Outline of
"AI Guidelines for Business Appendix Ver1.1"

Ministry of Internal Affairs and Communications
Ministry of Economy, Trade and Industry
(April 4, 2025)

Positioning of main part and appendix of AI Guidelines for Business

- The main part covers "the efforts to be made regarding AI (guiding principles = what)" based on "the ideal society while considering stakeholders' expectations (basic philosophies = why)" that are important for using AI safely and securely to maximize the benefits of AI.
- The appendix covers "the specific approach to be adopted (implementation = how)" to lead AI business actors to take actual implementation of the principles.

Main part（why, what）		Appendix (how)

The ideal society while considering stakeholders' expectations
(basic philosophies = why)

The efforts to be made regarding AI
(guiding principles = what)

The specific approach to be adopted
(implementation = how)

Structure of "AI Guidelines for Business"

- The descriptions in the Appendix correspond to those in the main part and serve as a supporting document for the reading of the main part and considerations and actions based on the main part.

Main part（why, what）		Appendix (how)	
For all AI business actors — Part 1	Definitions	1. Relevant to Part 1 [About AI]	A. Preconditions for AI B. AI's benefits and risks
Part 2	Society to aim for with AI, and matters each AI business actor works on A. Basic philosophies B. Principles C. Common Guiding Principles D. Common Guiding Principles for AI business actors involved in advanced AI systems E. Building AI governance	2. Relevant to Part 2 [E. Building AI Governance]	A. Building of AI governance and monitoring by management B. Examples of business operator's efforts at AI governance
For each AI business actor — Part 3	Matters Related to AI Developers * Includes additional matters described in "Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems" as well	3. Relevant to Part 3 [For AI Developers]	A. Descriptions of Part 3 "Matters Related to AI Developers" B. Descriptions of "Common Guiding Principles" in Part 2 C. Matters to be observed in developing advanced AI systems
Part 4	Matters Related to AI Providers	4. Relevant to Part 4 [For AI Providers]	A. Descriptions of Part 4 "Matters Related to AI Providers" B. Descriptions of "Common guiding principles" in Part 2
Part 5	Matters Related to AI Business Users	5. Relevant to Part 5 [For AI Business users]	A. Descriptions of Part 5 "Matters Related to AI Business Users" B. Descriptions of "Common Guiding Principles" in Part 2
Other references		6. Major precautions for referring to "Contract Guidelines on Utilization of AI and Data" 7. Checklist 8. Cross-actor virtual cases 9. References for overseas guidelines, etc.	The appendixes 7, 8, and 9 are Japanese only.

Appendix 1. Relevant to Part 1

- Appendix 1 provides a detailed explanation of the premises related to AI and AI's benefits and risks
- Appendix 1 leads to a deeper understanding of the content of this guideline through these. explanations.

A. Preconditions for AI	B. AI's benefits and risks

- Flow of training and use of AI
  - The general flow of training and use of AI
- Overview of AI system
  - Scope of AI system
- Value chain of AI from development to use
  - The general flow of applying AI and responses of each AI business actor
- Examples of AI systems and services
  - Specific examples of typical AI systems and services and the actors involved
- Patterns of AI companies
  - Specific examples of AI value chain at the time of business utilization
- About data providers
  - Definitions of data providers and the like, that are not covered by the guidelines

- AI's benefits
  - Focus on end-users who will be primarily benefited
- AI's risks
  - Typical examples (including assumptions)
  - Systematic Classification of Risk Examples (tentative ver.) (Technical risks and social risks)
  - Key common guiding principles corresponding to each risk example, and examples of countermeasures taken by AI business actors.

Appendix 1 A. Preconditions for AI

- To deepen the understanding of AI itself, the contexts in which it is utilized, and the roles of each AI business actor, terms such as "AI systems," "AI business actors," and "data providers" are explained with specific examples (the following is an excerpt of the content).

