to be respected in the development, programming and use of robots and AI and in the incorporation of such principles into Union regulations and codes of conduct, with the aim of shaping the technological revolution so that it serves humanity and so that the benefits of advanced robotics and AI are broadly shared, while as far as possible avoiding potential pitfalls”.\(^{59}\)

In addition, the report focuses on examining civil law issues concerning robots and AI, and at the same time, strongly recommends the establishment of an European Agency for robotics and artificial intelligence. Furthermore, registration of smart robots, strict liability for damages and intellectual property proposals have also been made as the report calls on the European Commission to formulate EU law stipulating compliance with property rights, interoperability, etc. At the same time, recommendations have also been made on the formulation of an ethics code of conduct for robot developers, which should include respect for basic rights, adherence to the precautionary principle, inclusiveness and accountability, safety, reversibility, maximization of privacy and benefits and minimization of harm.\(^{60}\)

\(c\) Other Guidelines

There are several non-governmental organizations which propose the establishment of guidelines on AI.

\(aa\) The Tenets of “Partnership on AI”

An example of the efforts made by the industry is the “Partnership on AI”. This partnership was established “to study and formulate best practices on AI technologies, to advance the public's understanding of AI, and to serve

\(^{59}\) European Parliament (note 47), General principles – V.

\(^{60}\) European Parliament (note 47), Introduction – O: “whereas the developments in robotics and AI can and should be designed in such a way that they preserve the dignity, autonomy and self-determination of the individual, especially in the fields of human care and companionship, and in the context of medical appliances, ‘repairing’ or enhancing human beings”.

580
as an open platform for discussion and engagement about AI and its influences on people and society”. It was conceptualized by major network entrepreneurs from the US as an open platform for studying and formulating best practices regarding AI technology, improving the public's understanding of AI, and discussing AI and its social impact. The tenets explain the organization's position on AI, stating many things, including that they intended to ensure that AI technologies benefit and empower as many people as possible, that they would educate and listen to the public and actively engage stakeholders to seek their feedback and that they would strive to create a culture of cooperation, trust, and openness among AI scientists and engineers.

bb) The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

As an initiative of European and American academic societies, the IEEE started the “Global Initiative on Ethics of Autonomous and Intelligent Systems”, which published the second version of its report titled “Ethically Aligned Design” in December 2017. The report presents topics for discussion to encourage engineers to think about ways to design AI in accordance with human moral values and ethical principles. In addition to the topics on economics, humanitarian and legal issues related to AI, issues specific to general purpose AI and lethal autonomous weapon systems are discussed in the report.

cc) The Asilomar AI Principles of the “Future of Life Institute”

Another set of guidelines is coming from the nonprofit organization “Future of Life Institute” (FLI) established by entrepreneurs and researchers. The FLI held a meeting on research and development of AI in Asilomar, USA in January 2017, which resulted in the Asilomar AI principles pub-
lished in February 2017. These principles address a variety of issues: the safety and security of AI systems; the transparency of AI systems in the event of an accident; the need for a satisfactory explanation auditable by a competent human authority regarding any involvement by an autonomous system in judicial decision-making; failure transparency when AI system causes harm; the compatibility of AI systems with ideals of human dignity, rights, freedoms, and cultural diversity etc.

In addition, the principles also touch on longer-term issues like the possibility of highly autonomous AI systems which may even recursively self-improve or self-replicate, ethical questions around superintelligence, and safety issues.

There are many discussions and opinions on regulations on AI and AI networks. There is a strong opinion that we should create laws and governmental bodies that regulate AI as soon as possible. However, there are also people who oppose the regulation of AI.

\[d\) Analysis\]

As explained above, the United States are seeking a soft law approach towards AI, whereas the EU is seeking more of a legislative approach to control AI and is considering to establish an agency to enforce the legal framework for AI. We are still in the early stages of deciding what kind of regulation is suitable, but it is worth starting to think about the approaches towards AI and its governance.

Introducing regulation might reduce the risk of AI being abused, but it might have the side-effect of suppressing the advancement of technology. It is still very difficult to define AI and its surrounding situation, thus, once regulation is put in place, there is a risk that the interpretation of such regulation would need to be stretched. Since many things are still uncertain, it may be better to first seek to introduce soft law like guidelines, because it will still induce the development of carefully designed AI.

---

65 https://futureoflife.org/ai-principles/.

66 Note 65.

67 See Scherer, Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies, Harv. J. L. & Tech. 29 (2016), 353, 393 ff. In this article, as a regulatory proposal, the “Artificial Intelligence Development Act” and a corresponding agency is proposed.

The Japanese government is following this approach and is seeking to implement governance based on guidelines, because setting up a strict regulation might hinder the development and research of AI.

2. Guidelines for an AI Networking Society

In this section, a proposal for guidelines for an AI networking society by the Japanese Government is presented. As for the research and development of AI, the Ministry of Internal Affairs and Communications of Japan issued “The Conference on Networking among AIs Report (2016): Impacts and Risks of AI Networking – Issues for the Realisation of Wisdom Network Society (WINS)”.69 This report was the first systematic review of AI networking issues in Japan. Referring to the OECD guidelines governing privacy and security, Japan thought that it was necessary to begin discussions and considerations towards formulating international guidelines and principles to govern the research and development of AI.

In late July of 2017, the “AI Network Society Promotion Conference” of the Information and Communication Policy Research Institute of the Ministry of Internal Affairs and Communications announced the “AI Development Guidelines for International Discussion”. This is a suggestion of guidelines originating in Japan aimed at being adopted in the international community as typified by the OECD.

