

Social Principles of Human-centric AI (Draft)

unofficial translation

TABLE OF CONTENTS

1. Introduction	1
2. Fundamental philosophy	3
3. AI Ready Society: Essential social revolution to achieve Society 5.0	5
4. Social Principles of Human-centric AI	6
4. 1. AI Social Principles	6
4. 2. Development and utilization principles of AI	9
5 Conclusions	10

This document is a draft of the social principles of human-centric AI produced by the council for social principles of human-centric AI. This principle is to be drawn up in March 2019 after soliciting domestic and international opinions.

1 Introduction

Our modern society is facing with the problems which endangered the survival of the human race, such as global environmental issues, widening disparity, and resource depletion. AI is considered to be the key technology to investigate these problems, to solve the 17 goals listed in the SDGs (Sustainable Development Goals), and to build a sustainable world.

In particular, Japan may be likened to a top runner of social problems, including declining birth rate and aging population, labor shortages, depopulation, and increased fiscal expenditure.

Establishing Society 5.0¹ to solve the social issues along with economic development, Japan aims to revitalize its society and economy, to be an internationally attractive society, and to contribute to the SDGs on a global scale, by the utilization of AI.

Similar to many other science and technology, while AI will bring great benefit to society, we must be cautious to develop and implement it appropriately, since it will have a large social impact on society. To use AI effectively to contribute to the society and to avoid or reduce the negative effects beforehand, we should promote, along with the research and development of the technologies related to AI, the transformation into the "AI -Ready Society" in which we can effectively and safely utilize AI, by redesigning society in all aspects including Japan's social system, industry structure, innovation system, governance, and its citizens' character.

In this document, we regard the term "AI" as "highly complex information systems"², instead of particular technologies, to discuss the effects of AI on society widely. We will present the social principles of AI and enumerate issues to be considered in AI research, development and implementation. To improve the coming Society 5.0, it is essential that relevant stakeholders interact and cooperate each other.

This document's structure is shown in Figure 1.

Society 5.0 is the shape of the future society that Japan aspires to, following the Information Society (Society 4.0). The core image of the Society 5.0 is a sustainable human-centric society in which AI, IoT (Internet of Things), robotics and other cutting-edge technologies are implemented to create unprecedented value, and diverse people can realize their diverse well-being with respect to the well-being of others.

There are various ideas on the definition of "AI (Artificial Intelligence)" between researchers, and there is no clear definition of AI at present. In addition, various technologies commonly referred to as "AI" are rarely used by a single unit, and they are commonly used as part of the information system. In this document, on the idea that these principles should be commonly taken into account in advanced and complex information systems, "AI" in this principle means "highly complex information systems."

