AIガイドライン比較表

［別添4］
（報告書（案）別紙2）

尊重すべき価値	**AI利活用ガイドライン／AI利活用原則案 AI Utilization guidelines/Draft AI Utilization Principles**	**国際的な議論のためのAI開発ガイドライン案 Draft AI R&D guidelines for international discussions**	**「人間中心のＡＩ社会原則」 Social Principles of Human-centric AI**	**人工知能学会倫理指針 Ethical Guideline**	**Ethics Guideline for Trustworthy AI**	**Recommendation of the Council on Artificial Intelligence**	**Ethically Aligned Design**	**Asilomar AI Principles**	**Tenets**
by	AIネットワーク社会推進会議（総務省）/Japan	AIネットワーク社会推進会議（総務省）/Japan	統合イノベーション戦略推進会議（人間中心のAI社会原則会議）/Japan	人工知能学会(JSAI)/Japan	European Commission (High Level Expert Group on AI(HLEG))	OECD	IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems	Future of Life Institute (FLI)	Partnership on AI
公開日		2017/7/28	2019/3/29	2017/2/28	2019/4/8	2019/5/22	2019/3/25(1st edition)	2017/2/	2016/9/28
過去（案など）の公開日	2018/7/17 利活用原則案公開		2018/12/27 案公開		2018/12/18 案公開		2016/12/13(ver.1), 2017/12/12(ver.2)		
URL	http://www.soumu.go.jp/main_content/000564147.pdf (eng)http://www.soumu.go.jp/main_content/000581310.pdf	http://www.soumu.go.jp/main_content/000499625.pdf (eng)http://www.soumu.go.jp/main_content/000507517.pdf	https://www.cas.go.jp/jp/seisaku/jinkouchinou/	http://ai-elsi.org/wp-content/uploads/2017/02/%E4%BA%BA%E5%B7%A5%E7%9F%A5%E8%83%BD%E5%AD%A6%E4%BC%9A%E5%80%AB%E7%90%86%E6%8C%87%E9%87%9D.pdf (eng) http://ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1.pdf	https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai	https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449	https://ethicsinaction.ieee.org/	https://futureoflife.org/ai-principles/	https://www.partnershiponai.org/tenets/
主要構成	目的、基本理念(6) AI利活用原則（10）＋解説 等	基本理念(5) AI開発原則（9）＋解説 等	2. 基本理念(3) 3. ビジョン(5) 4.1. AI社会原則(7) 4.2. AI開発利用原則 等	序文＋指針（9）	1.Foundations of Trustworthy AI(4 Principles) 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 人間中心	**基本理念：** 人間がAIネットワークと共生することにより、その恵沢がすべての人によってあまねく享受され、人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること	**基本理念** 1. 人間がＡＩネットワークと共生することにより、その恵沢がすべての人によってあまねく享受され、人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること。	**2(1).人間の尊厳が尊重される社会（Dignity）** 人間がAIを道具として使いこなすことによって、人間の様々な能力をさらに発揮することを可能とし、より大きな創造性を発揮したり、やりがいのある仕事に従事したりすることで、物質的にも精神的にも豊かな生活を送ることができるような、人間の尊厳が尊重される社会を構築する必要がある。		**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 人間の尊厳	**7) 尊厳･自律の原則 (human dignity and individual autonomy)** 利用者は、AIシステム又はAIサービスの利活用において、人間の尊厳と個人の自律を尊重する。 ・ AIシステム又はAIサービスにより意思決定や感情が操作されるリスク、AIシステム又はAIサービスに過度に依存するリスク ・ （AIシステム又はAIサービスを人間の脳や身体と連携させる場合）生命倫理に関する議論などの参照 ・AIを利用したプロファイリングを行う場合における不利益への配慮	**7) 倫理の原則 (Ethics)** 開発者は、AIシステムの開発において、人間の尊厳と個人の自律を尊重する。 ・ （人間の脳や身体と連携するAIシステムの開発の場合）生命倫理に関する議論などの参照 ・ 学習データに含まれる偏見などに起因して不当な差別が生じないための所要の措置 ・ AIシステムが人間性の価値を不当に毀損することがないよう留意	**4.1(1) 人間中心の原則** AIの利用は、憲法及び国際的な規範の保障する基本的人権を侵すものであってはならない。 AIは、人々の能力を拡張し、多様な人々の多様な幸せの追求を可能とするために開発され、社会に展開され、活用されるべきである。	**1) 人類への貢献(Contribution to humanity)** 人類の平和、安全、福祉、公共の利益に貢献し、基本的人権と尊厳を守り、文化の多様性を尊重する。人工知能を設計、開発、運用する際には専門家として人類の安全への脅威を排除するように努める。 **9) 人工知能への倫理遵守の要請(Abidance of ethics guidelines by AI)** 人工知能が社会の構成員またはそれに準じるものとなるためには、(1-8)に定めた人工知能学会員と同等に倫理指針を遵守できなければならない。	**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 AI research and technology is robust, reliable, trustworthy, and operates within secure constraints.
