Report on
Artificial Intelligence and Human Society

Unofficial translation

March 24th, 2017

Advisory Board on Artificial Intelligence and Human Society

Preface

Artificial Intelligence (AI) is expected to transform our society, not only by substituting for routinized tasks but also by supporting and enhancing human activities and decision making.

In its 5th Science and Technology Basic Plan (2016–2020), Japan introduced a new concept, namely “Society 5.0,” as a way by which to guide and mobilize action in science, technology, and innovation to achieve a prosperous, sustainable, and inclusive future that is, within the context of ever-growing digitalization and connectivity, empowered by the advancement of AI.

With its potential to equip and better shape our society with new services, businesses, social structures, values, and welfare, AI is perceived as a fabulous enabler, but its benefits to society will deeply depend on the way it will be implemented and used in real socioeconomic systems.

One may have concerns about the quick advancement of AI and its implementation in society. This is why the Advisory Board on Artificial Intelligence and Human Society was set up in May 2016 under the initiative of the Minister of State for Science and Technology Policy with the aim to assess different societal issues that could possibly be raised by the development and deployment of AI and to discuss its implication for society. The advisory board was expressly composed of members with various backgrounds in fields such as engineering, philosophy, law, economics, and social sciences. The advisory board finally published a report that includes a summary of the issues to be addressed regarding AI and human society.

I would like to share ideas and excerpts from this report with you to encourage communication and potential future cooperation. These issues should be discussed continuously as AI advances, ideally with international cooperation.

Yuko Harayama, PhD

Chair of the advisory board, and
Executive Member, Council for Science,
Technology and Innovation (CSTI), Cabinet Office, Japan

Board Members

Chair

Yuko Harayama	Executive Members of the Council for Science, Technology and Innovation
Noriko Arai	Professor, Director of Research Center for Community Knowledge, National Institute of Informatics
Arisa Ema	Project Assistant Professor, Science Interpreter Training Program, Komaba Organization for Educational Excellence, College of Arts and Sciences, The University of Tokyo
Shinya Ouchi	Professor, Graduate School of Law, Kobe University
Fumio Shimpo	Professor, Department of Policy Management, Keio University
Shoko Suzuki	Professor, Graduate School of Education, Kyoto University
Toru Nishikawa	President and CEO, Preferred Networks, Inc.
Kazuo Hashimoto	Professor, Center for Research Strategy, Waseda University
Izumi Hayashi	Attorney-at-Law, Sakurazaka Law Offices
Yutaka Matsuo	Project Associate Professor, Department of Technology Management for Innovation, The University of Tokyo
Noriyuki Yanagawa	Professor, Graduate School of Economics, The University of Tokyo
Masazumi Wakatabe	Professor, Department of Political Science and Economics, Waseda University

Observers:

- National Strategy Office of Information and Communications Technology, Cabinet Secretariat
- Intellectual Property Strategy Headquarters, Cabinet Office
- Ministry of Internal Affairs and Communications
- Ministry of Education, Culture, Sports, Science and Technology
- Ministry of Health, Labor and Welfare
- Ministry of Agriculture, Forestry and Fisheries
- Ministry of Economy, Trade and Industry
- Ministry of Land, Infrastructure, Transport and Tourism

Table of Contents

Executive Summary

Artificial Intelligence (AI) technologies that can perform portions of human intellectual activities (perception, recognition, decision making, inference, and actions based on the same) have been advancing quickly, especially big data and machine learning. They are being implemented in autonomous cars, diagnosis supports, conversation agents, and others. AI technologies are critical and fundamental for the realization of the Society 5.0 that Japan's government is promoting, and it is expected to contribute by giving solutions to social problems, such as the labor shortage caused by the low birthrate and aging society, and to enable every person to play a significant role by utilizing their own abilities. Although AI technologies bring tremendous benefits to human society, they might affect human society itself fundamentally by advancing without awareness because their development is so fast that it inevitably surpasses the pace of institutional and social adaptation. Thus, the influence of AI technologies on society should be discussed to ensure that these technologies are used safely and beneficially.

Ethical, Legal, and Social Implications (ELSI) surrounding AI technologies are attracting attention internationally, including Japan, and various people as well as researchers and engineers discuss opportunities and concerns regarding AI technologies with great interest. For example, academic organizations, non-profit organizations, governments, and international organizations are discussing and providing policy recommendations on increasing productivity, decreasing heavy work, accelerating science, and finding new medical treatments as good opportunities, and they address job loss and existential risks to humans as concerns.

The Advisory Board on Artificial Intelligence and Human Society (Advisory Board) focused on existing AI technologies, those that may foreseeable be realized in the near future, and the society in which those technologies are or will be implemented. The Advisory Board's objective was to clarify what benefits are expected, what issues are to be concerned, what issues are to be resolved, and what attitudes are beneficial. Digitalization that cannot be dissociated from AI technologies was included in the discussions. Given that AI technologies are being applied in various fields, the Advisory Board took a case-based approach that dealt with various cases in the four representative categories: mobility, manufacturing, personal services, and conversation/communication. The Advisory Board aimed to clarify common key issues of AI technologies and human society from a multi-stakeholder point of view, i.e., from the viewpoints of users, researchers, engineers, artists, people across generations, including children, businesses, and governments.

AI technologies support and augment human intellect and actions, and they execute parts of intellectual behaviors on behalf of humans. This is supposed to greatly benefit and empower human society and contribute to ensuring its sustainability. The extracted key issues are summarized as follows:

- Ethical issues: A significant action is to consider the balance between human decisions and AI-based decisions depending on the situations and objects to be judged. The balance change causes the emergence of a new sense of ethics. If users confirm AI services that enable them to manipulate someone's mind and/or to evaluate people, discussions of ethics might especially be needed. Careful attention contributes to proper AI progress that has the possibility to change

human concepts, since AI technologies augment human beings' senses and abilities. New evaluation procedures have required the observation of the values (e.g., originality, utility, and virtue) of the products made and actions performed by humans, by AI technologies, and through the cooperation of both.