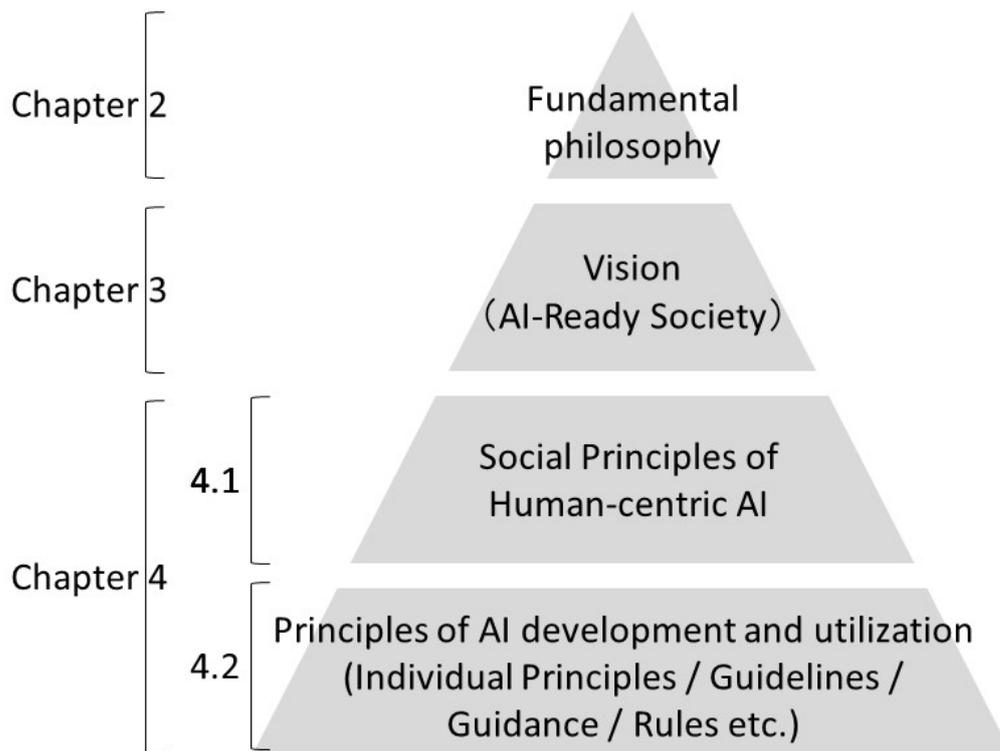


Figure 1: The Structure of This Document

2 Fundamental Philosophy

AI is expected to contribute to the realization of Society 5.0 significantly. We consider it to be important not only to deliver many benefits of AI to people and society, derived from its efficiency and convenience, but to utilize AI as public goods of human beings as a whole and to ensure the global sustainability described in SDGs, through qualitative changes generated by genuine innovations toward the ideal society.

We consider it to be essential to respect the following three values as philosophy to be pursued to be realized in coming society.

(1) Dignity

We should not construct a society in which we overly depend on AI, nor should AI be used to control human behavior by pursuing efficiency and convenience too much. We need to construct a society in which human dignity is respected, such as people demonstrating human abilities, showing great creativity, doing challenging work and living a lively life materially and spiritually, by using AI as their instruments.

(2) Diversity & Inclusion

The society in which people with diverse backgrounds, concepts on priorities and mentality can pursue their own idea of happiness, be flexibly included to create new values, is one of the ideals in present world worth challenged to establish. AI is a technology powerful enough for us to reach such an ideal. With appropriate development and deployment of AI, we need to change the framework of society to reach such a goal.

(3) Sustainability

We need to generate many new businesses and solutions by utilizing AI, to solve social disparities and to build sustainable societies that can deal with global environmental problems as climate change. Japan, as one of the world leaders in science and technology, has a duty to enhance its legacy by utilizing AI to develop such societies

3 AI Ready Society: Essential social revolution to achieve Society 5.0

As for the technologies that are expected to contribute to the realization of Society 5.0, among the most prominent are IoT, robotics, ultra-high-speed broadband networks and AI. Even if we can entrust AI for complex processes to some extent, it is necessary for humans to set the objective configuration that answers the question, " what is the purpose of using AI?" AI can be used not only to make society better, but even to achieve undesirable purposes, and be used in an inappropriate way unconsciously. Therefore, we need to consider the ideal way of "human," "social systems," "industrial structures," "innovation systems," and "governance," that can answer "For what do you want to use AI," while paying close attention to the interaction with technological progress.

(1)"Human"

In "AI - Ready society" that responds to the spread of AI in every corner of society, how humans deal with AI is the key to realizing a society where AI can be fully utilized. The ability and role expected by human beings are as follows.

People have to fully understand the advantages and disadvantages of AI. In particular, it is important for people to have the ability to recognize that biases are included in the data, algorithms, or both, which are the information resources of the AI, and that someone may use them for mala fide purposes. In addition, people should recognize that the data have three types of biases, namely, statistical bias, bias caused by social aspect and biases by maliciously introduced by some AI user.

By using AI, an environment to be realized is such that many people can engage in creative and/or highly productive jobs. To that end, it is expected that diverse people can acquire the ability to realize the various dreams and ideas aimed at by the support of AI whatever origin, culture, taste, etc. they have. An educational system must be built to realize this and a social system that contributes to achieving them.

