



Hitachi's Activities for the Ethical Use of AI in Its Social Innovation Business

Published: February 22, 2021

Lumada Data Science Lab.

Hitachi, Ltd.

Contents

Foreword	3
Introduction.....	5
1. The Context of AI Ethics	6
1.1 Background	6
1.2 Status of rules for AI ethics	10
2. About Hitachi's Principles guiding the ethical use of AI	13
2.1 Hitachi's activities to establish <i>Principles guiding the ethical use of AI</i>	13
2.2 Hitachi's basic concepts on <i>Principles guiding the ethical use of AI</i>	14
2.3 Key points when implementing AI in society	16
3. Hitachi's Activities for Observing the <i>Principles guiding the ethical use of AI</i>	22
3.1 Specific implementation of the <i>Principles guiding the ethical use of AI</i>	22
3.2 About internal AI ethics training	24
3.3 Future policy for revising the <i>Principles guiding the ethical use of AI</i>	25
Conclusion	26
Sources.....	27

Foreword

Hideaki Koizumi
Honorary Fellow, Hitachi, Ltd.

Homo sapiens, Earth's most prosperous species, reached the apex of the living world by forming an advanced society. The forming of a society must be founded not only on self-interest, but also on consideration for the positions and lives of others, that is, altruism. Altruism is linked with the concept of coexistence, especially with others regardless of time and space. It appears that concern for others, in other words, *warmheartedness*, is the essence of ethics. Aristotle left his *Nicomachean Ethics* around the concept of *phronesis*, and *The Metaphysics of Morals*, written by Kant in his twilight years, asserts that our duty is to promote the happiness of others.

Around a century ago, when Hitachi, Ltd. was first born in the Hitachi mine, a group of kindhearted people, whose primary consideration was the mine's coexistence with the surrounding population, toiled diligently to build a 156-meter smokestack on the mountain. Numerous areological observation huts (sometimes sondes were used) were placed in the surrounding mountains, and partially linked up to exhaust fume statuses and refinement and production amounts to build a management system that was open to the wider society. This system is well-known as a monumental achievement in Japan's modern environmental history.

This mine was where power systems driven by electricity and air pressure obtained through generation and transmission, electric elevators and trolleys, systems for finding mineral deposits, information and communication systems, and cultural facilities; namely, theaters, hospitals, and schools, were developed. In lofty terms, the aim of these activities was to promote *Human Security and Well-Being* in that regional community. This is likely where Hitachi established its philosophy of *coexistence between enterprises and society*.

The heart behind these activities did not stop beating there, but continued to pulse through the Hitachi Group, leading to the *Hitachi Social Innovation* initiative. While the world outside Japan saw the inception and spread of *Industry 4.0* by the German Academy of Technical Sciences (acatech), Japan birthed its own original concept of *Society 5.0*, incorporating the perspective of *society and humans* into the digital revolution. Hitachi's ideas, through the concept of *Society 5.0*, are now being transmitted all over the world.

Meanwhile, Japan contributes to the activities of the United Nations, one such activity being *STI for SDGs* (Science, Technology, and Innovation for Sustainable Development Goals). Hitachi is expected to contribute its own AI (artificial intelligence) with a clarified set of ethics and goals.

Finally, the idea *think globally, act locally* has led to the composition of this document, *Hitachi's*

Activities for the Ethical Use of AI in Its Social Innovation Business. This document brims with the passion of Hitachi's young talent earnestly tackling the issue of AI ethics. I hope that, with the wise counsel of many others, we can raise ourselves to an even higher standard than before.

With sincerest gratitude, I look forward to your continued guidance and support.

Introduction

Ever since the term was coined in the Dartmouth Conference of 1956, AI has spawned endless research and development. In the more recent context of vast, ever-growing amounts of data, AI is now being used in countless areas of business and everyday life, for example through machine learning and deep learning.

Anticipating that AI will be used for more advanced and diverse purposes in the future, Japan and other countries are considering conceptual frameworks and rules for using AI while taking ethical considerations into account.

The Japanese Cabinet Office's *Social Principles of Human-Centric AI* gives basic principles for handling AI as important elements in realizing economic development and solutions to social problems through the *Society 5.0* initiative. The EU similarly gives an ethical framework for using AI in its *Ethics guidelines for trustworthy AI*.

In the private sector, relevant organizations, and academic societies, guidelines on the handling of AI are being compiled based on a variety of stances on the development and use of AI.

This white paper introduces the Hitachi, Ltd. (hereinafter Hitachi) model of ethically conscious development and use of AI known as *Principles guiding the ethical use of AI in the Social Innovation Business* (hereinafter *Principles guiding the ethical use of AI*). Also introduced are some specific activities in line with these *Principles guiding the ethical use of AI*.

