

Comparison of AI guidelines

Values to be respected	AI Utilization Guidelines	Draft AI R&D guidelines for international discussions	Social Principles of Human-centric AI	Ethics Guideline for Trustworthy AI	Recommendation of the Council on Artificial Intelligence	Ethically Aligned Design	Asilomar AI Principles	Tenets
by	The Conference toward AI Network Society (MIC) /Japan	The Conference toward AI Network Society (MIC) /Japan	Integrated Innovation Strategy Promotion Council. (Council of Social Principles of Human-centric AI) /Japan	European Commission (High Level Expert Group on AI (AI HLEG))	OECD	IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems	Future of Life Institute (FLI)	Partnership on AI
Released on ..	2019/8/9	2017/7/28	2019/3/29	2019/4/8	2019/5/22	2019/3/25(1st edition)	2017/2/	2016/9/28
Previously released on:...(Draft doc, etc.)	2018/7/17 (Draft AI Utilization Principles)		2018/12/27 (Draft)	2018/12/18 (Draft)		2016/12/13(ver.1), 2017/12/12(ver.2)		
URL	(Overview) http://www.soumu.go.jp/main_content/000637844.pdf (JPN) http://www.soumu.go.jp/main_content/000637097.pdf	http://www.soumu.go.jp/main_content/000507517.pdf (JPN) http://www.soumu.go.jp/main_content/000499625.pdf	https://www.cas.go.jp/jp/seisaku/inkouchinou/pdf/humancentricai.pdf	https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai	https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0448	https://ethicsinaction.ieee.org/	https://futureoflife.org/ai-principles/	https://www.partnershiponai.org/tenets/
Structure	Purpose Basic philosophies (7) AI utilization principles (10) and their Comments	Purpose, Basic philosophies (5), AI R&D principles (9) and their comments	2. Basic philosophies (3) 3. Vision (5) 4.1. Social Principles of AI (7) 4.2. AI R&D and Utilization Principles etc.	1. Foundations of Trustworthy AI(4 Principles(P)) 2. Realising Trustworthy AI: Requirements(R: 7)+Technical and non-technical methods 3. Assessing Trustworthy AI, ...	Common understanding of terms 1. Principles for responsible stewardship of trustworthy AI (5) 2. National policies and international co-operation for trustworthy AI (5)	pillars (3) General Principles (GP: 8) Chapter (11 including GPs)	Principles (23)	Tenets (10)
Human-centered	Basic Philosophies To achieve a human-centered society where all human beings across the board enjoy the benefits from their life in harmony with AI networks, while human dignity and individual autonomy are respected.	Basic Philosophies 1. To achieve a human-centered society where all human beings across the board enjoy the benefits from their life in harmony with AI networks, while human dignity and individual autonomy are respected.	2(1). Dignity: A society that has respect for human dignity We need to construct a society where human dignity is respected and, by using AI as a tool, a society where people can better demonstrate their various human abilities, show greater creativity, engage in challenging work, and live richer lives both physically and mentally.	P1: Respect for human autonomy The fundamental rights upon which the EU is founded are directed towards ensuring respect for the freedom and autonomy of human beings. Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process..... The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems.	1.2. Human-centred values and fairness a) AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognised labour rights.			