Diversity, Inclusiveness, 多様性、包摂	**基本理念：** AIの利活用において利用者の多様性を尊重し、多様な背景と価値観、考え方を持つ人々を包摂すること	**基本理念** 1. 人間がＡＩネットワークと共生することにより、その恵沢がすべての人によってあまねく享受され、人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること。	**2(2) 多様な背景を持つ人々が多様な幸せを追求できる社会（Diversity & Inclusion）** 多様な背景と価値観、考え方を持つ人々が多様な幸せを追求し、それらを柔軟に包摂した上で新たな価値を創造できる社会は、現代における一つの理想であり、大きなチャレンジである。 **4.1(1) 人間中心の原則** AIの普及の過程で、いわゆる「情報弱者」や「技術弱者」を生じさせず、AIの恩恵をすべての人が享受できるよう、使いやすいシステムの実現に配慮すべきである。 **4.1.（2） 教育･リテラシーの原則** AIを前提とした社会において、我々は、人々の間に格差や分断が生じたり、弱者が生まれたりすることは望まない。	**4) 公正性(Fairness)** 人工知能の開発と利用において常に公正さを持ち、人工知能が人間社会において不公平や格差をもたらす可能性があることを認識し、開発にあたって差別を行わないよう留意する。人類が公平、平等に人工知能を利用できるように努める。	**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 持続可能な社会	**基本理念：** AIネットワーク化の進展とともに、AIの利活用により個人、地域社会、各国、国際社会が抱える様々な課題の解決を図り、持続可能な社会を実現すること	**目的：** AIネットワークが進展していく過程で、個人、地域社会、各国、国際社会の抱える様々な課題の解決に大きく貢献するなど、人間及びその社会や経済に多大な便益を広範にもたらすことが期待される。	**2(3) 持続性ある社会（Sustainability）** 我々は、AIの活用によりビジネスやソリューションを次々と生み、社会の格差を解消し、地球規模の環境問題や気候変動などにも対応が可能な持続性のある社会を構築する方向へ展開させる必要がある。科学・技術立国としての我が国は、その科学的・技術的蓄積をAIによって強化し、そのような社会を作ることに貢献する責務がある。	**8) 社会との対話と自己研鑽(Communication with society and self-development)** 人工知能に関する社会的な理解が深まるよう努める。社会には様々な声があることを理解し、社会から真摯に学び、理解を深め、社会との不断の対話を通じて専門家として人間社会の平和と幸福に貢献することとする。高度な専門家として絶え間ない自己研鑽に努め自己の能力の向上を行うと同時にそれを望む者を支援することとする。	**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.	