Hitachi's *Principles guiding the ethical use of AI* defines model concepts and actions. These incorporate knowledge on AI technologies from over 50 years of IT (Information Technology) achievement, and knowledge on ethics as it relates to science and technology drawn from high-quality OT (Operational Technology) and research in brain science and other areas.

To deal swiftly and accurately with increasingly diverse and complex problems and to promote its Social Innovation Business, Hitachi is developing, using, and applying trustworthy AI while observing its *Principles guiding the ethical use of AI*. In doing so, Hitachi is realizing a comfortable, resilient, and sustainable society that protects human dignity, and helping to improve quality of life around the world.

We hope that this document is widely read and serves as a useful reference to its readers.

AI (artificial intelligence): A processing module or a system that is constructed based on such processing module(s), that is capable of independently changing its output or processing method based on the data, information or knowledge learned.

1. The Context of AI Ethics

1.1 Background

(1) How AI contributes to society

In the past, Hitachi has sparked social innovation by developing systems in a variety of sectors including finance, medicine, electricity, rail, and water supply, thereby contributing to society-wide development. Most of these systems are supported by a combination of advanced operations relying on human judgment and rule-based automatic control.

Rule-based automatic control requires quantified rules. Advanced operations relying on human judgment largely depends on tacit knowledge such as experience and intuition, making it difficult to define clear rules and posing an obstacle to automation.

AI is now overcoming this obstacle to automation by using mechanisms such as deep learning to learn tacit knowledge that was previously only accessible to experts, enabling systems to autonomously make advanced decisions.

With its expert-level decisions, AI is expected to bring advancements in transport and production, as well as greater efficiency and sophistication in a variety of social settings. From an energy efficiency perspective, AI is also expected to help protect the natural environment. Moreover, AI will surely reduce the need to dispatch workers to dangerous sites by eliminating the need for human staff and automating systems, thus contributing to safety.

In Hitachi, AI has had a profound impact on problems that cannot be solved by humans alone, such as how to detect the vast garbage patches in the ocean, or to significantly improve the accuracy and throughput of blood-testing devices.

Recently, advancements in machine learning and deep learning have been widening the scope of AI use to the prediction and evaluation of human actions, as well as areas closely related to everyday life and safety such as social infrastructure.

Looking upon on our lives, we find that AI is already used in a variety of everyday settings, such

as recommendation functions on ecommerce sites and social media, camera-based facial recognition, and elevator control; AI is no longer a rarity. Even in business, the idea of AI as special or cutting-edge is becoming a thing of the past, and there is a growing need to use and apply AI to transform our businesses.

As AI becomes commonplace in a variety of social settings, the types and amounts of collected data are expected to dramatically increase in the future. We anticipate that this data will yield a great many benefits, further improving quality of life.

(2) Ethical concerns regarding advanced use of AI

Meanwhile, as the use of AI diversifies and becomes more advanced, AI is facing more scrutiny regarding its numerous ethical problems.

For example, concerns have been raised about human life or health being threatened by operations unintended by the developers, or system interruptions having major society-wide ramifications. Furthermore, in areas where personal information is processed, there have been cases of AI decision results effectively exacerbating discrimination, prejudice, and inequality. This is becoming a major social problem.

These problems mainly arise through These problems mainly arise through its mechanism of autonomously updating its decisions through learning, a mechanism characteristic of AI.

Although several concerns have been raised about AI from an ethical perspective, this document will feature four representative examples: *AI-driven automation and unmanned operation*, *quality of training data*, *addressing environmental changes*, and *ensuring transparency*.

The first example, *AI-driven automation and unmanned operation* were originally considered a major benefit of AI. Technological advancements are dramatically expanding the proportion of tasks formerly done by humans that can be taken over by AI. However, as *AI-driven automation and unmanned operation* replace human jobs in the realm of decision-making, it might become unclear where responsibility lies. Furthermore, there is still a need to debate whether decision-making related to human life and dignity can be left to AI alone.

To this end, we must clarify where responsibility lies in *AI-driven automation and unmanned operation*, and establish mechanisms allowing humans to intervene as necessary and periodically evaluate and revise automated processing.

The second example, *quality of training data*, is a concern that can occur during optimization, where evaluation criteria are changed based on *correct* data learned by AI. In past rule-based systems, input information was processed on rules pre-defined in development, so subsequent output information could be foreseen.

On the other hand, with AI, which characteristically uses learning-based optimization, low-quality training data, that is, biased or inappropriate data, can result in output at a later stage that could not be foreseen during development.