Human dignity	7) Principle of Human Dignity and Individual Autonomy Users should respect human dignity and individual autonomy in the utilization of AI systems or AI services. • Attention to the manipulation of human decision-making, emotions, etc. by AI • Reference to the discussion of bioethics, etc. in the case of linking AI systems with the human brain and body	7) Principle of Ethics Developers should respect human dignity and individual autonomy in the R&D of AI systems. • Encouraged that developers pay particularly due consideration to respecting human dignity and individual autonomy, in light of discussions on bioethics, etc. • Advisable that developers take precautions to ensure that AI systems do not unduly infringe the value of humanity. • Advisable that developers take precautions to ensure that AI systems do not unduly infringe the value of humanity.	4.1(1) The Human-Centric Principle The utilization of AI must not infringe upon the fundamental human rights guaranteed by the Constitution and international standards. AI should be developed, utilized, and implemented in society to expand the abilities of people and allow diverse people to pursue their own well-being.	P1: Respect for human autonomy The fundamental rights upon which the EU is founded are directed towards ensuring respect for the freedom and autonomy of human beings. Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process..... The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems.	1.2. Human-centred values and fairness Governments should call on AI actors to develop effective mechanisms to demonstrate that, throughout their lifecycle, AI systems respect human rights and democratic values, including freedom, dignity, autonomy, privacy, non-discrimination, fairness and social justice, and diversity [as well as core labour rights].	GP1. Human Rights: A/IS shall be created and operated to respect, promote, and protect internationally recognized human rights. GP2. Well-being A/IS creators shall adopt increased human well-being as a primary success criterion for development.	10) Value Alignment: Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation. 11) Human Values: AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity.	3) We are committed to open research and dialogue on the ethical, social, economic, and legal implications of AI. 6d) Maximize the benefits and address the potential challenges of AI technologies, by: Ensuring that <u>AI research and technology is robust, reliable, trustworthy, and operates within secure constraints.</u>
Diversity and Inclusiveness	Basic Philosophies To respect users' diversity, as well as an inclusiveness of people with diverse backgrounds, values, and ideas, throughout the process of the utilization of AI.	Basic Philosophies 1. To achieve a human-centered society where all human beings across the board enjoy the benefits from their life in harmony with AI networks, while human dignity and individual autonomy are respected.	2(2) Diversity & Inclusion: A society where people with diverse backgrounds can pursue their own well-being It is both an ideal in the modern world and a major challenge to create a society in which people with diverse backgrounds, values and ways of thinking can pursue their own well-being while <u>society creates new value by flexibly embracing them.</u> 4.1(1) The Human-Centric Principle In the process of AI deployment, each stakeholder should take into consideration the user-friendliness of the system in order to allow all people to enjoy the benefits of AI and avoid creating a digital divide with so-called "information poor" or "technology poor" people left behind. 4.1. (2) The Principle of Education/Literacy In a society premised on AI, we do not desire to create disparities or divisions between people or create those who are socially disadvantaged.	R5: Diversity, non-discrimination and fairness In order to achieve Trustworthy AI, we must enable inclusion and diversity throughout the entire AI system's life cycle. Besides the consideration and involvement of all affected stakeholders throughout the process, this also entails ensuring equal access through inclusive design processes as well as equal treatment. This requirement is closely linked with the principle of fairness.	1.1. Inclusive and sustainable growth and well-being Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as empowering human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and protecting natural environments, thus invigorating inclusive growth, sustainable development and well-being.		14) Shared Benefit: AI technologies should benefit and empower as many people as possible. 15) Shared Prosperity: The economic prosperity created by AI should be shared broadly, to benefit all of humanity. 23) Common Good: Superintelligence should only be developed in the service of widely shared ethical ideals, and for the benefit of all humanity rather than one state or organization.	
Sustainable society	Basic Philosophies To achieve a sustainable society by solving various problems with which individuals, local communities, countries, and the international community face through the utilization of AI, as the AI networking progresses.	Purpose : In the process of the evolution of AI networking, enormous benefits are expected for humans as well as the society and the economy in such manners as making significant contributions to solving various problems that individuals, local communities, countries, and the international community are confronted with.	2(3) Sustainability: A sustainable society We need to use AI to create a succession of new businesses and solutions, resolve social disparities, and develop a sustainable society that can deal with issues such as global environmental problems and climate change. Japan, as a leading science and technology-oriented country, has an obligation to strengthen its accumulated scientific and technological resources by utilizing AI and thereby contributing to the creation of such a sustainable society.	R6. Societal and environmental well-being In line with the principles of fairness and prevention of harm, the broader society, other sentient beings and the environment should be also considered as stakeholders throughout the AI system's life cycle. Sustainability and ecological responsibility of AI systems should be encouraged, and research should be fostered into AI solutions addressing areas of global concern, such as for instance the Sustainable Development Goals. Ideally, AI systems should be used to benefit all human beings, including future generations	1.1. Inclusive and sustainable growth and well-being Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as empowering human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and protecting natural environments, thus invigorating inclusive growth, sustainable development and well-being.		20) Importance: Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources.	