For the actual guidelines, as shown below, there are nine principles proposed and especially the second principle of transparency was considered important at the OECD conference.

1. Principle of collaboration – Developers should pay attention to the interconnectivity and interoperability of AI systems.
   [Principles mainly concerning mitigation of risks associated with AI systems]
   
---

69 Telecommunications Research Laboratory of the Ministry of Internal Affairs of Japan (note 1).
2. Principle of transparency – Developers should pay attention to the verifiability of inputs/outputs of AI systems and the explainability of their judgments.
3. Principle of controllability – Developers should pay attention to the controllability of AI systems.
4. Principle of safety – Developers should take it into consideration that AI systems will not harm the life, body, or property of the users or third parties through actuators or other devices.
5. Principle of security – Developers should pay attention to the security of AI systems.
6. Principle of privacy – Developers should take it into consideration that AI systems will not infringe the privacy of the users or third parties.
7. Principle of ethics – Developers should respect human dignity and individual autonomy in R&D of AI systems.

[Principles mainly concerning improvements in acceptance by users et al.]
8. Principle of user assistance – Developers should take it into consideration that AI systems will support the users and make it possible to give them opportunities for choice in appropriate manners.
9. Principle of accountability – Developers should make efforts to fulfill their accountability to stakeholders including AI systems' users.

The actual guidelines themselves are based on five basic philosophies. These are: 1) “human-centered society”, which means that human dignity and individual autonomy are respected; 2) the guideline will be shared as non-binding soft law; 3) ensuring an appropriate balance between benefits and risks of AI; 4) ensuring technical neutrality and not imposing excessive burdens on developers; and 5) to constantly review the guidelines and readily revise them as necessary through international discussions with related stakeholders.70

In addition, the guidelines contain an attachment describing roles that are expected to be taken by related stakeholders of relevant industries, academia, and governments. It states that each country’s government and international organizations are expected to make efforts to promote the dialogue among stakeholders and that stakeholders are also expected to make efforts to participate in the dialogue and share best practices conforming to the guidelines. It also encourages standardization bodies and

other related entities to prepare and release recommended models, to advocate for assistance to AI developer communities from each country's government, and to actively promote policies to support the R&D of AI.

As mentioned above, the principle of transparency would be the most important issue to be considered in many other countries. The transparency principle here does not mean that developers will be required to disclose algorithms, source codes, or learning data, and privacy and trade secrets are also considered. It is desirable that developers pay attention to the verifiability of the inputs and outputs of AI systems and the explainability of the judgement of AI systems, because AI systems will affect our lives in areas where we should enjoy freedom, privacy and other rights.

Controllability and safety issues are also important. With deep learning technology, it is not possible to predict how AI will evolve, but with these guidelines, developers will watch for the architecture of the learning process not to harm humans.

Overall, the nine AI R&D guidelines proposed by the Japanese government take a soft law approach and they aim to bring benefits to all stakeholders. It will also be the basis for discussion in comparison with other guidelines proposed by various organizations and governments. As far as the guidelines are creating better conditions for a better discussion and promoting the development of AI, it can be said that they fulfilled the role as a sort of protective umbrella against the risks for humans associated with AI, and this is the main objective of creating such guidelines.

IV. Conclusion

In effect, the discussion about the current AI is undertaken to ensure that human beings control AI and to make society more convenient for human beings. However, the fact that AI gathers all kinds of information about human beings and learns by itself suggests that AI can become indistinguishably similar to human beings and AI as data will likely to be similar to humans. Every piece of human information is gathered and analyzed. This can be said to be a risk, but we are trying to mitigate the risk by attempting to establish guidelines for AI R&D.

In the future, humans will live in AI networked societies that will inevitably involve AI and it will be difficult for human beings to compre-

71 The Conference toward AI network Society (note 70), 8.
72 The Conference toward AI network Society (note 70), 3.
hend and explain the reasoning of AI in order to reassure concerned parties. For that reason, it is necessary to give AI the ability to explain situations precisely and we should consider ways to develop such algorithms to enable AI to explain the process. To adequately implement AI, a sufficient degree of transparency will be necessary, which will require developers to provide information on technical characteristics so that product developers and users can select and utilize the appropriate technology.

Furthermore, it is possible that an AI automobile or an AI robot which performs surgery, both scenarios being very likely in the future, may cause a fatal accident and may end a person’s life. If such accidents cannot be completely prevented, then we should at least be able to retrospectively understand how and why they happened to be able to take appropriate measures. Therefore, one possibility is to simultaneously develop algorithms with the ability to elucidate why an AI solved the respective problem in the particular way. Humans are always seeking explanations and an answer to the question of “why” and in this regard, the principle of transparency in the Japanese government’s proposed guidelines should play an important role in promoting such an architecture.

For developing such algorithms for AI, an appropriate collaboration of developers and lawyers who consider the current and future regulatory framework is needed because making a legal framework without considering the actual technical problems would be useless.

However, all technical solutions for the risks associated to AI and AI networked society and all possible frameworks will also require to consider what humans are and the concept of responsibility, and to redesign human society. Moreover, there are still uncertainties and unknown issues with regards to AI and an AI networking society, which may require flexible changes to the regulatory framework of AI.

We should also continuously discuss how we can make AI and an AI networking society transparent, controllable, and acceptable to humans collaborating with the advancing technology. For now, since this area is still emerging, a soft law approach, such as guidelines, is a valid and feasible way towards a framework to govern the risks of AI.

75 The Conference toward AI network Society (note 70), 8.