AIガイドライン比較表

［別添4］
（報告書（案）別紙2）

尊重すべき価値	**AI利活用ガイドライン／AI利活用原則案 AI Utilization guidelines/Draft AI Utilization Principles**	**国際的な議論のためのAI開発ガイドライン案 Draft AI R&D guidelines for international discussions**	**「人間中心のＡＩ社会原則」 Social Principles of Human-centric AI**	**人工知能学会倫理指針 Ethical Guideline**	**Ethics Guideline for Trustworthy AI**	**Recommendation of the Council on Artificial Intelligence**	**Ethically Aligned Design**	**Asilomar AI Principles**	**Tenets**
International Cooperation 国際協力	**基本理念** • AIの利活用の在り方について、非拘束的なソフトローたる指針やベストプラクティスを国際的に共有すること ・AIネットワーク化の進展等を踏まえ、国際的な議論を通じて、本ガイドライン案を不断に見直し、必要に応じて柔軟に改定すること	**基本理念** ２．ＡＩの研究開発と利活用が今後急速に発展し、ネットワーク化されたＡＩシステムが国境を越えて人間及び社会に広範かつ多大な影響を及ぼすものと見込まれることから、ＡＩシステムの研究開発の在り方について、非拘束的なソフトローたる指針やそのベストプラクティスをステークホルダ間で国際的に共有すること。	**4.2. AI開発利用原則** AI開発利用原則については、現在、多くの国、団体、企業等において議論されていることから、我々は早急にオープンな議論を通じて国際的なコンセンサスを醸成し、非規制的で非拘束的な枠組みとして国際的に共有されることが重要であると考える。 **5. おわりに** 国際的な議論の場において、我が国は、本原則を世界各国と共有した上で、国際的な議論のリーダーシップをとり、コンセンサスの形成を目指すべきであり、それによってSDGsの実現を支えるSociety5.0の社会像を世界に示し、国際社会の協調的かつ創造的な新たな発展に寄与すべきである。		**Introduction:** 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.	**2.5 International cooperation for trustworthy AI** a) Governments, including developing countries and with stakeholders, should actively cooperate to advance these principles and to progress on responsible stewardship of trustworthy AI. …			
Proper utilization 適正な利用	**1) 適正利用の原則 (Proper utilization)** 利用者は、人間とAIシステムとの間及び利用者間における適切な役割分担のもと、適正な範囲及び方法でAIシステム又はAIサービスを利用するよう努める。 ・ 開発者等からの情報提供や説明を踏まえた適正な範囲・方法での利用 ・ 予防措置、事後対応（原因解明、再発防止措置等）における関係者間の協力	**8) 利用者支援の原則 (User assistance)** 開発者は、AIシステムが利用者を支援し、利用者に選択の機会を適切に提供することが可能となるよう配慮する。 ・ 利用者にとって操作しやすいインターフェース ・ 利用者に選択の機会（デフォルトの設定、理解しやすい選択肢の提示等）を適時適切に提供する機能 ・ 社会的弱者の利用を容易にするための取組 ・ 利用者に対する適切な情報提供					**GP4. Effectiveness** A/IS creators and operators shall provide evidence of the effectiveness and fitness for purpose of A/IS.		1) We will seek to ensure that AI technologies benefit and empower as many people as possible. 7) We believe that it is important for the operation of AI systems to be understandable and interpretable by people, for purposes of explaining the technology.
Education/literacy 教育・リテラシー	**1) 適正利用の原則－ア 適正な範囲・方法での利用** 利用者は、ＡＩの性質、利用の態様等に応じて、利用する前に、便益及びリスクを認識し、適正な用途を理解するとともに、必要な知識・技能を習得すること等が期待されるのではないか。		**4.1.（1） 人間中心の原則** 我々は、リテラシー教育や適正な利用の促進などのための適切な仕組みを導入することが望ましい。 **4.1.(2) 教育・リテラシーの原則** 我々は、以下のような原則に沿う教育・リテラシーを育む教育環境が全ての人に平等に提供されなければならないと考える。		2.2. Non-technical methods --> Education and awareness to foster an ethical mind-set	**2.4. Building human capacity and preparing for labour market transformation** 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.	**GP8. Competence** A/IS creators shall specify and operators shall adhere to the knowledge and skill required for safe and effective operation.		