Comparison of AI guidelines

Values to be respected	AI Utilization Guidelines	Draft AI R&D guidelines for international discussions	Social Principles of Human-centric AI	Ethics Guideline for Trustworthy AI	Recommendation of the Council on Artificial Intelligence	Ethically Aligned Design	Asilomar AI Principles	Tenets
International Cooperation	<p>Basic Philosophies</p> <p>To share the Guidelines, as non-binding soft law, and their best practices internationally among stakeholders.</p> <p>To constantly review the Guidelines and flexibly revise them as necessary through international discussions, considering the extent of the progress of AI networking.</p>	<p>Basic Philosophies:</p> <p>2. To share the Guidelines, as non-binding soft law, and their best practices internationally among stakeholders, as, with the rapid development of the R&D and utilization of AI, networked AI systems are expected to have broad and significant impacts on human beings and society beyond national borders.</p>	<p>4.2 AI R&D and Utilization Principles</p> <p>Since many countries, organizations, companies, and so on are currently discussing AI development and utilization principles, we believe it is important to create an international consensus quickly through open discussions, and share results internationally in a non-regulatory, non-binding framework.</p> <p>5. Conclusion</p> <p>In addition, Japan should share these principles with other countries around the world and take a leadership role in international discussions with the goal of establishing an AI-Ready Society worldwide. In doing so, Japan should present to the world a social image of Society 5.0 that supports the realization of the SDGs, and should contribute to cooperative and creative new development of the international community.</p>	<p>Introduction:</p> <p>Just as the use of AI systems does not stop at national borders, neither does their impact. Global solutions are therefore required for the global opportunities and challenges that AI systems bring forth. We therefore encourage all stakeholders to work towards a global framework for Trustworthy AI, building international consensus while promoting and upholding our fundamental rights-based approach.</p>	<p>2.5 International cooperation for trustworthy AI</p> <p>a) Governments, including developing countries and with stakeholders, should actively cooperate to advance these principles and to progress on responsible stewardship of trustworthy AI.</p> <p>...</p>			
Proper Utilization	<p>1) Principle of Proper Utilization</p> <p>Users should make efforts to utilize AI systems or AI services in a proper scope and manner, under the proper assignment of roles between humans and AI systems, or among users.</p> <ul style="list-style-type: none"> Utilization in the proper scope and manner on the basis of the provision of information and explanation from developers, etc. Cooperation among stakeholders to work on preventive or remedial measures including an elucidation of causes and measures to prevent recurrence. 	<p>8) Principle of User assistance</p> <p>Developers should take it into consideration that AI systems will support users and make it possible to give them opportunities for choice in appropriate manners.</p> <ul style="list-style-type: none"> To make efforts to make available interfaces that provide in a timely and appropriate manner the information that can help users' decisions and are easy-to-use for them. To make efforts to give consideration to make available functions that provide users with opportunities for choice in a timely and appropriate manner. To make efforts to take measures to make AI systems easier to use for socially-vulnerable people such as universal design. 				<p>GP4. Effectiveness</p> <p>A/IS creators and operators shall provide evidence of the effectiveness and fitness for purpose of A/IS.</p>	<p>1) We will seek to ensure that <u>AI technologies benefit and empower as many people as possible.</u></p> <p>7) We believe that it is important for the operation of AI systems <u>to be understandable and interpretable by people, for purposes of explaining the technology.</u></p>	
Education/literacy	<p>1) Principle of Proper Utilization – Utilization in the proper scope and manner</p> <p>Users are expected to recognize benefits and risks, understand proper uses, acquire necessary knowledge and skills and so on before using AI, according to the characteristics, usage situations, etc. of AI.</p>		<p>4.1. (1) The Human-Centric Principle</p> <p>It is desirable that we introduce appropriate mechanisms for literacy education and for the promotion of proper use of AI</p> <p>4.1.(2) The Principle of Education/Literacy</p> <p>From this point of view, we believe that an educational environment that fosters education and literacy in accordance with the principles must be provided equally to all people.</p>	<p>2.2. Non-technical methods</p> <p>--> Education and awareness to foster an ethical mind-set</p>	<p>2.4. Building human capacity and preparing for labour market transformation</p> <p>a) Governments should work closely with stakeholders to prepare for the transformation of the world of work and of society. They should empower people to effectively use and interact with AI systems across the breadth of applications, including by equipping them with the necessary skills.</p>	<p>GP8. Competence</p> <p>A/IS creators shall specify and operators shall adhere to the knowledge and skill required for safe and effective operation.</p>		