For this reason, AI should be designed from a different perspective than conventional systems.

The third example, *addressing environmental changes*, is the concern of the definition of *correct* changing according to transitions in the social environment. Unlike changes caused by AI training

data, this problem concerns changes on the user side.

Even if the quality of training data described in the second example was ethically sound at its time of creation, inappropriate decisions might be made while the AI is optimizing itself to new conditions.

Because this problem can occur in association with changes in social environment, it is important to monitor the definition of *correct* in the training data as with input and output data, and to devise operation management methods accordingly.

The fourth example, *ensuring transparency*, is the concern of difficulty identifying the causes of problems that occur in AI. Because AI optimizes evaluation rules based on training data, even if output results are found dubious, it is extremely difficult to analyze the causes leading up to the issue, and clarify what to improve or where responsibility lies. This difficulty is characteristic of AI.

XAI (Explainable AI; a technology for explaining AI functionality) is currently being considered as a solution to this problem, but advanced considerations are still needed, and this solution would face several hurdles before coming into general use.

(3) The expanding applicability of AI

Although the aforementioned concerns regarding the use of AI have mainly been debated in the IT area, AI is being increasingly used in the OT area relating to human safety and social infrastructure.

In typical OT areas such as the management of operations in the railway sector, optimization of power transmission and distribution plans in the energy sector, and advancement of equipment maintenance in health care and manufacturing, major decisions regarding social life and human safety are made primarily by humans. While the importance of human judgment will not change, if AI is expected to play an important role in decision-making in the future, special attention must be paid to the preceding four concerns.

Additionally, OT concerns many systems such as those controlling large machinery where lapses in judgment can lead to major incidents involving injury or death. This raises concerns about the hasty application of AI, which changes its rules autonomously as it learns. Mechanisms are needed for the use of AI that can be trusted with such operations.

1.2 Status of rules for AI ethics

In response to the expanding use of AI and its accompanying concerns, the governments of various nations, international institutions, and private organizations are considering rules for promoting the proper use of AI, including AI ethics.

There have been initiatives to compile what might be called *principles* describing the basic concepts behind the use of AI, the first of which is said to be the September 2016 *Tenets* released by the organization *Partnership on AI*. After the Future of Life Institute (FLI) announced the *Asilomar AI Principles* in January 2017, considerations also began on a national level, such as in Japan and some European countries. In March 2019 in Japan, the Cabinet Office established the *Social Principles of Human-Centric AI* as an AI philosophy for realizing *Society 5.0*.

May 2019 saw the release of the first set of AI principles agreed upon by multiple nations, the *OECD Principles on AI*. In June that same year at the G20 Summit, the *G20 AI Principles* were agreed upon as an accompanying document to the summit declaration.

Subsequently, additional initiatives were launched to elaborate on these principles. In Europe in April 2019, the European Commission presented the *Ethics guidelines for trustworthy AI*. In Japan, the Ministry of Internal Affairs and Communications announced the *AI Utilization Guidelines* at the Conference toward AI Network Society.

At this stage, there is no comprehensive legal system regulating AI, but as indicated in the European Commission's AI white paper *White Paper on Artificial Intelligence: a European approach to excellence and trust*, efforts are underway to present regulations specific to certain sectors such as employment and remote biometrics. There are cases, like in San Francisco, where the use of facial recognition technology by public institutions is regulated by the law.

Table 1: Main principles and guidelines on AI ethics

Title	Publisher	Overview
<i>Tenets</i> (September 2016)	Partnership on AI	This document presents eight tenets such as <i>pursuit of human interests, participation of related parties, and open investigation and dialogue</i> as AI concepts to be observed by members.
<i>Asilomar AI Principles</i> (January 2017)	FLI	This document presents 23 principles as research topics, ethics and values, and long-term challenges that must be addressed for AI to bring prosperity to people's lives.
<i>Social Principles of Human-Centric AI</i> (March 2019)	Cabinet Office (Japan)	In anticipation of AI's contribution to realizing <i>Society 5.0</i> , this document presents <i>a society that respects human dignity, a society where people from diverse backgrounds can pursue diverse forms of happiness, and a sustainable society</i> as three values to be observed as a guiding philosophy.
<i>Ethics guidelines for trustworthy AI</i> (April 2019)	European Commission	As a framework to realize trustworthy AI, this document presents three elements that serve as criteria throughout the entire system lifecycle: <i>legal, ethical, and resilient</i> . This document also shows that harmony among these elements is fostered by the efforts of society.
<i>OECD Principles on AI</i> (May 2019)	OECD	This document presents five principles for the responsible operation of trustworthy AI including <i>comprehensive growth, sustainable development, and the pursuit of happiness</i> and <i>laws, human rights, democracy, and respect for diversity</i> , and urges AI providers to promote and implement these principles.
<i>G20 AI Principles</i> (June 2019)	G20	This document presents the following as G20 AI principles: <i>principles for responsible stewardship to realize trustworthy AI, domestic policy and international cooperation to realize trustworthy AI</i> .
<i>AI Utilization Guidelines</i>	Ministry of Internal Affairs	This document summarizes items for users to keep in mind when using AI, compiling 10 principles for