Examples of AI systems and services (excerpt)

Case name	Used AI	Overview	AI developer	AI provider	AI business user	Non-business user
Recruitment AI	Text analysis	The recruit department of each foreign subsidiary of Company A Group **uses an AI service that provides reference information for conducting the applicant screening process based on applicants' applications.** The AI development department of Company A has created an AI model that receives past application data and decision of acceptance (judgment on whether to employ each applicant) from the recruit department of Company A (AI business user; including the recruit departments of overseas group companies) and processes them through machine learning (classification model) for supporting in making acceptance decisions.	Company A (Development department)	Company A (System department and human resource development department)	Company A Group (Recruit department)	Applicants for recruitment
Unmanned convenience store	Image analysis	Company J, which holds convenience store franchises across Japan, operates **unmanned convenience stores in which image recognition AI is used.** In the unmanned convenience stores, AI calculates the price for items taken by each customer and carries out the payment process for all the items through digital money, etc., when the customer leaves the store. An AI system for unmanned convenience stores developed by Company X is incorporated into the AI service.	Company X	Company J (AI system development department and convenience store business division)	Convenience stores	Customers of convenience stores
Cancer diagnosis AI	Text and image analyses	Using the multimodal learning, this system **imports "information of the medical history, genes, etc., of a patient (data 1)" and "endoscopic image (data 2)" to highlight areas that are highly possibly affected by cancer in real time during an endoscopic examination.** It enables physicians to observe output images and diagnose potential cancer. Company A has developed AI and provides the cancer diagnosis AI system to health facilities.	Company A (AI development department)	Company A (Healthcare IT service department)	Health facilities (System department and gastroenterology)	Patients examined

Patterns of AI business actors

5
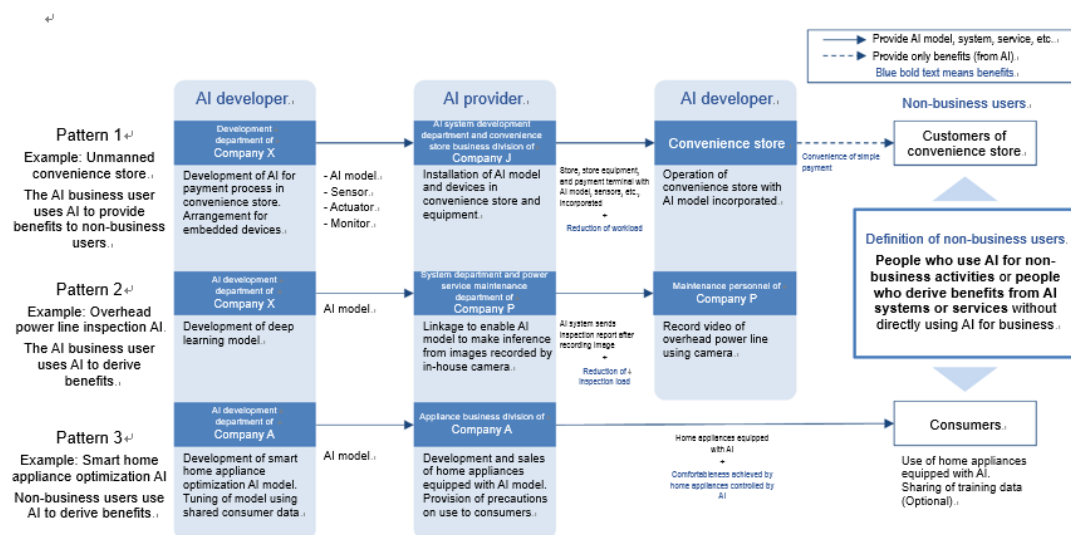
Appendix 1 B. AI's benefits and risks (AI's benefits)

- Focusing on the end-users who will benefit from AI, the benefits of AI are organized by industry and business operation.

	Development	Marketing	Sales	Logistics/distribution	Customer support	Legal	Finance	HR
Examples of benefits available conventionally (Improved by generative AI)	Automation of code verification and documentation	Automatic distribution of ad emails	Support after order intake Automatic transmission of emails, etc.	Optimization of production and inventory based on demand prediction	Automatic response through chat bot	Translation	Automatic creation of financial statements	Automation of payroll calculation, etc.
	Extraction and verification of similar code and data	Personalized ad based on data	Sales prediction for each channel and need	Optimization of delivery route	Creation of FAQ based on past inquiries	Review of legal text	Future prediction based on past records, and detection of malpractices	Human resources demand matching based on résumés, etc.
Examples of benefits unique to generative AI	Generation of training data, coding assistant, brainstorming for new products	Automatic creation of sales promotion (marketing materials, sales copy, etc.)	Automatic creation of sales talk script	Assistant for negotiation for logistics conditions	Automatic generation and summarization of transcription of support	Automatic generation of draft of contract based on stipulations	Response to in-house inquiry according to context	Holding human resources interview according to context

Appendix 1 B. AI's benefits and risks (AI's risks)

- AI's risks have been systematically organized to help AI business actors comprehensively understand and consider countermeasures for these risks.