Human intervention and Controllability	<p>1) Principle of Proper Utilization – Human Intervention</p> <p>Regarding the judgment made by AI, in cases where it is necessary and possible, humans may be expected to make decisions as to whether to use the judgments of AI, how to use it etc.</p>	<p>3) Principle of controllability</p> <p>Developers should pay attention to the controllability of AI systems.</p> <ul style="list-style-type: none"> Encouraged that developers make efforts to conduct verification and validation in advance (to conduct experiments in a closed space such as in a laboratory or a sandbox in which security is ensured, at a stage before the practical application in society.) Encouraged that developers pay attention to whether the supervision (such as monitoring or warnings) and countermeasures (such as system shutdown, cut-off from networks, or repairs) by humans or other trustworthy AI systems are effective. 	<p>4.1. (1) The Human-Centric Principle</p> <p>When using AI, people must judge and decide for themselves how to use it.</p>	<p>R1. Human agency and oversight</p> <p>AI systems should support human autonomy and decision-making, as prescribed by the principle of respect for human autonomy. This requires that AI systems should both act as enablers to a democratic, flourishing and equitable society by supporting the user's agency and foster fundamental rights, and allow for human oversight.</p>	<p>1.2. Human-centred values and fairness</p> <p>b) AI actors should implement mechanisms and safeguards, such as capacity for human determination, that are appropriate to the context and consistent with the state of art.</p>		<p>16) Human Control:</p> <p>Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.</p>	
Proper data	<p>2) Principle of Data Quality</p> <p>Users and data providers should pay attention to the quality of data used for learning or other methods of AI systems.</p> <ul style="list-style-type: none"> Attention to the quality of data (e.g., the accuracy and completeness of data) used for learning or other methods of AI. Attention to security vulnerabilities of AI by learning inaccurate or inappropriate data 			<p>R3. Privacy and Data Governance</p> <p>Closely linked to the principle of prevention of harm is privacy, a fundamental right particularly affected by AI systems. Prevention of harm to privacy also necessitates adequate data governance that covers the quality and integrity of the data used, its relevance in light of the domain in which the AI systems will be deployed, its access protocols and the capability to process data in a manner that protects privacy.</p>				

Comparison of AI guidelines

Values to be respected	AI Utilization Guidelines	Draft AI R&D guidelines for international discussions	Social Principles of Human-centric AI	Ethics Guideline for Trustworthy AI	Recommendation of the Council on Artificial Intelligence	Ethically Aligned Design	Asilomar AI Principles	Tenets
Collaboration among AI systems	3) Principle of Collaboration AI service providers, business users, and data providers should pay attention to the collaboration of AI systems or AI services. Users should take it into consideration that risks might occur and even be amplified when AI systems are to be networked. <ul style="list-style-type: none"> Attention to the interconnectivity and interoperability of AI systems Address to the standardization of data formats, protocols, etc. Attention to problems caused and amplified by AI networking 	1) Principle of collaboration Developers should pay attention to the interconnectivity and interoperability of AI systems. <ul style="list-style-type: none"> To make efforts to address the standardization of data formats and the openness of interfaces and protocols including API. To pay attention to risks of unintended events as a result of the interconnection or interoperations between AI systems that they have developed and other AI systems, etc. To make efforts to promote open and fair treatment of license agreements for and their conditions of intellectual property rights 			2.5 International cooperation for trustworthy AI c) Governments should promote the development of multi-stakeholder, consensus-driven global technical standards for interoperable and trustworthy AI.			5) We will engage with and have representation from stakeholders in the business community <u>to help ensure that domain-specific concerns and opportunities are understood and addressed.</u>