There are sufficient human resources who acquired the application skills to design and implement AI system with the basic education of data and AI through a cross-disciplinary, combined, and integrated framework in a wide range of fields. It is expected that such human resources will be the driving force for all activities of society and that the capabilities of those people will contribute to establish the AI based human living environment. By improving the living environment like this, a social system that enables many people to live a richer and more fulfilling life has been realized.

(2)"Social Systems"

It is expected that by using AI, we can accelerate the evolution of services and solutions and gain various benefits from efficiency improvement and individualization. To take proper advantage of this change in society, it is necessary that the entire social system, including medical, financial, insurance, transportation, energy, and many other fields, can flexibly change and cope with AI evolution. This includes not only simple efficiencies in consideration of socially accepted existing purpose (such as improved convenience and liberation from simple labor), but also the realization of new value that may be generated by the diversification and mobilization of the objective and prevent the widening potential negative aspects (such as inequality and widening disparity, social exclusion) by the evolution of AI.

We need to implement flexible and robust architecture designs that have mechanisms to expand, interconnect, and improve order formation, in both the hardware and software of the social system. Besides, we need to develop a common data utilization architecture for various social systems, especially for guaranteeing interconnection and collaboration.

(3)"Industry Structures"

For diverse people to realize variety of dreams and ideas, it is necessary that labor, employment, and entrepreneur environment be flexible and open internationally. Therefore, companies need to commit to fair competition and promote flexible working styles. Human creativity needs to be demonstrated through industry. And investment for start-ups needs to be promoted.

(4)"Innovation Systems"

It is necessary to create an environment in which innovation is relentless. It could come from universities/research institutes/companies, even ordinary citizens. Anyone should be able to participate in the research and development, utilization and evaluation of AI regardless of one's occupation or position, and to stimulate each other.

Therefore, all kinds of data including real space data, should be instantaneously and securely available at the AI analyzable level. It is required that everyone can provide data at ease and can obtain the benefits from the data that they provided.

This desirable development is accelerated by creating an environment in which the user and researcher can develop and utilize AI at ease, and R&D and utilization cycle should be accelerated. It is also expected that AI utilization will create new ideas and possibilities, and widen the potential of innovation significantly.

(5)"Governance"

Social change and technological development will naturally require the continuous update of the goal settings of the aforementioned 'human', 'social systems', 'industrial structures', and 'innovation systems'.

Therefore, we should establish an agile system in which the diverse stakeholders - the government, industries, universities, research institutions, and the general public – should collaborate to investigate, assess, decide, and implement the rules, the systems, the standards, and the behavior norms in order to achieve the appropriate governance.

The diverse stakeholders should be cautious not to exclude the social minorities when they build the systems to solve technological/social problems. The governance can be achieved not only by law but also by technological architectures. Industries should play the independent roles to implement those architectures flexibly and effectively. In addition to the domestic governance, we should implement the international collaboration systems to cope with the problems that cross national borders.

4 Social Principles of Human-centric AI

We recognize that in order to realize "AI-Ready society" and to promote appropriate and proactive social implementation of AI, it is important to establish basic principles that each stakeholder should bear in mind.

We will systematize these basic principles into "Social Principles of AI" which society (especially a legislative or administrative organ such as a state) should pay attention to, and " Development and Utilization Principles of AI " which researchers, developers and user enterprises of AI should pay attention to, so that AI can be accepted and used appropriately by society. "Social Principles of AI" necessary for realizing a society with the three basic principles listed in Chapter 2 and " Development and utilization principles of AI " to be considered by developers and user enterprises are as follows:

4. 1 Social Principles of AI

Social principles of AI are the principles on social frameworks that should be implemented within national and local governments and multilateral cooperation in an "AI-Ready Society."

(1) Human-centric

Utilization of AI should not infringe upon fundamental human rights that are guaranteed by the Constitution and international norms.