- Legal issues: Determining the locus of responsibility for accidents involving AI technology along with preparing insurance for probabilistic risks contributes to social acceptance and helping users understand the risks of utilizing AI technologies. Exploitation of big data while considering information privacy protection requires the consideration of appropriate institutional frameworks (laws, guidelines, and contracts). Considering the rights to and incentives for the creation of AI technologies remains a subject of further study for multi-stakeholders. In-depth analysis and basic research (e.g., social sciences) play an important role in reconsidering fundamental concepts, such as human responsibilities, that the modern law is based on.
- Economic issues: Individuals changing their work style and updating their abilities propose to harmonize each person's abilities with a creative job/task. These changes also require that companies reconsider their decision-making techniques and staff (re)assignment to take advantage of work flexibility. At the government level, combining educational and employment policies is one of the effective procedures for mobilizing labor, revitalizing the economy, and preventing economic disparities.
- Educational issues: The significant issues are understanding the advantages and limitations of the present AI technologies, properly utilizing AI technologies, and performing creative activities in collaboration with AI technologies. An educational policy functions according to discussions of how to efficiently reform curriculums based on evidence that shows the technologies' limitations, the critical human abilities differentiated from present AI technologies, and the essential human abilities to be acquired.
- Social issues: There is a need to create space for dialogue among people with different visions and ideas and to consider common, fundamental social values. One of the fundamental issues is the need to facilitate provisions against the "AI divide," unbalanced social costs relative to AI, and discrimination that occurs because of the technologies' procession. Continuous assessments of social pathology, conflict, and dependence on AI technologies will offer solutions to these issues.
- Research and development (R&D) issues: Researchers and engineers are required to engage in R&D with professional ethics while observing ethical codes and guidelines. Technologies for cyber security and privacy protection must be advanced. Basic research to ensure that AI technologies are controllable and transparency by explaining the processes and logics of calculations made by AI technologies contribute to their social implications. Advancing AI technologies that are beneficial for human society requires basic sciences, including social sciences to study societal issues that may arise in the future, observing the social acceptance of their stochastic behaviors such as deep learning, and creating an environment that supports open science to enhance the diversity of AI

technologies.

We have already initiated international discussion on these key issues and are engaged in the pursuit of this process because we consider clarifying the relationship between AI technologies and human society to be a prerequisites to realizing a sustainable society; we also consider addressing the fundamental question, “What values are shared by humans all over the world,” to be unavoidable. The Advisory Board wishes for this report to be shared by not only AI researchers/engineers, governments, and business sectors but also by humanities and social scientists, education suppliers, and the general public in order to open debate and precipitate taking action.

Introduction

1.1 Impacts of AI technologies on human society

In 1956, a proposal was made to consider various research on "thinking machines" as a new research horizon, "Artificial Intelligence (AI)," at the Dartmouth conference held at Dartmouth College in the United States of America. Research had previously been conducted on whether computers could be invented or intelligence artificially realized, but AI was officially set as a research horizon at this conference. The realization of AI is the researcher's goal. Complete AI does not exist yet, and it is difficult to say what is complete. Technologies that can replace some intellectual activities (perception, recognition, reasoning, learning, decision making, behaviors/actions based on the same) have been developed in this research process. In this report, these technologies are defined as “Artificial Intelligence technologies (AI technologies)” and are the subjects of study.

In March 2016, AI technology won the world championship of Go with 4 wins and 1 loss. This progress was a great surprise. One of the reasons is that Go is a very difficult game, and since there are numerous variations of the following measures, it was thought that the computer’s comprehensive solution search would be unable to win the world championship. In other words, AI technologies exceed humans, even concerning games that require high intellect. Around 2014, vehicles with automatic driving (auxiliary) function (Level 2) were sold on the open market. These vehicles can be driven automatically with multiple functions such as acceleration, steering, and braking through information gathered via cameras, radars, and other sensors in a limited situation, such as on highways. In the summer of 2016, it was reported that a cancer patient for whom an adequate treatment could not found recovered by using an alternative treatment that utilized AI technologies.

In addition, corporate activity, the way of working, and everyday life are changing with, digitalization, such as the use of rideshares. Among the optimization of information in digitalization, there are many cases that cannot be strictly separated from AI technologies.

Currently, the government of Japan aims for the realization of its Society 5.0[1], which can overcome social problems such as the declining birth rate and the increasing age of the population by using science and technology and which can ensure sustainable development for the future. Therefore, AI technology is regarded as an important fundamental technology with the Internet of Things (IoT), which accelerates robotics and digitalization.

The progress of AI technologies in recent years has held surprises for society, and the government as well as companies invest in R&D tremendously as a driving force for the realization of a sustainable society and as core technologies for industrial strengthening.

1 Society 5.0 is a society in which the different needs of members are distinctly met through the provision of relevant goods and services in the desired amount to those people who need them and in which the entire population has access to top-notch services, and lives convenient, vibrant lives that take into account their differences, such as age, gender, religion or language (The 5th Science and Technology Basic Plan, Cabinet Office, Japan).

1.2 Expectations and concerns regarding AI technologies

AI technologies have brought tremendous benefits to human society. For example, autonomous cars are expected to improve safety by reducing accidents caused by drivers' carelessness and mistakes. Furthermore, there will be a change in which people become free from tedious labor and spend more time on high-value-added jobs and hobbies as machines become increasingly automated. A new lifestyle is upcoming that enables users to discover new knowledge efficiently, as shown by the automated cancer treatment suggestion, if human society can accelerate intellectual tasks that require an enormous amount of specialized knowledge and a great deal of time. Therefore, AI technologies offer to make human society more convenient, more prosperous, more secure, and safer.

On the other hand, those people who are concerned about the advancement of AI technologies have the following questions: “Who is responsible for accidents caused by an autonomous car?”; “Will AI steal our jobs?”; “Will our children lose motivation for learning as AI technologies become common?”; and “Will we be manipulated by AI before we know it?” The aim of this report is to clarify the current situation and issues to be considered, rather than impetuously trying to guarantee the safety and security of AI. Some concerns involve outcomes that are unlikely while others would require individuals (both users and developers), businesses, and governments to consider appropriate actions.