Table 3. Systematic Classification of AI-related Risk Examples（tentative ver.）

- The table below does not cover all AI risks and includes hypothetical cases, and it is expected to be recognized as just one example.
- The table below also includes risks that require responses and discussions from society as a whole, including governmental and public institutions.

Major categories	Subcategories	Risk examples
Technical Risks (=risks primarily associated with AI systems)	Risks during the learning and input stages of AI	Attacks on AI systems such as data poisoning attacks
	Risks during the output stage of AI	Biased outputs, discriminatory outputs, and inconsistent outputs
		Incorrect outputs due to Hallucinations and similar issues
	Risks during the postresponse stage	Black-boxing and inadequate explanations of decisions
Societal Risks (=existing risks that may also arise in AI or be amplified by AI)	Risks related to ethics and law	Inappropriate use of personal information
		Occurrence of accidents related to lives, etc.
		Discrimination in triage
		Excessive dependence
		Misuse
	Risks related to economic activities	Infringement of intellectual property rights, etc.
		Financial loss
		Leak of confidential information
		Unemployment of workers
		Concentration of data and profits
		Infringement of qualifications, etc.
	Risks related to the information space	Distribution and diffusion of disinformation
		Negative influence on democracy
		Filter bubble and echo chamber phenomena
		Loss of diversity and inclusion
		Reproduction of biases
	Risks related to the environment	Energy consumption and environmental load

Appendix 1 B. AI's benefits and risks (AI's risks)

- To further connect the identified risks to the consideration of countermeasures by AI business actors, the key common guiding principles corresponding to each risk and examples of countermeasures by AI business actors have been described[*1].

Table 4. Mapping of AI's Risk Examples, Common Guiding Principles, and Important Matters for Each AI Business Actor

- The table below does not cover all AI risks and includes hypothetical cases, and it is expected to be recognized as just one example.
- The table below also includes risks that require responses and discussions from society as a whole, including governmental and public institutions.

Risk Examples	Key common guiding principles corresponding to each risk example	Important matters for each AI business actor in addition to common guiding principles		
		Part 3. AI Developer	Part 4. AI Provider	Part 5. AI Business User
Attacks on AI systems such as data poisoning attacks	5) Ensuring security	i. Deployment of mechanisms for security measures ii. Consideration for the latest trends	i. Deployment of mechanisms for security measures ii. Handling of vulnerabilities	i. Implementation of security measures
Biased outputs, discriminatory outputs, and inconsistent outputs	1) Human-centric (1) Human dignity and autonomy of individuals (3) Countermeasures against disinformation			
Incorrect outputs due to Hallucinations and similar issues	2) Safety	i. Proper data training ii. Development that takes into consideration the lives, bodies, properties, and minds of humans and the environment iii. Development contributing to proper use (of AI)	i. Actions against risks that consider the lives, bodies, properties, and minds of human and the environment ii. Provision contributing to proper use (of AI)	i. Proper use (of AI) that considers safety
	3) Fairness	i. Consideration for bias in data ii. Consideration for bias in algorithms, etc., of AI models	i. Consideration for bias in configurations and data of AI systems and services	i. Consideration for bias in input data or prompt
	8) Education/literacy			
Black-boxing and inadequate explanations of decisions	6) Transparency	i. Ensuring verifiability ii. Providing relevant stakeholders with information	i. Documentation of system architectures and the like ii. Providing relevant stakeholders with information	i. Providing relevant stakeholders with information

Technical Risks

*1 : It is advisable to refer to the relevant sections of Parts 3 to 5 of the main text and Appendices 3 to 5 for measures and specific methods for each entity.

Appendix 2. "Section 2. E. Building AI Governance"

- Appendix 2 includes "Behavioral Goals" and "Practice Guidelines" for establishing "I Governance, as well as corresponding hypothetical "Practical Examples" and examples of actual company initiatives.
- Specific examples will enable each business to consider how to establish its own AI governance.

A. Creating AI governance and monitoring by management

- Behavioral Goals
  - The general and objective goals
- Practice Guidelines
  - Based on the "Governance Guidelines for Implementation of AI Principles," incorporating elements from national and international guidelines, ISO, etc.
- Practical Examples
  - Includes examples based on hypothetical cases
  - Incorporates examples of responses to the latest trends in generative AI, etc.