(August 2019)	and Communications (Japan)	specific measures expected to be implemented, such as <i>proper use, safety, security, and fairness</i> .
<i>White Paper on Artificial Intelligence</i> (February 2020)	European Commission	This document presents risks and examples of ideal use while touching on topics such as <i>direction by humans, trustworthiness and safety, and privacy</i> as the ideal image of innovation in the use of AI and the building of trustworthy ecosystems.

2. About Hitachi's *Principles guiding the ethical use of AI*

2.1 Hitachi's activities to establish *Principles guiding the ethical use of AI*

(1) Hitachi's past efforts to foster AI ethics

Until now, Hitachi has striven to foster an ethical code on AI technologies as Hitachi solves a variety of problems in society and enterprises, based on many years of experience conducting AI technological research and product development in the IT and OT areas.

Even in Lumada, Hitachi's initiative for accelerating digital innovation, a high ethical code on the use of AI is proving essential. This perspective must be considered when creating new value by combining on-site experience with the ability to capture data on and flexibly respond to changes in the world.

With this high ethical code at its foundation, Lumada can be applied to a wide range of areas. For example, Lumada has already provided many solutions in finance, industry, distribution, electricity, and mobility, and Hitachi is working to accelerate the further development, distribution, and use of AI technologies going forward.

(2) Contributing to institutional considerations such as in governments and international organizations

Hitachi has already been contributing to institutional considerations such as in public institutions and principal organizations, leveraging its background knowledge in the societal implementation and research and development of AI.

In Japan, Hitachi has been participating in the Conference toward AI Network Society held by the Ministry of Internal Affairs and Communications since 2016, contributing to discussions on the ideal future form of AI included in the aforementioned *AI Utilization Guidelines*.

Internationally, Hitachi has participated in ISO/IEC JTC 1/SC 42 as a member of ISO (International Organization for Standardization), and has held debates on the international standard of use of AI and machine learning. Hitachi works daily to stay abreast of the latest knowledge on AI ethics discussions inside and outside Japan, for example by submitting memorandums to the European Commission and participating in the WEF (World Economic Forum).

2.2 Hitachi's basic concepts on *Principles guiding the ethical use of AI*

This section introduces three basic concepts that Hitachi considers prerequisites of the *Principles guiding the ethical use of AI*.

- (1) AI must contribute to realizing a sustainable society.**
- (2) AI must permeate society based on human-centric thought.**
- (3) AI must continuously and flexibly adapt to social and environmental changes over the long term.**

(1) AI must contribute to realizing a sustainable society.

To realize a sustainable, diverse, and inclusive society that leaves no one behind, the September 2015 United Nations Summit adopted the *SDGs* (Sustainable Development Goals).

Hitachi abides by these concepts even in its development and application of AI, believing that it is necessary to contribute to realize a sustainable society by creating social, environmental, and economic values through providing solution and products..

Hitachi conducts design, development, and testing based on clearly defined goals and effects of using and applying AI, to enable AI to solve diverse, society-wide problems and help improve human quality of life.

(2) AI must permeate society based on human-centric thought.

The use of AI must not violate privacy or infringe on freedom, fairness, equity, or other matters of individual rights and dignity. Because the Social Innovation Business Hitachi promotes aims to improve the quality of life of the individuals constituting society, it is crucial to ensure values such as privacy and fairness when using AI.

While continuing to work to protect privacy, Hitachi also pledges to respect human rights within the *Hitachi Group Codes of Conduct*. as well as not to discriminate based on factors such as gender, sexual orientation, age, nationality, race, ethnicity, ideology, creed, religion, social status, family origin, illness, or disability in its diverse range of enterprise activities. In the implementation of AI in society, Hitachi will strive to ensure AI's acceptance in society by following this philosophy, placing human-centric thinking at the core, and performing evaluation and testing to ensure that the AI does not infringe on individual rights or dignity..

(3) AI must continuously and flexibly adapt to social and environmental changes over the long term.

People's values and philosophies change along with their society and environment. Systems applying AI must maintain universal values while continuing to provide high value in line with

changes in people's values, without working against human interests based on antiquated values.