Although utilizing AI technologies would raise concerns and require some actions in the short term, they could offer more benefits if users properly deal with the risks and concerns that arise. Japan has a mature society and is also referred to as the forerunner of finding solutions to emerging issues. AI technologies are expected to contribute to economic growth and safe, secure and prosperous lives, overcoming challenges such as the low birth rate, greater longevity, and natural disasters, and thereby to realize a sustainable society.

1.3 Features of AI technologies differ from previous technologies

Although human societies have previously accommodated and accepted new technologies with both concern and expectation, there is a question as to why current society primarily faces troublesome technologies relative to AI.

What distinguishes AI technologies from previous technologies is that it is enabled to do things that have been assumed to only be doable by humans. In addition, given appropriate goals, AI technologies can learn from data and self-improve rapidly. It should also be noted that AI technologies are invisible, unlike electric equipment, cars, and cell phones, and they work in the background inside of existing items and services. Users might find it difficult to assess whether AI technologies are actually used and how much they are used in those since they work as part of a program that cannot be directly, physically observed. AI technologies can use existing tools and technologies as they are in pretty much the same manner as humans. As a result, users cannot simply confirm whether AI technologies are implemented at a glance. It might occur to people that AI has already prevailed in a wide variety of technologies and scenes or that alleged AI products do not actually use AI technologies. Taken together, AI technologies can prevail and advance before users know it, since their development is rapid and their behavior is hard to see from

the outside. The characteristics of AI technologies described above should be considered when addressing issues and determining appropriate actions to take in resolving them.

1.4 Purposes of Advisory Board

The Advisory Board on Artificial Intelligence and Human Society (hereafter, the Advisory Board) was established in May 2016 under the Minister of State for Science and Technology Policy in order to advance the R&D and utilization of AI technologies. AI technology is one of the most significant technologies that facilitates the realization of Society 5.0. The Advisory Board has worked especially to clarify benefits, issues, challenges, and future directions with special attention to AI technologies that exist or will be realized in the near future and a society in which such technologies prevail.

Given that AI technologies are being applied in various fields, the Advisory Board adopted a case-based approach that took up various cases in the four representative categories: mobility, manufacturing, personal services, and conversation/communication. These services are based on not only AI technologies but other technologies that cannot be clearly divided from the former, defined as digitalization. In this report, AI technologies include digitalization, and the Advisory Board aimed to clarify significant issues regarding AI technologies and human society from multi-stakeholder points of view, i.e., from the viewpoints of users, developers, people across generations, including children, artists, businesses, and governments, and thereby serving as a starting point for continuous discussions moving forward.

Overview of activities on AI and human society

Along with the rapid progress of AI technologies and intellectual machines, several global activities that study the impact of AI on human society are concurrently being conducted. Characteristically, in addition to scientists, engineers and government agencies are involved in AI research, but researchers and entrepreneurs in other fields as well as the general public also have a strong interest in these issues.

2.1 Global trends

In 2008–2009, the Association for the Advancement of Artificial Intelligence (AAAI) created an expert group, led by the president, and examined the long-term influence the development of AI had on society (2008–2009 AAAI Presidential Panel on Long-Term AI Future). In its development, the project "One Hundred Year Study of Artificial Intelligence" was launched in 2014 at Stanford University (USA). The project members are continuously investigating and examining opportunities and concerns from the viewpoint of laws and ethics about the long-term impact of AI development on society. The report of 2016[2] examines the influence of AI research for each field, such as transportation, home, healthcare, and others, assuming a typical North American city until 2030, and it also makes policy recommendations. In addition, by 2030 it is stated that there will be no concern that humankind will be destroyed by AI. The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems published “Ethically Aligned Design, Version 1,” which aimed to provide ethical issues and candidate recommendations about AI and autonomous systems and to provide recommendations for IEEE Standards based on Ethically Aligned Design.[3]

In addition, a voluntary organization, "Future of Life Institute (FLI)," was established in Boston (USA) in 2014, and enables not only researchers in various fields such as humanities, social science, and natural sciences but also actors, students, and entrepreneurs with various backgrounds to gather and deal with existential risks that the development of AI can bring to human beings as well as biotechnology, nuclear weapons, and climate change. Funding for AI R&D, organization of conferences,[4] and enlightenment activities are promoted using donations. Future risks of Artificial General Intelligence (AGI) and safety concerning AI (AI safety) are also involved in the study.

In the "Future of Humanity Institute (FHI)" (established in 2005) of Oxford University (UK), the existential risks to mankind, including AI safety, are studied. In the "Center for the Study of Existential Risks (CSER)" (established in 2012) of Cambridge University, prediction and evaluation of global catastrophic risks due to the development of science

2 Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, Kevin Leyton-Brown, David Parkes, William Press, AnnaLee Saxenian, Julie Shah, Milind Tambe, and Astro Teller. "Artificial Intelligence and Life in 2030." One Hundred Year Study on Artificial Intelligence: Report of the 2015–2016 Study Panel, Stanford University, Stanford, CA, September 2016. Doc: http://ai100.stanford.edu/2016-report (retrieved on February 24, 2017).

3 http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html (retrieved on February 24, 2017).

4 Beneficial AI 2017. https://futureoflife.org/bai-2017/ (retrieved on February 24, 2017).

and technology, especially AGI, are studied.

Activities for organizing an Institutional Review Board (IRB) on AI R&D in a private company as well as for making a partnership of companies to publicize correctly AI's influence on society and to share information to prevent bad effects on society and ethics have been reported.

Government agencies and international organizations commence similar studies. For example, in the United States, the Office of Science and Technology Policy (OSTP) held workshops four times with participation from citizens and studied the benefits and risks that AI technology brings to society (the report was released in October 2016).[5] In the Organization for Economic Co-operation and Development (OECD), the Committee on Digital Economy Policy mainly began attempting to examine necessary policies and the benefits that digitalization including AI technology brings to society.