Human intervention 人間の判断の介在 Controllability 制御可能性	**1) 適正利用の原則－イ 人間の判断の介在** AIによりなされた判断について、必要かつ可能な場合には、その判断を用いるか否か、あるいは、どのように用いるか等に関し、人間の判断を介在させることが期待される。	**3) 制御可能性の原則 (Controllability)** 開発者は、AIシステムの制御可能性に留意する。 ・ 事前の検証及び妥当性の確認、サンドボックスにおける実験 ・ 人間や信頼できる他のAIシステムによる監督・対処	**4.1.（1） 人間中心の原則** AIの利用にあたっては、人が自らどのように利用するかの判断と決定を行うことが求められる。	**5) 安全性(Security)** 専門家として、人工知能の安全性及びその制御において責任を認識し、人工知能の開発と利用において常に安全性と制御可能性、必要とされる機密性について留意し、同時に人工知能を利用する者に対し適切な情報提供と注意喚起を行うように努める。	**R1. Human agency and oversight** 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.	**1.2. Human-centred values and fairness** 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.		**16) Human Control:** Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.	
Proper data 適正な学習（学習データの質）	**2) 適正学習の原則 (Data Quality)** 利用者及びデータ提供者は、AIシステムの学習等に用いるデータの質に留意する。 ・ 学習等に用いるデータの質（正確性や完全性など）				**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.				
Collabolation among AI systems AI間の連携	**3) 連携の原則 (Collabolation)** AIサービスプロバイダ、ビジネス利用者及びデータ提供者は、AIシステム又はAIサービス相互間の連携に留意する。また、利用者は、AIシステムがネットワーク化することによってリスクが惹起・増幅される可能性があることに留意する。 ・ 提供するAIシステム又はAIサービスの相互接続性と相互運用性 ・ データ形式やプロトコル等の標準化への対応 ・ AIネットワーク化により惹起・増幅される課題	**1) 連携の原則 (Collaboration)** 開発者は、AIシステムの相互接続性と相互運用性に留意する。 ・ 国際的な標準や規格への準拠 ・ データ形式の標準化、インターフェイスやプロトコルのオープン化への対応 ・ 標準必須特許等のライセンス契約及びその条件についてのオープン・公平な取扱い				**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 to help ensure that domain-specific concerns and opportunities are understood and addressed.

AIガイドライン比較表

［別添4］
（報告書（案）別紙2）

尊重すべき価値	**AI利活用ガイドライン／AI利活用原則案 AI Utilization guidelines/Draft AI Utilization Principles**	**国際的な議論のためのAI開発ガイドライン案 Draft AI R&D guidelines for international discussions**	**「人間中心のＡＩ社会原則」 Social Principles of Human-centric AI**	**人工知能学会倫理指針 Ethical Guideline**	**Ethics Guideline for Trustworthy AI**	**Recommendation of the Council on Artificial Intelligence**	**Ethically Aligned Design**	**Asilomar AI Principles**	**Tenets**
Safety 安全性	**4) 安全の原則（Safety）** 利用者は、AIシステム又はAIサービスの利活用により、アクチュエータ等を通じて、利用者等及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する。 ・（生命・身体・財産に危害を及ぼし得る分野での利活用において）AIシステムの点検・修理及びAIソフトのアップデートを行うことなどによる危害の防止 ・危害が発生した場合に備えた事前措置	**4) 安全の原則（Safety）** 開発者は、AIシステムがアクチュエータ等を通じて利用者及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する。 ・事前の検証及び妥当性の確認 ・本質安全や機能安全に資するための措置 ・（生命・身体・財産の安全に関する判断を行うAIシステムについて）ステークホルダに対する設計の趣旨などの説明	**4.1(4) セキュリティ確保の原則** AIを積極的に利用することで多くの社会システムが自動化され、安全性が向上する。一方、少なくとも現在想定できる技術の範囲では、希少事象や意図的な攻撃に対してAIが常に適切に対応することは不可能であり、セキュリティに対する新たなリスクも生じる。社会は、常にベネフィットとリスクのバランスに留意し、全体として社会の安全性及び持続可能性が向上するように務めなければならない。	**1) 人類への貢献(Contribution to humanity)** 人類の平和、安全、福祉、公共の利益に貢献し、基本的人権と尊厳を守り、文化の多様性を尊重する。人工知能を設計、開発、運用する際には専門家として人類の安全への脅威を排除するように努める。 **2) 法規制の遵守(Abidance of laws and regulations)** 専門家として、研究開発に関わる法規制、知的財産、他者との契約や合意を尊重しなければならない。他者の情報や財産の侵害や損失といった危害を加えてはならず、直接的のみならず間接的にも他者に危害を加えるような意図をもって人工知能を利用しない。 **5) 安全性(Security)** 専門家として、人工知能の安全性及びその制御における責任を認識し、人工知能の開発と利用において常に安全性と制御可能性、必要とされる機密性について留意し、同時に人工知能を利用する者に対し適切な情報提供と注意喚起を行うように努める。	**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) 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. c) 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.	**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:** 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. **22) Recursive Self-Improvement:** AI systems designed to recursively self-improve or self-replicate in a manner that could lead to rapidly increasing quality or quantity must be subject to strict safety and control measures. 18) AI Arms Race: An arms race in lethal autonomous weapons should be avoided.	6e) Maximize the benefits and address the potential challenges of AI technologies, by: Opposing development and use of AI technologies that would violate international conventions or human rights, and promoting safeguards and technologies that do no harm.
Security セキュリティ	**5) セキュリティの原則（Security）** 利用者及びデータ提供者は、AIシステム又はAIサービスのセキュリティに留意する。 ・その時点での技術水準に照らした合理的な対策 ・侵害が発生した場合に備えた事前措置 ・セキュリティ対策のためのサービス提供、インシデント情報の共有 ・AIの学習モデルに対するセキュリティ脆弱性への留意	**5) セキュリティの原則（Security）** 開発者は、AIシステムのセキュリティに留意する。 ・情報の機密性、完全性、可用性に加え、信頼性、頑健性にも留意 ・事前の検証及び妥当性の確認 ・セキュリティ・バイ・デザイン	**4.1(4) セキュリティ確保の原則** AIを積極的に利用することで多くの社会システムが自動化され、安全性が向上する。一方、少なくとも現在想定できる技術の範囲では、希少事象や意図的な攻撃に対してAIが常に適切に対応することは不可能であり、セキュリティに対する新たなリスクも生じる。社会は、常にベネフィットとリスクのバランスに留意し、全体として社会の安全性及び持続可能性が向上するように務めなければならない。		**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: Working to protect the privacy and security of individuals. 6d) Maximize the benefits and address the potential challenges of AI technologies, by: Ensuring that AI research and technology is robust, reliable, trustworthy, and operates within secure constraints.
Privacy プライバシー	**6) プライバシーの原則（Privacy）** 利用者及びデータ提供者は、AIシステム又はAIサービスの利活用において、他者又は自己のプライバシーが侵害されないよう配慮する。 ・AIの利活用における最終利用者及び第三者のプライバシーの尊重 ・学習等に用いるパーソナルデータの収集・前処理・提供におけるプライバシーの尊重 ・自己等のプライバシー侵害への留意及びパーソナルデータ流出の防止	**6) プライバシーの原則（Privacy）** 開発者は、AIシステムにより利用者及び第三者のプライバシーが侵害されないよう配慮する。 ・事前のプライバシー影響評価 ・プライバシー・バイ・デザイン	**4.1(3) プライバシー確保の原則** AIを前提とした社会においては、個人の行動などに関するデータから、政治的立場、経済状況、趣味・嗜好等が高精度で推定できることがある。これは、単なる個人情報を扱う以上の憤重さが求められる場合があることを意味する。パーソナルデータが本人の望まない形で流通したり、利用されたりすることによって、個人が不利益を受けることのないよう、各ステークホルダーは、以下の考え方に基づいて、パーソナルデータを扱わなければならない。	**3) 他者のプライバシ尊重(Respect for the privacy of others)** 人工知能の利用および開発において、他者のプライバシーを尊重し、関連する法規に則って個人情報の適正な取扱いを行う義務を負う。	**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: Working to protect the privacy and security of individuals.