AI should be developed and utilized and implemented in society to expand the abilities of people and to pursue the diverse concepts of happiness of diverse people. In the AI utilized society, it is desirable that we implement appropriate mechanisms of literacy education and promotion of proper uses, so as not to over-depend on AI or not to ill-manipulate human decisions by exploiting AI.

AI can expand human abilities and creativity not only by replacing part of human task but also by assisting human as an advanced instrument.

When using AI, people must judge and decide for themselves how to use AI. Appropriate stakeholders involved in the development, provision, and utilization of AI should be responsible for the result of AI utilization, depending on the nature of the issue.

In order to avoid creating digital divide and allow all people to reap the benefit of AI regardless of their digital expertise, each stakeholder should take into consideration to user-friendliness of the system in the process of AI deployment.

(2) Education

In a society premised on AI, we have to eliminate disparities, divisions, or socially weak people. Therefore, policy-makers and managers of the enterprises involved in AI must have an accurate understanding of AI, the knowledge for proper use of AI in society and AI ethics, taking into account the complexity of AI and the possibility that AI can be misused intentionally. The AI user should understand the outline of AI and be educated to utilize it properly because AI is much more complicated than the already developed conventional tools. On the other hand, from the viewpoint of AI's contributions to society, it is important for the developers of AI to learn about the social sciences, business models, and ethics, including normative awareness of norms and wide range of liberal arts not to mention the basis possibly

generated by AI.

- From the above point of view, it is necessary to establish an educational environment that provides AI literacy according to the following principles, equally to every person.
- In order to get rid of disparity between people having a good knowledge about AI technology and those being weak in it, opportunities for education such as AI literacy are widely provided in early childhood education and primary and secondary education. The opportunities of learning about AI should be provided for the elderly people as well as workforce generation.
- Our society needs an education scheme by which anyone should be able to learn AI, mathematics, and data science beyond the boundaries of literature and science. Literacy education provides the following contents: 1) Data used by AI are usually contaminated by bias, 2) AI is easy to generate unwanted bias in its use, and 3) The issues of impartiality, fairness, and privacy protection which are inherent to actual use of AI.
- In a society in which AI is widely used, the educational environment is expected to change from the current unilateral and uniform teaching style to one that matches the interests and skill level of each individual person. Therefore, the society probably shares the view that the education system will change constantly to the above mentioned education style, regardless of the success experience in the educational system of the past. In education, it is especially important to avoid dropouts. For this, it is desirable to introduce an interactive educational environment which fully utilizes AI technologies and allows students to work together to feel a kind accomplishment.
- In order to develop such an educational environment, it is desirable that companies and citizens work on their own initiative, not to burden administrations and schools (teachers).

(3) Privacy

In society premised on AI, it is possible to estimate each person's political position, economic situation, hobbies / preferences, etc. with high accuracy from data on the data subject's personal behavior. This means, when utilizing AI, that more careful treatment of personal data is necessary than simply utilizing personal information. To ensure that people are not suffered disadvantages from unexpected sharing or utilization of personal data through the internet for instance, each stakeholder must handle personal data based on the following principles.

- Companies or government should not infringe individual person's freedom, dignity and equality in utilization of personal data with AI technologies.
- AI that uses personal data should have a mechanism that ensures accuracy and legitimacy and enable the person herself/himself to be substantially involved in the management of her/his privacy data. As a result, when using the AI, people can provide personal data without concerns and effectively benefit from the data they provide.
- Personal data must be properly protected according to its importance and sensitivity. Personal data varies from those unjust use of which would be likely to greatly affect rights and benefits of individuals (Typically thought and creed, medical history, criminal record, etc.) to those that are semi-public in social life. Taking this into consideration, we have to

pay enough attention to the balance between the use and protection of personal data based on the common understanding of society and the cultural background.