2.2 Trends in Japan

The Japanese Society for Artificial Intelligence (JSAI) has put importance on connectivity with society taking into consideration the high presence and strong impact of AI in the general society along with the opportunity that changes to the journal's name and cover design would become an important topic in society as a trigger.[6] The Ethics Committee was established in December 2014 with the aim of discussing and considering the relationship between AI and society, then it released the Ethical Guidelines in February 2017. The "Acceptable Intelligence with Responsibility (AIR)," where researchers in fields such as science, technology and society (STS), and AI gather, mainly examines dialogue during on-site visits and interviews about the social acceptance of AI. The study group on robot law in the Information Network Law Association is investigating international research trends of social institutions and consumer protection. Efforts are being made from such various perspectives.

The "Fifth Science and Technology Basic Plan,"[7] formulated by the Council for Science, Technology, and Innovation (CSTI) and started in 2016, is a plan to implement a systematic and consistent science and technology policy from a long-term perspective. That AI technology is an important part of realizing Society 5.0, the importance of deepening the relationship between science and technology innovation and society, and ethical, legal and social efforts are mentioned in the plan. In February 2016, the Ministry of Internal Affairs and Communications set up "The Committee for AI Networking" to evaluate the impacts and risks of AI networking on society and the economy. The meeting then released the principles for AI R&D. Since October 2016, "The Committee for Promoting AI Network Society" has continued making guidelines for AI R&D.

5 https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf (retrieved on February 24, 2017).

6 Matsuo et al. (2015), Introduction of JSAI Ethics Committee. Journal of Japanese Society for Artificial Intelligence, 30(3), 358–364 (in Japanese).

7 http://www8.cao.go.jp/cstp/english/index.html (retrieved on February 24, 2017).

Will AI surpass human intelligence?

Artificial Intelligence (AI) has been presented as beings that are a possible threat to humans or that feel conflicted just as humans do, due to their intelligence either equaling or surpassing that of humans in the science fiction (SF) world. There has recently been much discussion on AI and human society all over the world, and a part of such discussion seems to be based on the image of AI that has been created in the SF world. In what follows, we consider the probability of such AI threats becoming reality in order to identify ways in which to further deepen discussion on AI and human society.

AI that tries to control or extinguish the human race has been described in SF; e.g., the movies *The Terminator* and *The Matrix* as well as Japanese cartoon "Phoenix: A Tale Of The Future." The movies *2001: A Space Odyssey*, *Alien*, *Blade Runner*, and *Ex Machina* presented AI that attempts to harm humans.

Movies like *Bicentennial Man* and *A.I.* depicted AI that feels conflicted, just as humans. The movie *Her* presented romance between a man and an AI woman. Regarding such AIs, there have recently been discussions on how to deal with extraneous things that consist of machineries and computer programs and yet seem to have human-like minds as well as on whether such AI should be developed [1].

The probability of AI with great intellect becoming reality may seem plausible given the recent rapid advancement of AI that has enabled its victory over humans, e.g., the quiz champions [2], the Go champion [3], and the experienced fighter pilot [4] and that approach a human level even in the fields of creation, e.g., recipes [5] and novels [6].

However, any actual and current AI technologies are basically executed functions of computer programs, i.e., a description of a concrete computation process that reflects the objective given by human programmers. Basically, it only works according to the human programmers' intention, since it cannot set its own objectives, unlike the AI of SF described above, though recent advancement such as deep learning has made it possible for computers to learn from data. AI sometimes might not work according to its human programmers' intention if it is based on machine learning or statistical methods or if it contains bugs. However, it is almost impossible for AI to set its own objective independent of human programmers and behave in a rational manner toward that objective. The objectives given to actual AI are specific and concrete tasks by which the human programmers' desired output is obtained from a given input; actual AI only performs specific and concrete tasks like image recognition, speech recognition, and automated driving. So far, existing AI does not have general intelligence, as humans do; it can only automate the execution of specific and concrete tasks.

It cannot be completely denied that entirely new hardware and software will be invented through technological advancement, which will remove the current AI's limitations and create artificial general intelligence (AGI) or Superintelligence (a hypothetical AI that possesses intelligence far surpassing that of the brightest and most gifted human minds [7]), which will be a threat to humankind. We should be aware of the potential risk of harmful people using AI to case damage human society. However, the probability that AI will have its own intentions and could spell the end

of human race will be quite low over the coming decades.

A White House report, *Preparing for the Future of Artificial Intelligence*, published October 2016, says that general AI will not be achieved for at least decades [8]. The One Hundred Year Study on Artificial Intelligence (AI100) [9], where Stanford University plays a central role in discussing AI and human society, has recently published a report, *Artificial Intelligence and Life in 2030* [9], that states that "Contrary to the more fantastic predictions for AI in the popular press, the Study Panel found no cause for concern that AI is an imminent threat to humankind." Andrew Ng, an expert in machine learning in Baidu, states that "The reason I say that I don't worry about AI turning evil is the same reason I don't worry about overpopulation on Mars" [10].

On the other hand, the Future of Life Institute (FLI) [11], where existential risks including AI are being studied, mentions, as the reason why they mainly study AI with intelligence equaling humans or Superintelligence, that we should start studying such AI immediately given its huge impact on humankind [12].

Our view is that, for further deepening of the discussion on AI and human society, it is important to balance the opposing positions mentioned above: the one in which discussion of AI and human society should be realistic, pragmatic, and based on a precise understanding of the current states of technologies, and the other in which we should consider the worst-case scenario regardless of the probability of its being reality.