fairness, equity, removal of discrimination 公平性	**8) 公平性の原則（Fairness）** AIサービスプロバイダ、ビジネス利用者及びデータ提供者は、AIシステム又はAIサービスの判断にバイアスが含まれる可能性があることに留意し、また、AIシステム又はAIサービスの判断によって個人が不当に差別されないよう配慮する。 ・学習等に用いられるデータの代表性やデータに内在する社会的なバイアス ・アルゴリズムによるバイアスへの留意 ・AIシステムよりなされた判断に対する人間の判断の介在	**7) 倫理の原則（Ethics）** 開発者は、AIシステムの開発において、人間の尊厳と個人の自律を尊重する。 ・（人間の脳や身体と連携するAIシステムの開発の場合）生命倫理に関する議論などの参照 ・学習データに含まれる偏見などに起因して不当な差別が生じないための所要の措置 ・AIシステムが人間性の価値を不当に毀損することがないよう留意	**4.1(6) 公平性、説明責任及び透明性の原則** AIの設計思想の下において、人々がその人種、性別、国籍、年齢、政治的信念、宗教等の多様なバックグラウンドを理由に不当な差別をされることなく、全ての人々が公平に扱われなければならない。	**4) 公正性(Fairness)** 人工知能の開発と利用において常に公正さを持ち、人工知能が人間社会において不公平や格差をもたらす可能性があることを認識し、開発にあたって差別を行わないよう留意する。人類が公平、平等に人工知能を利用できるように努める。	**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.				
Transparency 透明性 Explainability 説明可能性	**9) 透明性の原則（Transparancy）** AIサービスプロバイダ及びビジネス利用者は、AIシステム又はAIサービスの入出力の検証可能性及び判断結果の説明可能性に留意する。 ・生命、身体、財産に危害が及ぼし得る分野で利活用する場合におけるAIの入出力等のログの記録・保存 ・個人の権利・利益に重大な影響を及ぼす可能性のある分野で利活用する場合におけるAIシステム又はAIサービスの判断結果の説明可能性の確保 ・行政機関が利用する際の透明性の確保 （※アルゴリズム、ソースコード、学習データの開示を想定するものではない。）	**2) 透明性の原則（Transparancy）** 開発者は、AIシステムの入出力の検証可能性及び判断結果の説明可能性に留意する。 ・生命、身体、自由、プライバシー、財産などに影響を及ぼす可能性のあるAIシステムにおける入出力の検証可能性及び判断結果の説明可能性 （※アルゴリズム、ソースコード、学習データの開示を想定するものではない。）	**4.1(6) 公平性、説明責任及び透明性の原則** ・AIを利用しているという事実、AIに利用されるデータの取得方法や使用方法、AIの動作結果の適切性を担保する仕組みなど、状況に応じた適切な説明が得られなければならない。 ・人々がAIの提案を理解して判断するために、AIの利用・採用・運用について、必要に応じて開かれた対話の場が適切に持たれなければならない。	**6) 誠実な振る舞い(Act with integrity)** 人工知能が社会へ与える影響が大きいことを認識し、社会に対して誠実に信頼されるように振る舞う。専門家として虚偽や不明瞭な主張を行わず、研究開発を行った人工知能の技術的限界や問題点について科学的に真摯に説明を行う。	**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)	**1.3. Transparency and explainability** 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. **1.4. Robustness, security and safety** 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.	**GP5. Transparency** The basis of a particular A/IS decision should always be discoverable.	**4) Research Culture:** A culture of cooperation, trust, and transparency should be fostered among researchers and developers of AI. **7) Failure Transparency:** If an AI system causes harm, it should be possible to ascertain why. **8) Judicial Transparency:** Any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority.	7) We believe that it is important for the operation of AI systems to be understandable and interpretable by people, for purposes of explaining the technology.