For this reason, Hitachi believes it is important to accurately recognize social and environmental changes and maintain and manage AI lifecycles to ensure AI quality over the long term.

2.3 Key points when implementing AI in society

Hitachi contributes to social innovation through its IT systems, such as telecommunications, and its OT systems for social infrastructure, industry, and distribution such as power, water supply, and railways. While past applications of AI have focused on IT systems, OT systems will be increasingly considered going forward. In view of this, this section introduces seven key points regarding Hitachi's activities in the consideration of specific implementations, while touching on the differences between IT and OT.

① Safety

In all systems, regardless of the distinctions between IT and OT systems, ensuring safety is a universal Hitachi value. Even when using AI, safety evaluation and measures must be performed while considering the respective features of IT and OT.

Especially in OT systems involved in the control of large equipment such as power generators, railway cars, and machine tools, the control of equipment is usually closely linked with human safety. Even in the IT area, systems such as clinical decision support systems and electronic commerce systems have an impact on major decisions made regarding human life and property.

Leveraging its knowhow gained from experience developing countless systems in the past, Hitachi is continuously taking measures to guarantee the necessary level of safety and security when applying AI. This includes combining concepts such as incorporating a *safety over efficiency* mindset in AI, fail-safe mechanisms for safety controls in the event of erroneous AI decisions, fail-soft systems for appropriate degradation, and fault tolerance for maintaining functionality.

② Privacy

The proper handling of personal data is a particularly important factor in IT systems. To realize AI that does not infringe on individual privacy and interests, Hitachi applies its knowhow regarding privacy measures gained over years of experience to continuously ensure effective privacy protections. With the application of AI in a variety of settings, privacy will become even more important in the future, as with security (mentioned later in this document). For this reason, many measures will be needed when using personal data in learning data, for example the minimization of such use, processing to prevent identification of individuals, data encryption, and restriction of staff authorized to access data.

When applying AI to OT systems, most information handled by AI involves equipment operations, for example the temperature and voltage of controlled equipment. However, privacy still cannot be ignored; privacy must be considered in some cases such as solutions that use AI analysis of camera footage to assist human labor.

③ Fairness, equality, and Prevention of discrimination

AI decision results must be evaluated to ensure that they do not promote discrimination or prejudice on the basis of race, gender, or other factors. As mentioned previously, Hitachi pledges to respect human rights within the *Hitachi Group Code of Conduct*, and pledges not to discriminate in its diverse range of enterprise activities.

To this end, both the input and output of learning data must be considered. For input, Hitachi is devising mechanisms to exclude bias from types and volumes (by element) of learning data, and data defined as correct. For output, Hitachi considers the possibility of unintended discrimination or prejudice resulting from AI-learned data and definitions of *correct*, and continuously performs characteristic evaluations and measures in each system to ensure fairness. This practice covers not only AI design and development, but also the building of comprehensive mechanisms including those for operation. In addition to ensuring the standards required globally, Hitachi strives to be flexible and maintain fairness in its AIs based on the customs and characteristics of regions where systems are applied.

④ Proper and responsible development and use

Hitachi presents specific and long-term system use cases to prevent the use of AI outside its original purpose; that is, use that deviates from the purpose of the design and operational conditions. Hitachi is working to eliminate the possibility of abuse or erroneous use of AI.

For the proper and responsible development and use of AI, it is important to revise functions not only at the beginning of operation, but also continuously according to changes in social circumstances and service details.

For example, IT systems allow relatively many opportunities to revise entire services in the operation cycle, allowing fast, flexible, on-the-spot adaptation of applied AI to functional expansions to target systems and changes in values.

Meanwhile, some OT systems, in particular social infrastructure systems, are assumed to operate continuously 24 hours a day, 365 days a year. This means that opportunities to revise AI can be limited after operation has commenced. In addition, the operation cycle is longer than for general IT systems, often lasting 10 to 20 years or longer. Measures must be devised to account for this, for example restricting learning data at the time of development to elements unlikely to change over the long term, and limiting the scope in which decision results change as the AI learns.

⑤ Transparency, Explainability, and Accountability

Hitachi acts with utmost care regarding the aforementioned *safety, privacy, fairness, equality, and prevention of discrimination, proper and responsible development and use*, and other considerations. However, Hitachi is aware of the difficulty of completely and continuously adhering to such principles amidst technological advancements and constantly changing social conditions. For this reason, Hitachi is continuously verifying the basis of AI decision results and ensuring transparency (such as explainability) to improve future decisions based on verifying results. Hitachi believes this provides long-term peace of mind regarding the use of AI, and is a prerequisite for AI to contribute to realizing a sustainable society.