B. Examples of business operators' AI governance initiatives

- Examples of business operators
  - Columns of 9 entities' (8 companies and 1 local government) AI governance initiatives based on the "Governance Guidelines for Implementation of AI Principles"
  - Describes perspectives that many companies encounter when promoting AI governance

List of entities Featured in the Column (in order of appearance)

- ABEJA
- NEC Group
- Toshiba Group
- Panasonic Group
- Fujitsu Group
- Softbank
- NTT DATA
- Ubie, Inc.
- Kobe City

Appendix 2 A. Building of AI governance and monitoring by management (Content)

- Support the establishment of AI governance by ensuring that each AI business understands the significance of and utilizes the governance behavioral goals.

Category	Behavioral Goals ※ Some are further subdivided like 「3-1-1」
1. Environmental and risk analysis	1-1 Understanding benefits/risks 1-2 Understanding social acceptance of AI 1-3 Understanding company's AI know-how
2. Goal setting	2-1 Setting AI governance goals
3. System design	3-1 Requiring evaluation of goal deviation and measures to minimize it 3-2 Improving literacy of those in charge of the AI management system 3-3 Enhancing AI management through cooperation between AI business actors and divisions 3-4 Reducing burden related to incidents involving AI Business Users and non-business users through preventive and prompt action
4. Operation	4-1 Ensuring that the operation of AI management system is explainable 4-2 Ensuring that the operation of each AI system is explainable 4-3 Considering proactive disclosure of AI governance practices
5. Evaluation	5-1 Verifying AI management system functions 5-2 Considering opinions of outside stakeholders
6. Environment and risk reanalysis	6-1 Reimplementing Behavioral Goals 1-1 to 1-3 at an appropriate time

Appendix 2 A. Building of AI governance and monitoring by management (Structure)

- Organize "Practice Guidelines" and "Practical Examples" for each "Behavioral Goal."
- Enables each AI business to consider AI governance by referring to them according to the type of AI to be employed and the degree of risk.

Appendix Content

Behavioral Goal 1-1 [Understanding benefits/risks]
AI business actors will, under the leadership of the management team, clarify the purpose of development, provision and use of AI, specifically understand that there are not only benefits offered by AI, but also unintended risks, report them to the management, share with the management and update the understanding at an appropriate time.

[Practice Guidelines]
AI business actors will, under the leadership of the management team, take the following measures:
- Specifically understand the benefits and risks including unintended risks in a way that is relevant to a company's own business.
- Establish a system for promptly reporting to / sharing with the management.
"Risks" include the following examples:
- Common risks of AI in general: Output of biased results or discriminative results, filter bubbles, echo chambers, disinformation, handling of inappropriate personal data, data pollution attacks, obscuring, leak of confidential information, abuse of AI system services, energy consumption and environmental burden, reproduction of bias, etc.

.
.
.

[Practical Examples]
[Practical Example i: Understanding benefits and risks]
It is important that AI business actors examine not only benefits, but also risks under the leadership of the management (including the examination the management itself implements by taking the lead instead of leaving it up to an officer in charge or staff in charge), share the results of examination and update the understanding at an appropriate time.
.
.

Descriptions

Behavioral Goal
- Provides general and objective goals that are important for AI businesses to work toward
- Provides material for each AI business to consider when reviewing its policies

Practice Guidelines
- Summarizes important matters and points to keep in mind to implement the above action goals
- Provides material for each AI business to consider specific actions

Practical Examples
- Provides examples of hypothetical cases in practice
- Makes it easier for each business operator to take action by having a specific example

11

Appendix 3-5. For each AI business actor (Content)

- Attachments 3 through 5 provide detailed explanations in the main part for each AI business actor.
- It can be used in conjunction with the main part to refer to and discuss specific approaches related to the main part's content.