References

1. Matsuo et al., Artificial Intelligence and Ethics, Journal of the Japanese Artificial Intelligence, 31(5), 635-641. (2016) (In Japanese)
2. Stephen Baker, Final Jeopardy: Man vs. Machine & the Quest to Know Everything, Houghton Mifflin Harcourt. (2011)
3. https://deepmind.com/research/alphago/
4. http://www.huffingtonpost.jp/engadget-japan/ai-fighter_b_10746878.html
5. http://www.huffingtonpost.jp/2014/07/08/chef-watson_n_5568914.html
6. http://www.fun.ac.jp/~kimagure_ai/
7. Nick Bostrom. Superintelligence: Paths, Dangers, Strategies. Oxford University Press. (2014)
8. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf
9. https://ai100.stanford.edu/
10. https://ai100.stanford.edu/2016-report
11. https://www.wired.com/brandlab/2015/05/andrew-ng-deep-learning-mandate-humans-not-just-machines/
12. http://futureoflife.org/ http://futureoflife.org/2015/10/12/ai-faq/

Approach

The Advisory Board aimed to create concrete discussions on probable situations, rather than abstract discussions on concerns or the blueprints of the distant future. Therefore, the board adopted an approach based on cases to estimate the influence on society by focusing on existing AI technologies as well as those foreseeable in the near future. The board members, experts in various categories, and business executives participated in the discussion, and some examples were extracted considering ethics, legality, economics, and others, as imaging the near future.

In addition, public comments, a brainstorming workshop with university students who were interested in AI and human society, and public workshops at Miraikan (National Museum of Emerging Science and Innovation) with the general public, including children, were conducted to widely collect opinions and/or examples. As a result, common issues relative to several categories were identified. These issues enabled the members to consider a many-sided discussion about human society, for example, to analyze employment problems from economic, educational, and R&D viewpoints.

3.1 Categories of cases

Applying AI technologies to various life scenes in the near future would change personal lifestyle and business factors. The four categories of mobility, manufacturing, personal services included medical care and finance, and conversation/communication are especially suggested to greatly impact immediate lifestyle and business. Therefore, these categories were selected for the discussions in consideration of characteristic Japanese industries, and many examples were collected.

- Mobility: covering moving persons and goods, e.g., autonomous cars, ride-share services, and delivery systems by drone.
- Manufacturing: covering manufacturing and production, e.g., fully automated factories, power-assisting suits that extend abilities of operations, highly accurate manufacturing, and reproduction of arts and expert techniques.
- Personal services including medical care and finance: covering ICT services for individuals, e.g., e-commerce, recommending services with web browsing, medical care services applying AI technologies, robot advisers, transaction support systems and/or credit administrations based on AI technologies.
- Conversation/Communication: covering human communications using ICT or robotics and human-machine communications, e.g., anthropomorphic (spoken) dialogue agents, chat bots, and machine translation services.

3.2 Overview of issues

A statement on science and technology's relationship with human society has required ELSI assessments since the Human Genome Project. AI technologies supposed that industries and labor could economically change in perspective. It was implied that

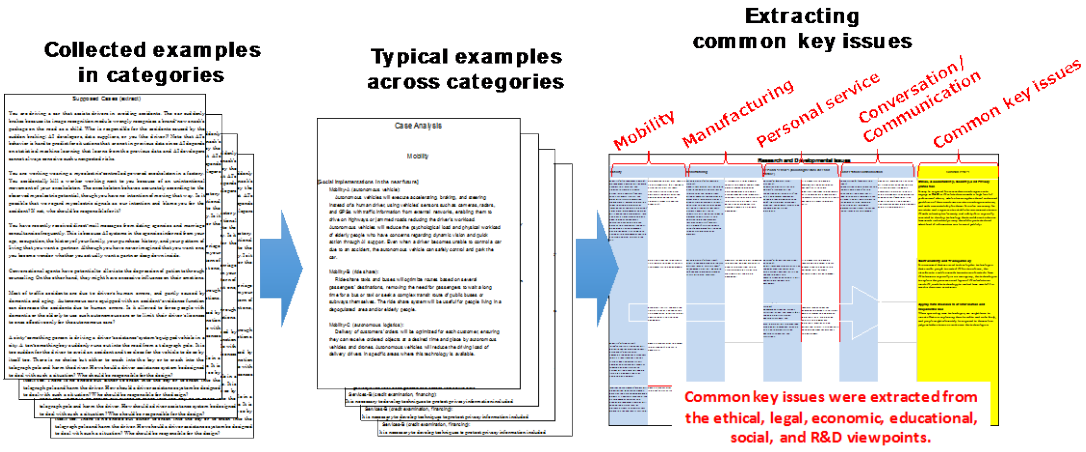
reforming the education system would effectively work for children if a point is made to include computer science and AI study. Furthermore, appropriate AI progress suggests applying the board's conclusion to AI R&D. Thus, the board adopted three additional perspectives (Economic, Educational, and R&D issues) and facilitated discussion considering six points of view.

- Ethical issues: ethics and morality relative to outputs of AI technologies and those values.
- Legal issues: laws and/or contracts relative to utilizing AI technologies, responsibility for AI products, and protection of personal information.
- Economic issues: changes in the economy, companies, and labor by the popularization of a lifestyle based on AI technologies as well as required policies.
- Educational issues: identifying how to improve the education system and resolve problems for schools, colleges, universities, education suppliers, and students.
- Social issues: the relationship between users and the implementation of AI technologies as well as indirectly resultant social problems.
- R&D issues: the implementation by AI researchers and engineers of AI technologies and educating users about how to avoid the disadvantages of AI products and/or services.

3.3 Extracting common key issues

According to the discussions of many examples from the four categories, the objective of the Advisory Board was to provide common key issues from the six points of view. Typical cases of implemented AI technologies were selected in order to create a matrix in which the columns indicated categories and the rows explained examples. By analyzing the matrix, key issues were extracted.

Although this approach enabled the members to discuss the key issues, the result supposed that care would be taken not to arrive at an exhaustive extraction of all of the key issues. In addition, the examples had different implementation time-scales and market values. The Advisory Board also had not dealt with items such as anti-social or unexpected usage of AI technologies, autonomous AI networking products, and others. Therefore, the above-mentioned issues still require further study. The Advisory Board focused on realistic and significant examples that are current or foreseeable for the near future, and they summarized critical issues to contribute to continuing discussions.


Collected examples in categories
Typical examples across categories
Extracting common key issues
Mobility
Manufacturing
Personal service
Conversation/ Communication
Common key issues
Common key issues were extracted from the ethical, legal, economic, educational, social, and R&D viewpoints.