AIガイドライン比較表

［別添4］
（報告書（案）別紙2）

尊重すべき価値	**AI利活用ガイドライン／AI利活用原則案 AI Utilization guidelines/Draft AI Utilization Principles**	**国際的な議論のためのAI開発ガイドライン案 Draft AI R&D guidelines for international discussions**	**「人間中心のＡＩ社会原則」 Social Principles of Human-centric AI**	**人工知能学会倫理指針 Ethical Guideline**	**Ethics Guideline for Trustworthy AI**	**Recommendation of the Council on Artificial Intelligence**	**Ethically Aligned Design**	**Asilomar AI Principles**	**Tenets**
Accountablitiy アカウンタビリティ	**10) アカウンタビリティの原則 (Accountability)** 利用者は、ステークホルダに対しアカウンタビリティを果たすよう努める。 ・アカウンタビリティを果たす努力 ・AIシステム又はAIサービスに関する利用方針の通知・公表	**9) アカウンタビリティの原則 (Accountability)** 開発者は、利用者を含むステークホルダに対しアカウンタビリティを果たすよう努める。 ・利用者にAIシステムの選択及び利活用に資する情報の提供 ・開発原則1)～8)の趣旨に鑑み、AIシステムの技術的特性についての情報提供や説明、ステークホルダとの対話を通じた意見聴取 ・AIサービスプロバイダなどとの情報共有・協力	**4.1(6) 公平性、説明責任及び透明性の原則** ・AIを利用しているという事実、AIに利用されるデータの取得方法や使用方法、AIの動作結果の適切性を担保する仕組みなど、状況に応じた適切な説明が得られなければならない。 「AI-Readyな社会」においては・・・結果に対する説明責任（アカウンタビリティ）が適切に確保されると共に、技術に対する信頼性（Trust）が担保される必要がある。	**5) 安全性(Security)** 人工知能を利用する者に対し適切な情報提供と注意喚起を行うように努める。 **6) 誠実な振る舞い(Act with integrity)** 専門家として虚偽や不明瞭な主張を行わず、研究開発を行った人工知能の技術的限界や問題点について科学的に真摯に説明を行う。 **7) 社会に対する責任(Accountability and Social Responsibility)** 潜在的な危険性については社会に対して警鐘を鳴らさなければならない。(以下略) **8) 社会との対話と自己研鑽 (Communication with society and self-development)** 人工知能に関する社会的な理解が深まるよう努める。	**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.	**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.	**GP6. Accountability** A/IS shall be created and operated to provide an unambiguous rationale for all decisions made.	**3) Science-Policy Link:** There should be constructive and healthy exchange between AI researchers and policy-makers. **4) Research Culture:** A culture of cooperation, trust, and transparency should be fostered among researchers and developers of AI. **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.	2) We will educate and listen to the public and actively engage stakeholders to seek their feedback on our focus, inform them of our work, and address their questions. 3) We are committed to open research and dialogue on the ethical, social, economic, and legal implications of AI. 4) We believe that AI research and development efforts need to be actively engaged with and accountable to a broad range of stakeholders. 5) We will engage with and have representation from stakeholders in the business community to help ensure that domain-specific concerns and opportunities are understood and addressed. 6c) Maximize the benefits and address the potential challenges of AI technologies, by: Working to ensure that AI research and engineering communities remain socially responsible, sensitive, and engaged directly with the potential influences of AI technologies on wider society. 8) We strive to create a culture of cooperation, trust, and openness among AI scientists and engineers to help us all better achieve these goals.