A. Descriptions of Part 3-5	B. Descriptions of "Common Guiding Principles" in Part 2	C. Matters to be observed in developing advanced AI systems ※Only in Appendix3. For AI developers

- **Points**
  – Supplements important points in addition to those described in the main part
- **Specific Methods**
  – Specific explanations with reference to other guidelines, etc.
- **References**
  – List other references (e.g.)
    - Digital Agency "Data Quality Guidebook (β Edition)" (June 2021)
    - National Institute of Advanced Industrial Science and Technology "Guidelines for Machine Learning Quality Management, 4th Edition" (December 2023)
    - Consortium of Quality Assurance for Artificial-Intelligence-based Products and Services "AI Product Quality Assurance Guidelines" (April 2024)
    - NIST, "AI Risk Management Framework Playbook" (January 2023)
    - Council of Europe, "Risk and Impact Assessment Method for AI Systems from the Perspective of Human Rights, Democracy, and the Rule of Law (HUDERIA)" (November 2024)

- Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems

Appendix 3-5. For each AI business actor (Structure)

- For each important matter of each AI business actor, "Points," "Specific methods," and "References" are provided in the following structure:

Appendix Content

A. Descriptions of Part 3 "Matters Related to AI Developers"

[Contents of the main part (repeat)]
<u>During data preprocessing and training</u>
D-2) i. Proper data training

- Properly collect training data through privacy-by-design, etc., and if it contains <u>third-parties</u>' personal data, data requiring attention to intellectual property rights, etc., ensure that such data is properly handled in compliance with laws and regulations throughout the lifecycle of AI ("2) Safety," "4) Privacy protection," "5) Ensuring security").
- Implement proper protective measures before and across training by, for example, considering the deployment of any data management and restriction function that controls access to data ("2) Safety," "5) Ensuring security").

[Points]
  In order to improve the quality of AI models, it is important for AI Developers to pay close attention to the quality of data used for AI training, etc.
- Pay close attention to the quality (accuracy, integrity, etc.) of the data used for AI training, etc., in view of the characteristics and applications of the AI to be used.

[Specific methods]
- Verify that the data does not contain personal data, confidential information, rights including copyrights or legally protected interests.
  ➢ Extraction of unique expressions

[References]
- National Institute of Advanced Industrial Science and Technology "Guidelines for Machine Learning Quality Management, 4th Edition" (December 2023)
- NIST, "AI Risk Management Framework Playbook" (January 2023)

B. Descriptions of "Common guiding principles" in Part 2

  Although not mentioned in the Main Part, "Part 3 Matters Related to AI Developers," specific methods for the Main Part, "Part 2" "Common guiding principles," which are especially important for AI Developers, are explained here.

Descriptions

Repeat of main part content

Points
- Supplements important points in addition to those described in the main part

Specific methods
- Specific explanations with reference to other guidelines, etc.

References
- Provide references to "Points" and "Specific methods"

Descriptions corresponding to Part 3-5 in the main part

Descriptions corresponding to Part 2

Specific approaches of each AI business actor regarding the Common Guiding Principles

13

Appendix 6. Major precautions for referring to "Contract Guidelines on Utilization of AI and Data"

- Appendix 6 contains the precautions for referring to the "Contract Guidelines on Utilization of AI and Data" as of March 2025, the first edition of which was formulated and published in June 2018, due to changes in the situation regarding the development and use of AI and new technological advancements.
- By clearly defining the rights and obligations of the parties through contracts, transactions related to AI can be facilitated and unnecessary disputes can be prevented.

Contract Guidelines on Utilization of AI and Data

+

Trends in international society

Emergence of new technology

Contract Checklist for AI Utilization and Development

Appendix 6. Major precautions for referring to "Contract Guidelines on Utilization of AI and Data"

The following matters should be considered in light of changes in the situation since the release of the Contract Guidelines:

(1) Diversification of contract models
- Contracts related to the use and development of AI can be broadly classified into the following three types, with each type having different contractual considerations and negotiation points.
- Type 1: General-purpose AI Service Usage
- Type 2: Customization
- Type 3: New Development

(2) Risk distribution under a complex value chain
- Responsibilities based on diversity or complexity of value chain needs to be considered.

(3) Development, provision and use of AI, and accountability
- The following are beneficial points to consider contractually in relation to accident risk; Organizing new types of risks, Implementation of reasonable explanations, Presentation of objective evidence.

Appendix 7. Checklist
(Content)

- Appendix 7 provides a "checklist" and "worksheet for specific approaches" to ensure the planning and implementation of initiatives to reduce risks and reap benefits from AI.

| Main part・Appendix 1～5 | » | Checklist | » | Worksheet for specific approaches |

Understand the importance of AI governance and what is expected of each AI business actor by reading the main part and the Appendix.

Confirm the approaches ("What") of each AI business actor by using the checklist.

Use the "worksheet for specific approaches" to discuss the specific approaches ("how") of each entity.