Summary of the issues to be addressed regarding AI technologies and human society

AI technologies support and augment human intellect and actions, and they execute parts of intellectual behaviors on behalf of humans. Thus, it supposes to greatly benefit and empower human society and to contribute to ensuring its sustainability if utilized safely and beneficially. However, AI researchers/engineers, governments, and business sectors as well as humanities and social scientists, education suppliers, and the general public need to consider an appropriate relationship between AI technologies and human society and to prepare for it to realize and support a sustainable society. The Advisory Board has attempted to identify key issues to be addressed as summarized below.

4.1 Ethical issues

The changing relationship between humans and AI technologies and the emerging new sense of ethics

Humans have utilized various tools and machines to make choices and decisions depending on circumstances. The advancement of AI technologies is increasing the cases in which they, using big data, can make accurate and quick decisions, semi-automatic operations, and statistically appropriate choices. When AI technologies support human choices and decisions, there are many benefits, such as improvement in accuracy and speed as well as independence from human cognitive bias and prejudice.[8] However, it is important to consider the balance between human decisions and AI-based decisions depending on the situations and objects to be judged.[9] Since relationships between humans and AI services/products/machines have gradually changed as AI technologies have advanced, it is likely that there is an emergence of a new sense of ethics based on these evolving relationships.

Concerns about manipulating emotion, faith, and behavior and ranking or selecting AI technologies without awareness

AI technologies are becoming able to support and make decisions and actions that only humans have previously been able to perform. Many people have concerns and anxieties about AI's potential manipulation or operation of their minds and behavior, the evaluation or ranking of people by AI technologies, and AI influencing people's emotions, affections,

8 Cognitive bias is a confirmation bias or a fundamental ascription error. Confirmation bias is the cognitive tendency to overestimate self-serving hypothetical information and emphasize confirmation. However, it is also known that AI technology based on machine learning deviates from judgment based on the biases of learning data (algorithm bias), requiring to prevent inequity based on the algorithm bias.

9 For instance, there is an unsolved ethical problem regarding whether one driver of a car should be sacrificed rather than several passengers in another car in an inevitable emergency based on utility; humans may want to save their own lives while AI technologies can be set to give priority to many lives over fewer lives based on utility. Thus, at the current stage, how we consider the balance between human and AI-based decisions is significant.

and faith.[10] Ethical discussions might especially be needed if these are conducted without people's awareness.

Revisiting the concept of humanity

The future blueprints show that AI technologies[11] augment human beings' senses and abilities regarding space, time, and the body. According to this, changing concepts of human ability and emotion are supposed. With interactions such as augmenting senses, there is an opportunity to revisit the concept of humanity by taking account of these AI's potential.[12]

Considering the value of products and actions relating to AI technologies: The diversity in values and future prospects

The application of AI technologies has enhanced productivity quantitatively and qualitatively. AI technologies simply produce objects that otherwise either could be made only by artists/experts or would require high costs and/or a long time to generate. This indicates that everyone should have access to such high-quality items. Increasingly, new evaluation procedures have been required to observe the values (e.g., originality, utility, and virtue) of products made and actions performed by humans, AI technologies, and cooperation between both. The objective is to provide assessment results about how those values are accepted in society. Furthermore, it is also important to provide opportunities for dialogue among various people. Cooperation between humans and AI technologies can lead to the augmentation of human ability being a basis of a new sense of values. To realize this objective, continuous discussions about various choices and a diversity of values are strongly demanded based on recognition of individuals' differences in values and future prospects.

4.2 Legal issues

Clarifying the locus of responsibility and utilizing insurance: Considering the risks of using and not using AI technologies

Considering legal issues contributes to the acceleration of AI technology utilization and acceptance safely in society. Previously, statistical reports showed that most traffic accidents were caused by human errors and carelessness. Although autonomous cars enable users to expect traffic accidents to decrease and thus create a safer society, one may be concerned about who is responsible for accidents caused by autonomous car systems. Society's nearly implemented AI technologies require clear determination of the locus of responsibility for risks, accidents, rights infringement, benefits, and achievements. For human society to accept and benefit from AI technologies, it could be useful to clarify the locus of responsibility according to the levels of technological advancement (e.g., levels 0 to 4 for automated driving technology)[13] and to deal with uncertain, probabilistic risks

10 Similar concerns have been discussed on TV and poster advertisements, and companies keep voluntary restrictions. When applying AI technologies, careful attention and consideration are required since services and advertisements are optimized for individuals.

11 Change by daily use of automated machines such as autonomous cars or AI-based prostheses.

12 What is human ability? Does it include daily-used AI technologies or prostheses?

13 It is said that drivers are responsible for the accidents caused by automated driving at the levels of 0 to 2.

through insurance[14]. This is also important in preventing businesses from becoming intimidated by or overreacting to reputation risks. Similarly, discussion from another point of view facilitates solving legal issues, such as human society's need to discuss the risks of losing opportunities and credibility by not using AI technologies.[15]

Exploitation of big data with consideration of information privacy protection

The ability to exploit big data would make AI technologies more useful. It is necessary to consider appropriate institutional frameworks (laws, guidelines, and contracts) to avoid the chilling effects of privacy invasion and to balance the usefulness of AI technologies with privacy issues. A mature society has to discuss access rights to personal information data, data portability, and related security issues with international cooperation. One of the prospects supposes government services based on AI technologies to embody the above-mentioned considerations.

Considering the rights of and incentives for creations by AI technologies

The exploitation of AI technologies encourages the easy creation of high-value products[16] by various people, such as algorithm developers, learned-data providers, service providers, and final creators. It is necessary to consider who retains the property rights to a creation or a learned model produced by AI technologies or the collaboration between AI technologies and humans (i.e., issues of data ownership). Furthermore, to facilitate the development and utilization of AI technologies, human society is required to find an appropriate method of assignment of rights (incentives) to algorithm developers, algorithm users, and data providers by means of appropriate contracts and guidelines on a case-by-case basis.