Safety	4) Principle of Safety Users should take into consideration that AI systems or AI services in use will not harm the life, body, or property of users or third parties through the actuators or other devices. <ul style="list-style-type: none"> Take into consideration that AI will not harm the life, body, or property through the actuators or other devices, by inspecting and repairing AI, updating AI software, etc. as necessary. Consider in advance measures to be taken, in case AI might harm the life, body, or property 	4) Principle of Safety Developers should take it into consideration that AI systems will not harm the life, body, or property of users or third parties through actuators or other devices. <ul style="list-style-type: none"> to conduct verification and validation in advance to contribute to the intrinsic safety (reduction of essential risk factors such as kinetic energy of actuators) and the functional safety (mitigation of risks by operation of additional control devices such as automatic braking) to explain the designers' intent of AI systems and the reasons for it to stakeholders 	4.1(4) The Principle of Ensuring Security The active use of AI automates many social systems and greatly improves safety. On the other hand, at least within the scope of currently available technologies, it is not always possible for AI to respond appropriately to rare events or deliberate attacks. Therefore, the use of AI poses a new set of risks to security. Society should always be aware of the balance between the benefits and risks, and endeavor to improve social safety and sustainability as a whole.	R2. Technical robustness and safety A crucial component of achieving Trustworthy AI is technical robustness, which is closely linked to the principle of prevention of harm. Technical robustness requires that AI systems be developed with a preventative approach to risks and in a manner such that they reliably behave as intended while minimising unintentional and unexpected harm, and preventing unacceptable harm. This should also apply to potential changes in their operating environment or the presence of other agents (human and artificial) that may interact with the system in an adversarial manner. In addition, the physical and mental integrity of humans should be ensured.	1.4. Robustness, security and safety a) <u>AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk.</u> c) <u>AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias.</u>	GP7. Awareness of Misuse A/IS creators shall guard against all potential misuses and risks of A/IS in operation.	5) Race Avoidance: Teams developing AI systems should actively cooperate to avoid corner-cutting on safety standards. 6) Safety: AI systems should be safe and secure throughout their operational lifetime, and verifiably so where applicable and feasible. 17) Non-subversion: <u>The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends.</u> 22) Recursive Self-Improvement: <u>AI systems should be designed to improve themselves, and to improve the performance of other AI systems, in a way that respects human values and the well-being of society.</u>	6e) Maximize the benefits and address the potential challenges of AI technologies, by: <u>Opposing development and use of AI technologies that would violate international conventions or human rights, and promoting safeguards and technologies that do no harm.</u>
Security	5) Principle of Security Users and data providers should pay attention to the security of AI systems or AI services. <ul style="list-style-type: none"> Take reasonable measures in light of the technology level at that time. Consider measures to be taken against security breaches of AI in advance. Provide services for security measures to end users and share incident information with end users. Attention to security vulnerabilities of AI learning model 	5) Principle of Security Developers should pay attention to the security of AI systems. <ul style="list-style-type: none"> To pay attention, as necessary, to the reliability and robustness of AI systems, in addition to: confidentiality; integrity; and availability of information that are usually required for ensuring the information security of AI systems. To make efforts to conduct verification and validation in advance. "security by design" 	4.1(4) The Principle of Ensuring Security The active use of AI automates many social systems and greatly improves safety. On the other hand, at least within the scope of currently available technologies, it is not always possible for AI to respond appropriately to rare events or deliberate attacks. Therefore, <u>the use of AI poses a new set of risks to security.</u> Society should always be aware of the balance between the benefits and risks, and endeavor to improve social safety and sustainability as a whole.	R.2 Technical robustness and safety A crucial component of achieving Trustworthy AI is technical robustness, which is closely linked to the principle of prevention of harm. Technical robustness requires that AI systems be developed with a preventative approach to risks and in a manner such that they reliably behave as intended while minimising unintentional and unexpected harm, and preventing unacceptable harm. This should also apply to potential changes in their operating environment or the presence of other agents (human and artificial) that may interact with the system in an adversarial manner. In addition, the physical and mental integrity of humans should be ensured.	1.4. Robustness, security and safety + Reference -Digital Security Risk Management for Economic and Social Prosperity(revised in 2015) http://www.oecd.org/sti/ieconomy/digital-security-risk-management.pdf	GP7. Awareness of Misuse A/IS creators shall guard against all potential misuses and risks of A/IS in operation.		6a) Maximize the benefits and address the potential challenges of AI technologies, by: <u>Working to protect the privacy and security of individuals.</u> 6d) Maximize the benefits and address the potential challenges of AI technologies, by: <u>Ensuring that AI research and technology is robust, reliable, trustworthy, and operates within secure constraints.</u>