Through the *Hitachi Group Code of Conduct*, in order to develop and maintain relationships of trust with a variety of stakeholders, Hitachi pledges to conduct fair and highly transparent disclosure of information. A characteristic problem of AI is the extreme difficulty of showing the process by which output results are derived. However, Hitachi will strive to overcome this problem through XAI research and other initiatives, thus ensuring accountability even in the realm of AI.

⑥ Security

To operate AI properly, security considerations such as *preventing data leaks, data tampering manipulation, and data destruction* are important elements in both IT and OT systems.

Hitachi pledges to fight threats such as cyberattacks and take the proper organization-wide measures, within the *Hitachi Group Code of Conduct*.

To address cyberattacks, a variety of measures must be taken according to the characteristics of each system. For example, IT systems generally place particular emphasis on ensuring confidentiality, while OT systems emphasize ensuring availability. For security measures, a variety of institutions have presented guidelines, for example NIST (National Institute of Standards and Technology) and NISC (National center of Incident readiness and Strategy for Cybersecurity). In addition, Hitachi is drawing from its years of experience with security measures to provide continuous, system-wide protection by implementing a variety of analyses and measures according to the characteristics of each system.

⑦ Compliance

As mentioned previously, there are a variety of guidelines and related rules for the use of AI. Hitachi acts with considerations about those rules. Also, Hitachi is working to build and operate AI in compliance with the applicable treaties, laws, and ordinances of target countries and regions.

(3) About Hitachi's *Principles guiding the ethical use of AI*

Hitachi has established and begun implementing the *Principles guiding the ethical use of AI in Social Innovation Business*. These principles will comprehensively cover the aforementioned knowledge Hitachi has gained from the activities of a variety of AI-ethics-related organizations inside and outside Japan, and key points required of IT and OT systems.

- Hitachi's philosophy on AI ethics

The preamble to Hitachi's *Principles guiding the ethical use of AI* presents Hitachi's philosophy on AI ethics. This shows that Hitachi aims to contribute to solving social problems and improving quality of life through the use of AI. Hitachi defines its principles by integrating an understanding of ethics as it relates to science and technology drawn from research in brain science and other areas, and AI knowledge cultivated over years of experience with IT and OT systems.

Table 2: Preamble to Hitachi's *Principles guiding the ethical use of AI*

It is the aim of Hitachi, Ltd. (hereinafter *Hitachi*), through the proper use of AI in its Social Innovation Business, to further enhance social, environmental, and economic values while contributing to the creation of a comfortable and sustainable society based on human dignity and higher quality of life (QoL) around the world.

To realize this, Hitachi will promote the ethical use of AI by integrating an understanding of ethics as it relates to science and technology drawn from research in brain science and other areas, and knowhow in AI technology based on over 50 years of achievements in IT (information technology), in the high quality OT (operational technology; control and operation) which Hitachi has worked to provide for more than a century. The use of AI at Hitachi will be based on these guiding principles which include standards of conduct to be respected in each of the three stages of planning, societal implementation, and maintenance, as well as seven items to be addressed which are commonly applied to all three stages.

- Three standards of conduct for the use of AI

Hitachi's *Principles guiding the ethical use of AI* present the following three standards of conduct as standards for the development, implementation, and long-term use of AI based on the concepts given in the preceding section 2.2. The use of AI is essential to promoting the Social Innovation Business. The *Three standards of conduct* define policies for suppressing a variety of inherent risks in AI and continuously achieving a high level of safety and security.

Table 3: Three standards of conduct in Hitachi's *Principles guiding the ethical use of AI*

I Development and use of AI will be planned for the realization of a sustainable society

It is important to ensure that the reason for using AI in services, solutions or products is appropriate from the planning stage, in order to suppress the inherent ethical risks in AI while generating new value. Hitachi will use AI to resolve issues in society, realize a comfortable, resilient, and sustainable society and to improve the quality of life of people around the world.

II AI will be societally implemented with a human-centric perspective

To ensure that decisions made by AI respect the rights of individuals and contribute to the interest of society, it is important that AI is societally implemented in a responsible manner and its harmonious coexistence with humans is ensured. Hitachi will societally implement AI from a human-centric perspective according to the principles of freedom, fairness and equity, and endeavor to verify that it functions as intended.

III Maintain and manage AI to sustain provided value over the long term

It is important that the AI continues to consistently provide value over the long term after it is societally implemented. Hitachi will endeavor to maintain and manage the value provided by the AI in a way that is responsive to and acceptable to societal and environmental changes.