Interpretation and revision of laws and the basic science of underlying legal concepts

Continuous discussions contribute to appropriately revising laws (on transportation, business, medicine, labor,[17] and others) with the changes to jobs and employment caused by AI technologies. There is the possibility of fundamentally reconsidering the underlying concepts of laws, such as human responsibility. For example, the existing laws do not have answers to clearly determine the locus of responsibility for an output by AI technologies or collaboration of humans and AI technologies, as AI technologies based on machine learning are socially implemented.[18] The process of accepting AI technologies requires

14 The government needs to discuss whether a mandatory insurance as social insurance should be implemented and who pay for its premium.

15 For example, a patient could recover from a disease if an AI-utilized diagnosis has been applied, but he/she could not be given such diagnosis.

16 Industrial robots that can work either at the level of or better than human experts as well as AI technologies that can create pieces of work by modeling the style of famous human artists are already being put to practical use.

17 If self-employment is becoming popular as predicted, the present labor laws may change because they mainly concern employees' rights at work.

18 If someone is wearing a myoelectric-controlled powered exoskeleton while working in a factory, and a nearby coworker is accidentally harmed because of an unintentional movement of the exoskeleton, who is responsible for this accident? Do we possibly regard myoelectric signals as the user's intention and

human society to advance as well as the technologies. Thus, discussions and social sciences are required to reconsider fundamental concepts, such as human responsibilities, that modern laws are based on.[19]

4.3 Economic issues

AI technologies have promoted economic and industrial activities, and they generate additional employment in new jobs such as providing data for machine learning. The comparative advantage of AI technologies drastically changes the power relationships in business, just as the small number of companies that successfully exploited big data on the Internet gained extensive power in the information society. Careful awareness is effective at avoiding industrial monopoly, especially with regard to its influence on society. It is also anticipated that many companies can reduce business costs and improve their business impetus, since AI technologies services/products require less labor power to operate. However, quick and appropriate actions are needed since an economically inefficient situation might occur at the transition phase when AI technologies are being implemented ethically, legally, and societally.

Changes in tasks and the way people work caused by AI technologies: Individual workers

AI technologies have increasingly become capable of doing automated jobs/tasks in place of humans. Consequently, many people are required to focus on more creative activities, for instance, changing jobs. One issue that needs to be addressed is how to harmonize an individual's abilities with a creative job/task. In the case where it is difficult to change jobs, that is, human resources do not meet with job requirements, both of unemployment and a labor shortage will occur at the same time. To avoid these problems, an individual has to acquire the ability to work creatively. As new businesses increasingly implement AI technologies, work styles also change in various ways that enable more people to work on their own account.

Changes in employment systems and companies due to the utilization of AI technologies: Companies

AI technologies are now crucial for industries and companies if they are to compete internationally. The change of work style can be contributed to AI technologies facilitating the reduction of tedious, prolonged, and exhausting jobs/tasks and the increase high-value work, as previously mentioned. This also requires companies to reconsider their manner of decision making and staff (re)assignment in order to take advantage of work flexibility that is unconstrained by time and space, e.g., teleworking. Companies should have impetus to make quick decisions and take action when acquisition of human resources that can

---

therefore blame the user? Otherwise, is the data provider or algorithm developer responsible for the accident? There is not yet a clear answer.

19 For instance, autonomous cars are expected to decrease traffic accidents and make human mobility much safer than it is at present. However, very rare accidents would occur in unexpected situations that differ from learned data, where the locus of responsibility is unclear. Insurance is effective for protecting people from economical loss in such accidents. However, how we accept such accidents/events personally and socially is a substantial, basic question for the social sciences requiring further study and discussion.

develop or utilize AI technologies or reassignment of labor are needed. For employee reassignment, it is effective for companies to provide education opportunities.

Industrial policies facilitating AI technology utilization, and educational and employment policies enabling labor mobility: The government

At the government level, it is necessary to formulate policies that provide opportunities for people to learn abilities that enable labor mobility[20] in order to facilitate economic growth through AI technologies and ensure that is are a variety of work styles suitable for individuals.[21] The government also has to contribute to determining how to harmonize an individual's abilities with a creative job/task. Combining educational and employment policies is one of the effective procedures for enabling labor movement. In addition, the government has to appropriately put macroeconomic policies and safety nets in perspective. The procedures for the fair distribution of profits based on AI services/products, economic revitalization, and prevention of economic disparities should be proposed through consideration of the benefits. Since AI technologies are beneficial for Japan in confronting labor shortage, policies that enhance industrial competitiveness should be accelerated. Those policies will be more effective if users provide their opinions about companies' activities and the government's policies.

4.4 Educational issues

Cultivating individuals' ability to utilize AI technologies

AI services/products work appropriately if users understand their benefits and risks, learn how to identify responsibilities, and operate them perfectly to keep them under control. Significant issues are need to understand the advantages and limits of the current AI technologies, to perfectly utilize AI technologies, and to perform creative activities in collaboration with AI technologies.

Enhancing essential human abilities that AI technologies cannot perform

Education policy functions according to discussions about how to efficiently reform the curriculum based on evidence that shows the limitations of technologies. For example, a deep understanding of semantics, the utilization of experience-based imagination in novel situations, the ability to identify a problem that should be solved, the ability to communicate and collaborate, and the ability to explore novel information actively and to discuss and incorporate the opinions of others are all abilities that current machine-learning AI technologies seem unable to perform, and they are expected to become more important. Enhancing these abilities differentiates humans from AI technologies and makes humans perform creative tasks by utilizing AI technologies, which leads to the realization of a sustainable society with high productivity and less labor. Education for children is especially urgent because it takes time, and the development of AI technologies is so rapid. It is important to consider what abilities should be still learned by humans for proper brain development even though the activities enabled by said abilities can be performed instead by AI technologies.

20 "Labor mobility" refers to individual workers changing jobs across companies, industries, and job categories in the labor market.

21 This issue is related to educational issues. Continuing education to update the knowledge and skills of the labor force is increasingly important in addition to education for children.