Privacy	6) Principle of Privacy Users and data providers should take into consideration that the utilization of AI systems or AI services will not infringe on the privacy of users' or others. <ul style="list-style-type: none"> Respect for the privacy of end users and third parties in the utilization of AI Respect for the privacy of others in the collection, pre-process, and provision of personal data used for learning or other methods of AI. Attention to the infringement of the privacy of users' or others Prevention of personal data leakage 	6) Principle of Privacy Developers should take it into consideration that AI systems will not infringe the privacy of users or third parties. <ul style="list-style-type: none"> To make efforts to evaluate the risks of privacy infringement and conduct privacy impact assessment in advance. "privacy-by-design" 	4.1(3) The Principle of Privacy Protection in a society premised on AI, it is still possible to gauge each person's political position, economic situation, personal hobbies, personal preferences, and so forth with great accuracy based on data about matters such as data subject's individual behavior. <u>This means, when utilizing AI, that more careful discretion may be required than the mere handling of personal data in accordance with the level of importance and sensitivity of the data. Each stakeholder must handle personal data based on the following principles to ensure that no individuals are disadvantaged from the unexpected distribution or use of personal data in undesirable ways.</u>	R3. Privacy and Data Governance Closely linked to the principle of prevention of harm is privacy, a fundamental right particularly affected by AI systems. Prevention of harm to privacy also necessitates adequate data governance that covers the quality and integrity of the data used, its relevance in light of the domain in which the AI systems will be deployed, its access protocols and the capability to process data in a manner that protects privacy.	Reference -Guidelines on the Protection of Privacy and Transborder Flows of Personal Data(revised in 2013) http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf	GP3. Data Agency A/IS creators shall empower individuals with the ability to access and securely share their data, to maintain people's capacity to have control over their identity.	12) Personal Privacy: People should have the right to access, manage and control the data they generate, given AI systems' power to analyze and utilize that data. 13) Liberty and Privacy: The application of AI to personal data must not unreasonably curtail people's real or perceived liberty.	6a) Maximize the benefits and address the potential challenges of AI technologies, by: <u>Working to protect the privacy and security of individuals.</u>
Fairness, equity, removal of discrimination	8) Principle of Fairness AI service providers, business users, and data providers should pay attention to the possibility of bias inherent in the judgement of AI systems or AI services, and take into consideration that individuals and groups will not be discriminated unfairly by those judgment. <ul style="list-style-type: none"> Attention to the representativeness of data used for learning or other methods of AI and the social bias inherent in the data Attention to unfair discrimination by algorithm Human intervention in the judgment made by AI (from the viewpoint of fairness) 	7) Principle of Ethics Developers should respect human dignity and individual autonomy in the R&D of AI systems. <ul style="list-style-type: none"> Encouraged that developers pay particularly due consideration to respecting human dignity and individual autonomy, in light of discussions on bioethics, etc. Advisable that developers take precautions to ensure that AI systems do not unduly infringe the value of humanity. Advisable that developers take precautions to ensure that AI systems do not unduly infringe the value of humanity. 	4.1(6) The Principle of Fairness, Accountability, and Transparency Under AI's design concept, <u>all people are treated fairly without unjustified discrimination on the grounds of diverse backgrounds</u> such as race, gender, nationality, age, political beliefs, religion, and so on.	1.5 Diversity, non-discrimination and fairness In order to achieve Trustworthy AI, we must enable inclusion and diversity throughout the entire AI system's life cycle. Besides the consideration and involvement of all affected stakeholders throughout the process, this also entails ensuring equal access through inclusive design processes as well as equal treatment. This requirement is closely linked with the principle of fairness.	1.2. Human-centred values and fairness a) AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognised labour rights.			