- Seven items to be addressed for observing the *Principles guiding the ethical use of AI*

Seven items to be addressed are defined for observing the aforementioned three standards of conduct, based on the key points given in the preceding section 2.3. Hitachi applies these items to areas including the research and development of actual IT and OT systems.

Table 4: Seven items to be addressed for observing Hitachi's *Principles guiding the ethical use of AI*

<p><u>1. Safety</u></p> <p>Hitachi will endeavor to realize and manage AI that protects human rights including health and lives, property, dignity, reliability and trust of users and related parties, by verifying that AI and the system or solution using the AI operates as with the intended quality. Furthermore, Hitachi will at the same time endeavor to realize and operate AI that prevents the destruction or deterioration of the global environment and provides a safe life to people.</p> <p><u>2. Privacy</u></p> <p>Hitachi will endeavor to realize and manage AI that protects privacy by ensuring the proper handling of personal data used as input data for AI learning, evaluation, and operation as well as of the output data from AI.</p> <p><u>3. Fairness, Equality, and Prevention of discrimination</u></p> <p>Hitachi will endeavor to realize and manage AI that works for the interests of a diverse and inclusive group of stakeholders, while ensuring that it prevents discrimination or avoids bias based on factors such as race, gender or citizenship.</p> <p><u>4. Proper and responsible development and use</u></p> <p>Hitachi will endeavor to ensure the proper use of AI by pursuing development based on the potential risks in each specific use case so that it does not deviate from the purpose or operating conditions for which it was designed. Further, Hitachi will endeavor to ensure proper AI operation by informing users and operators of the use policy, conditions and so on. In addition, Hitachi will endeavor to ensure proper AI maintenance and management by continuing to check for any changes in the operating conditions, mindset of people in relation to the AI and its decision, and conditions in society.</p> <p><u>5. Transparency, Explainability and Accountability</u></p> <p>Hitachi will endeavor to ensure the transparency of the AI by validating and ensuring that</p>

the reasons behind the decision results produced are explainable. Hitachi will endeavor to be accountable for AI and its decision in response to the purpose and conditions of use of AI.

6. Security

Hitachi will endeavor to realize security-oriented AI and its operations including countermeasures at the system or operations level to prevent information leaks, data tampering, system crashes and service interference.

7. Compliance

Hitachi will realize AI and its operations that comply with the applicable laws and regulations of the countries and regions in which the AI is to be used.

3. Hitachi's Activities for Observing the *Principles guiding the ethical use of AI*

3.1 Specific implementation of the *Principles guiding the ethical use of AI*

Hitachi is engaged in activities to bring the *Principles guiding the ethical use of AI* into use. For example, Hitachi is helping apply the principles to actual systems through a professional team in LDSL (Lumada Data Science Lab), the center of numerous advanced Collaborative Creation projects in the Social Innovation Business. Internal training sessions are also being held.

Specifically, Hitachi evaluates in advance whether AI deployment and operation will follow the *Principles guiding the ethical use of AI*, and ensures proper operation through the relevant lifecycle. Checklists and related documents for efficient and effective evaluation are provided and disseminated throughout the company (as described later).

To run its dedicated team, Hitachi receives help from its internal privacy protection advisory committee^{#1} to make efficient use of expert knowledge on privacy as well as Hitachi Consulting Co., Ltd., which provides AI ethics consulting^{#2}. This effectively aggregates the AI ethics knowledge of the entire Hitachi Group.

Additionally, to objectively evaluate activities relating to AI ethics and improve operation, Hitachi has founded the AI Ethics Advisory Board consisting of experts from outside the company. The AI Ethics Advisory Board not only makes suggestions for projects, but also gives advice on improving activities relating to the *Principles guiding the ethical use of AI*.^{#3}

(Reference) AI Ethics Advisory Board members (as of February 2021, in order of the fifty Japanese syllables)

Fumio Shimpo	(Professor at the Faculty of Policy Management, Keio University)
Shoko Suzuki	(Professor at the Graduate School of Education, Kyoto University)
Hiroshi Nakagawa	(Team Leader at the RIKEN Center for Advanced Intelligence Project (AIP))
Miki Nagata	(Representative of the Telecommunications Consumer Network)
Ryoji Mori	(Attorney at Cyber Law Japan Eichi Law Offices)

-
- #1 Hitachi's Privacy Activities Regarding the Use of Personal Data (Hitachi, Ltd.)
https://www.hitachi.co.jp/products/it/bigdata/bigdata_ai/personaldata_privacy/
 - #2 AI ethics consulting (Hitachi Consulting)
https://www.hitachiconsulting.co.jp/business/public/ai_logic/
 - #3 The AI Ethics Advisory Board was founded in February 2021. While the Board was not involved in creating the *Principles guiding the ethical use of AI*, future plans to make overall improvements to AI ethics activities, including the principles, will be implemented according to the Board's advice.