4.5 Social issues

Freedom to use (or not use) AI technologies and people's dialogue on common social values

The social benefits from AI technologies are numerous, such as the realization of social security and safety, improvement of productivity to counter labor shortages, a decreasing birthrate, an aging population, and the facilitation of participation by various people (inclusiveness) with individually optimized AI technology supports. Thus, AI technologies are crucial to the realization of Society 5.0. However, like many other tools and technologies, AI technologies' utilization cannot be socially enforced. It may be necessary to take into consideration the need to ensure the freedom to use AI technologies, based on an individual's faith, and avoiding social conflict between users and non-users of AI technologies. AI technologies work as a part of Information Technologies (IT) or software programs, so users cannot simply confirm AI services/products by their appearance. Thus, a discussion is required about whether AI technologies should be always explicit. Furthermore, Society 5.0 demands the avoidance of social conflicts between AI services/products users and non-users. This also requires continuous dialogue among people with different visions and ideas, including experts, regarding opposing opinions in order to consider common, fundamental social values.

AI divide, the unbalanced burden of social costs relative to AI, and the prevention of discrimination

To maximize the benefits from AI technologies, in addition to appropriate knowledge of the AI technologies themselves, users need digital goods and services literacy and knowledge of data privacy. However, all people cannot acquire or maintain this knowledge and literacy, and it might be a causal factor in the so-called "AI divide." For example, "rideshare," backed by AI optimization technologies, could offer a new means of transport at a low cost comparative to taxis; therefore, it is supportive of socially disadvantaged people. However, access to these services require a minimum familiarity with digital devices, so those without literacy may be excluded from the benefit of rideshare services. As ridesharing becomes popular, the traditional taxi services may become expensive or diminished. Therefore, it is necessary to take this into consideration when making policies to avoid generating an imbalanced social cost burden and a new differential caused by literacy, knowledge, and assets. Potential discrimination based on the output of personal profiling by AI technologies must be prevented.

New social pathology, conflict, and dependence on AI technologies

With increasing opportunities to use AI technologies in social contexts, there is a possibility of generating social pathology and new social problems, such as excessive rejection, overconfidence, and dependence on AI technologies. Recommendation and personal optimization by AI technologies may limit available information for individuals and increase the tendency for people to regard the limited information as universal. It is, therefore, necessary to provide accurate information and opportunities for dialogue and training.

4.6 Research and Development issues

Ethics, accountability, and visualization

Researchers and engineers are required to engage in R&D in AI-related areas with a high level of professional ethics[22] while observing the ethical codes and guidelines of their academic societies and organizations[23] and with accountability for them. AI technologies have features that users are hardly aware of; they use the technology yet do not know how it actually works inside the products/services. Thus, R&D is recommended regarding the appearance of AI technologies and to visualize how much AI technologies are used in decisions or actions.

Security, privacy protection, controllability, and transparency

Scientists and engineers are required to establish environments with robust cyber-security and safety in which to use AI technologies. It is especially essential to develop technology that enables us to choose how much personal data to share, the level of individual privacy to be protected, and what kind of information can be used publicly. R&D should be conducted to develop technologies that enable people to control the safety features of AI technologies, to explain the processes and logics of calculations inside AI technologies, to provide interfaces that smoothly perform transitions of control from AI to human, especially in emergencies.

Appropriate disclosure of information: Promoting the humanities, social sciences, and research collaboration

AI technologies based on machine learning produce statistically valid outputs, and they statistically benefit human society. For this paradigm to be accepted in society, scientists and engineers are required to explain it appropriately. When spreading new technologies, researchers and engineers might have to invest effort in explaining their benefits and risks fairly. To discuss the relationship between AI technologies and human society adequately and to design and realize a better future society, researchers in the humanities and social sciences should acquire up-to-date knowledge of new technologies and utilize them in their research. Scientists and engineers should collaborate with researchers in the humanities and social sciences for pursuing socially beneficial AI technologies.

Diversity of AI technologies for social diversity

While AI technologies are currently advancing in deep learning and machine learning,

22 “Principles in AI R&D Guidelines” have been proposed by Japan’s Ministry of Internal Affairs and Communications (MIC): (1) transparency, (2) user assistance, (3) controllability, (4) security, (5) safety, (6) privacy, (7) ethics, and (8) accountability (interim report published in April 2016; Annual report published in June 2016).

23 The Japanese Society of Artificial Intelligence published “Ethical Guidelines” in February 2017. The Institutional Review Board (IRB) on AI will be established in research institutes, universities, and companies in the near future.

there are various basic theories and technologies. In the future, new theories will emerge and further promote AI technologies. The government needs to promote basic sciences and create an environment that supports open science to enhance R&D in AI technology diversity. This will contribute to the advancement, robustness, and safety of AI technologies. Such technological diversity seems suited for social diversity.

Conclusion

Japan, with its energy and resource constraints and demographic pressure, is placed among developed countries on the front line in seeking new societal models, ensuring sustainable and inclusive growth, and maximizing the wellbeing of its citizens. To take the lead in this endeavor, the full potential of science, technology and innovation (STI) should be explored, and in this perspective, AI is considered to be a priority. Accordingly, the Japanese government proposed the concept of Society 5.0, which fully utilizes STI for humans and society.

Recognizing that AI has tremendous potential to generate benefits for society while it also may engender safety concerns or other side effects, the Advisory Board proceeded with an in-depth discussion of cases of AI technologies deployed today or likely to be developed in the foreseeable future with the aim of extracting key challenges to be addressed to ensure the intelligent use of AI for the benefit of human society.

Recent acceleration of the advancement of AI technologies makes it difficult for institutional and social adaptation to keep pace, which leads the government to address the question of transition management. The Advisory Board concluded that the government should enact policies relative to industrial competitiveness, employment, and education about social sciences for future design while promoting AI technologies.

We have already initiated international discussion on these key issues and are engaged in pursuing this process. We consider clarifying the relationship between AI technologies and human society to be a prerequisite to the realization of a sustainable society, and we believe that addressing the fundamental question, “What values are shared by humans all over the world,” is unavoidable.

The Advisory Board wishes for this report to be shared by AI researchers/engineers, governments, and business sectors as well as researchers in the humanities and social sciences, education suppliers, and the general public so they can open debates and take actions.