(4) Security

Positive utilization of AI means that many social systems will be automated, and the safety of the systems will be improved. On the other hand, within the scope of today's technologies, it is impossible for AI to respond appropriately to rare events or deliberate attacks. Therefore, there is a new security risk for the use of AI. Society should always be aware of the balance of benefits and risks, and should work to improve social safety and sustainability as a whole.

- Society must promote broad and deep research and development in AI (from immediate measures to deep understanding), such as the proper evaluation of risks in the utilization of AI and research to reduce risks. Society must also pay attention to risk management, including cybersecurity awareness.
- Society should always pay attention to sustainability in the use of AI. Society should not, in particular, be uniquely dependent on single AI or a few specified AI.

(5) Fair Competition

A fair competitive environment must be maintained to create new businesses and services, to keep economic growth sustainable, and to solve social issues.

- There must not be unreasonable data-collection or infringement of sovereignty under a dominant position of a particular country by concentrating AI resources.
- There must not be unreasonable data-collection or unfair competition under a dominant position of a particular company by concentrating AI resources.
- By using AI, influence for wealth and society should not be overly biased on some stakeholders in the society.

(6) Fairness, Accountability, and Transparency

- Under the "AI-Ready society", when using AI, fair and transparent decision-making and accountability for the results should be appropriately ensured, and trust in technology should be secured, in order that people using AI will not be discriminated on the ground of the person's background or treated unjustly in light of human dignity.
- Under the AI design concept, all people must be treated fairly without unjustified discrimination on the grounds of diverse backgrounds such as race, sex, nationality, age, political beliefs, religion, etc.
- Appropriate explanations should be provided such as the fact that AI is being used, the method of obtaining and using the data used in AI, and the mechanism to ensure the appropriateness of the operation results of AI according to the situation AI is used.
- In order for people to understand and judge AI proposals, there should be appropriate opportunities for open dialogue on the use, adoption and operation of AI, as needed.

- In order to ensure the above viewpoints and to utilize AI safely in society, a mechanism must be established to secure trust in AI and its using data.

(7) Innovation

- To realize Society 5.0 and continuous innovation in which people evolve along with AI, it is necessary to account for national, industry-academia, and public-private borders, race, sex, nationality, age, political and religious beliefs, etc. Beyond these boundaries, through a Global perspective we must promote diversification and cooperation between industry-academia-public-private sectors, through the development of human capabilities and technology.
- To encourage mutual collaboration and partnership between universities, research institutions and private sectors, and the flexible movement of talent.
- To implement AI efficiently and securely in society, methods for confirming the quality and reliability of AI and for efficient collection and maintenance of data utilized in AI must be promoted. Additionally, the establishment of AI engineering should also be promoted. This engineering includes methods for the development, testing and operation of AI.
- To ensure the sound development of AI technology, it is necessary to establish an accessible platform in which data from all fields can be mutually utilized across borders with no monopolies, while ensuring privacy and security. In addition, research and development environments should be created in which computer resources and high-speed networks are shared and utilized, to promote international collaboration and accelerate AI research.
- To promote implementation of AI technology, governments must promote regulatory reform to reduce impeding factors in AI related fields.

4. 2. Development and Utilization Principles of AI

We recognize that developers and user enterprises of AI should establish and comply with the development and utilization principles of AI based on the fundamental philosophies and "Social Principles of AI" mentioned above.

Since the development and utilization principles of AI are currently being discussed in many countries, organizations, and companies, we emphasize it is important to build an international consensus through open discussions as soon as possible and to share it internationally as a non-regulatory and non-binding framework.

5 Conclusions

To lead the world in building the "AI-Ready Society," Japan should share these principles with the government, related companies, organizations, and so on, and reflect them in government policy.

In addition, Japan should share these principles with other countries and take on the leadership of international debate and aim to form a consensus. Thus Japan should show the social image of Society 5.0 supporting the realization of SDGs to the world, and should contribute to the cooperative and creative development of the international community.

These principles will be flexibly evolved and developed in the future, depending on the progress of AI related technologies, social changes, world situation changes, and many other factors.