Comparison of AI guidelines

Values to be respected	AI Utilization Guidelines	Draft AI R&D guidelines for international discussions	Social Principles of Human-centric AI	Ethics Guideline for Trustworthy AI	Recommendation of the Council on Artificial Intelligence	Ethically Aligned Design	Asilomar AI Principles	Tenets
Transparency, Explainability	<p>9) Principle of Transparency AI service providers and business users should pay attention to the verifiability of inputs/outputs of AI systems or AI services and the explainability of their judgments.</p> <ul style="list-style-type: none"> Recording and preserving the inputs/outputs of AI, in the case of using AI in fields where AI systems might harm the life, body, or property Ensuring explainability, in the case of using AI in fields where the judgments of AI might have significant influences on individual rights and interests Ensuring transparency when AI is used in administrative agencies (※Not intended to ask for the disclosure of algorithm, source code, or learning data.) 	<p>2) Principle of transparency Developers should pay attention to the verifiability of inputs/outputs of AI systems and the explainability of their judgments.</p> <ul style="list-style-type: none"> desirable that developers pay attention to the verifiability of the inputs and outputs of AI systems as well as the explainability of the judgment of AI systems that might affect the life, body, freedom, privacy, or property of users or third parties. (※Not intended to ask for the disclosure of algorithm, source code, or learning data.) 	<p>4.1(6) The Principle of Fairness, Accountability, and Transparency · <u>Appropriate explanations should be given on a case-by-case basis depending on the application of AI and each particular situation, including such things as when AI is being used, how the AI data is obtained and used, and what measures have been taken to ensure the appropriateness of results obtained from AI operations.</u></p> <ul style="list-style-type: none"> In order for people to understand AI's proposals and make judgments on them, there should be appropriate opportunities for an open dialogue, as required, regarding the use, adoption, and operation of AI. 	<p>1.4 Transparency This requirement is closely linked with the principle of explicability and encompasses transparency of elements relevant to an AI system: the data, the system and the business models. (Traceability, Explainability)</p>	<p>1.3. Transparency and explainability <u>AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art.</u></p> <p>1.4. Robustness, security and safety <u>b) AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outcomes and responses to inquiry, appropriate to the context and consistent with the state of art.</u></p>	<p>GP5. Transparency The basis of a particular A/IS decision should always be discoverable.</p>	<p>4) Research Culture: A culture of cooperation, trust, and transparency should be fostered among researchers and developers of AI.</p> <p>7) Failure Transparency: If an AI system causes harm, it should be possible to ascertain why.</p> <p>8) Judicial Transparency: Any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority.</p>	<p>7) We believe that it is important for the operation of AI systems to be <u>understandable and interpretable by people, for purposes of explaining the technology.</u></p>
Accountability	<p>10) Principle of Accountability Users should make efforts to fulfill their accountability to the stakeholders.</p> <ul style="list-style-type: none"> Efforts to fulfill accountability Notification and publication of usage policy on AI systems or AI services 	<p>9) Principle of Accountability Developers should make efforts to fulfill their accountability to stakeholders including AI systems' users.</p> <ul style="list-style-type: none"> Encouraged that developers make efforts to provide users with the information that can help their choice and utilization of AI systems. Encouraged that, taking into account the R&D principles (1) to (8) set forth in the Guidelines, developers make efforts: (a) to provide users et al. with both information and explanations about the technical characteristics of the AI systems they have developed; and (b) to gain active involvement of stakeholders in such manners as to hear various views through dialogues with diverse stakeholders. it is advisable that developers make efforts to share the information and cooperate with providers et al. 	<p>4.1(6) The Principle of Fairness, Accountability, and Transparency · <u>Appropriate explanations should be given on a case-by-case basis depending on the application of AI and each particular situation, including such things as when AI is being used, how the AI data is obtained and used, and what measures have been taken to ensure the appropriateness of results obtained from AI operations.</u></p> <ul style="list-style-type: none"> In an "AI-Ready Society", it is necessary to ensure ... appropriate accountability for the results, and trust in the technology, so that people who use AI are not subject to undue discrimination with regard to personal background, or to unfair treatment in terms of human dignity. 	<p>R7. Accountability The requirement of accountability complements the above requirements, and is closely linked to the principle of fairness. It necessitates that mechanisms be put in place to ensure responsibility and accountability for AI systems and their outcomes, both before and after their development, deployment and use.</p>	<p>1.5. Accountability AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art.</p>	<p>GP6. Accountability A/IS shall be created and operated to provide an unambiguous rationale for all decisions made.</p>	<p>3) Science-Policy Link: There should be constructive and healthy exchange between AI researchers and policy-makers.</p> <p>4) Research Culture: A culture of cooperation, trust, and transparency should be fostered among researchers and developers of AI.</p> <p>9) Responsibility: Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications.</p>	<p>2) We will educate and listen to the public and actively engage stakeholders <u>to seek their feedback on our focus, inform them of our work, and address their questions.</u></p> <p>3) We are <u>committed to open research and dialogue on the ethical, social, economic, and legal implications of AI.</u></p> <p>4) We believe that AI research and development efforts <u>need to be actively engaged with and accountable to a broad range of stakeholders.</u></p> <p>5) We will engage with and have representation from stakeholders in the business community <u>to help ensure that domain-specific concerns and opportunities are understood and addressed.</u></p>