3.2 About internal AI ethics training

To introduce an appropriate ethical code in the societal implementation of AI, it is not enough merely to establish organizations and systems. It is also important to ensure that all parties involved in AI understand the importance of AI ethics and put AI ethics into practice.

For this reason, Hitachi is promoting internal training on the checklists, workflows, and related documents for explaining and implementing the *Principles guiding the ethical use of AI* as an initiative to disseminate AI ethics throughout the company. This includes the definition of AI, certain guidelines regarding factors not easily quantified such as discrimination and transparency, and how to approach system consideration.

Hitachi also conducts practical training on the potential problems and social impacts of AI, legal systems, and business trends. Possible measures are taught from the viewpoint of *who* must implement them *when*, and *how* to apply them, thus deepening employees' understanding of the importance of AI ethics.

Because the perspectives of fairness and transparency are particularly difficult to incorporate into prior considerations for OT systems, meticulous training is conducted with the aid of case studies on actual problems, including problems staff are expected to encounter.

These activities include lectures by experts and eLearning courses for widespread dissemination. Hitachi will strive to deploy AI implementations not only in Japan but worldwide, while tailoring implementations to the circumstances of each country.

In the future, Hitachi aims to share its philosophy of AI ethics with society as a whole, and to ramp up its activities to realize a better AI society. For example, Hitachi will conduct dissemination activities targeting customers and parties outside the company.

3.3 Future policy for revising the *Principles guiding the ethical use of AI*

As philosophies behind AI ethics change with social conditions, it is important to continue updating these *Principles guiding the ethical use of AI* according to the rules and current events of each country.

Hitachi will revise the *Principles guiding the ethical use of AI* roughly once every six months, led by LDSL. These revisions will account for the status of application of AI ethics leading up to the present, problems occurring inside and outside Hitachi, and the status of rules established or revised in national and international institutions and industries.

These principles are re-published upon each revision. Hitachi will continue to promote widespread awareness of its latest activities as it works to build better AI from an ethical perspective.

Conclusion

AI has already grown to become an essential part of people's lives, and AI technology has continued to advance. This document summarizes the AI ethics concepts and activities Hitachi has been working on to address this reality.

AI will come to be used for many more purposes, and ethical codes surrounding AI are expected to continue changing into the future.

Going forward, Hitachi will maintain a high ethical code as it develops and applies AI technologies as a key player in the Social Innovation Business. Hitachi will spread its philosophy on AI ethics far and wide, continuing to improve itself while incorporating a diverse range of perspectives with the aim of permeating our entire society with even more beneficial AI.

Sources

EC: *Ethics Guidelines for Trustworthy AI*

<https://ec.europa.eu/futurium/en/ai-alliance-consultation>

EC: *White Paper on Artificial Intelligence: a European approach to excellence and trust*

https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

FLI: *Asilomar AI Principles*

<https://futureoflife.org/ai-principles-japanese/>

G20: *G20 AI Principles*

<https://www.g20-insights.org/wp-content/uploads/2019/07/G20-Japan-AI-Principles.pdf>

OECD: *OECD Principles on AI*

<https://www.oecd.org/going-digital/ai/principles/>

Cabinet Office: *Social Principles of Human-Centric AI*

<https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/aigensoku.pdf>

Ministry of Internal Affairs and Communications: *AI Utilization Guidelines*

https://www.soumu.go.jp/main_content/000624438.pdf

Hitachi, Ltd.: *Hitachi's Privacy Activities Regarding the Use of Personal Data*

https://www.hitachi.co.jp/products/it/bigdata/bigdata_ai/personaldata_privacy/

Hitachi Consulting: *AI Ethics Consulting*

https://www.hitachiconsulting.co.jp/business/public/ai_logic/

About this document

As of February 2021, this document summarizes activities for the ethical use of AI, focusing on Lumada Data Science Lab., which gathers many cutting-edge collaborative-creation projects in the Social Innovation Business.

Information contained in this document might be subject to change without notice.

Hitachi's Activities for the Ethical Use of AI in Its Social Innovation Business

First published: February 22, 2021

Author	Hitachi, Ltd. Hitachi Consulting Co., Ltd.	Lumada Data Science Lab.
Editor and publisher	Hitachi, Ltd. 1-280 Higashi-Koigakubo, Kokubunji, Tokyo 185-8601	Lumada Data Science Lab.