15

Appendix 7. Checklist Utilization

- All AI business actors confirm an overview of their own efforts by using "Appendix 7. Checklist A [for all AI business actors]."
- If the business actor is concerned with advanced AI systems, refer also to "Appendix 7. B Checklist [for AI business actors concerned with advanced AI systems]."

Japanese only

別添7 A　チェックリスト　令和6年4月19日
[全主体向け]

本チェックリストは、AI事業者ガイドライン「第2部C.共通の指針」を要約したものです。
事業者に求められる重要な取組事項のチェックにご活用ください
※高度なAIシステムに関係する事業者は、
「チェックリスト[別添7 B.高度なAIシステムに関係する事業者向け]」も実施ください

チェック項目

- □ **人間中心**の考え方を基に、憲法が保障する又は国際的に認められた人権を侵すことがないようにしているか？
- □ AIに関わる全ての者の生命・身体・財産、精神及び環境に危害を及ぼすことがないよう**安全性**を確保しているか？
- □ 潜在的なバイアスをなくすよう留意し、それでも回避できないバイアスがあることを認識しつつ、回避できないバイアスが人権及び多様な文化を尊重する**公平性**の観点から許容可能か評価しているか？
- □ **プライバシー**を尊重・保護し、関係法令を遵守しているか？
- □ 不正操作によってAIの振る舞いに意図せぬ変更又は停止が生じることのないように、**セキュリティ**を確保しているか？
- □ **透明性**を確保するために、AI自体やAIシステム・サービスの情報をステークホルダーに対し合理的で技術的に可能な範囲で提供しているか？
- □ データの出所、AIの意思決定等のトレーサビリティに関する情報やリスクへの対応状況等について、関連するステークホルダーに対して合理的な範囲で**アカウンタビリティ**を果たしているか？
- □ **AIガバナンスやプライバシーに関するポリシー**等を策定しているか？
- □ 上記の実現のため、各事業者の状況に応じた**具体的なアプローチは検討**しているか？

検討には「**具体的なアプローチ検討のためのワークシート**」をご活用ください

別添7 B　チェックリスト
[高度なAIシステムに関係する事業者向け]

本チェックリストは、AI事業者ガイドライン「第2部D.高度なAIシステムに共通の指針」を要約したものです。
高度なAIシステムに関係する事業者に該当する事業者に求められる重要な取組事項のチェックにご活用ください(高度なAIシステムを開発するAI開発者にのみ適用される内容もあるため、AI提供者及び…

チェック項目

- □ 高度なAIシステムの市場導入前及び開発全体を通じて、AIライフサイクル全体にわたる**リスクを特定、評価、軽減**するための適切な措置を講じているか？
- □ 市場投入後に**脆弱性、インシデント、悪用パターンを特定、緩和**しているか？
- □ 十分な透明性の確保やアカウンタビリティの向上のため高度なAIシステムの**能力、限界、適切・不適切な使用領域を公表**しているか？
- □ 産業界、政府、市民社会、学界を含む関係組織間で、**責任ある情報共有とインシデントの報告**に努めているか？
- □ **リスクベースのアプローチにもとづくAIガバナンス及びリスク管理方針を策定、実施、開示**しているか？
- □ AIのライフサイクル全体にわたり、**物理的セキュリティ、サイバーセキュリティ、内部脅威を含む、強固なセキュリティ管理に投資**し、実施しているか？
- □ 技術的に可能な場合は、AIが生成したコンテンツを識別できるように、**電子透かしやその他の技術等、信頼性の高いコンテンツ認証及び来歴のメカニズムを開発、導入**しているか？
- □ 社会的、安全、セキュリティ上の**リスクを軽減するための研究を優先**し、効果的な軽減策に優先的に投資しているか？
- □ **気候変動、健康・教育などの世界の最大の課題**に対処するため、高度なAIシステムの開発を優先しているか？
- □ 国際的な技術規格の開発を推進しているか？
- □ **適切なデータ入力措置**と、個人データ及び知的財産の保護を実施しているか？
- □ 誤情報の拡散等のAI固有リスクに関する**デジタルリテラシー**の向上や脆弱性の検知への協力と情報共有等、高度なAIシステムの**信頼でき責任ある利用**を促進し、貢献しているか？
- □ …上記の実現のため、各事業者の状況に応じた**具体的なアプローチは検討**しているか？

検討には「**具体的なアプローチ検討のためのワークシート**」をご活用ください

- The checklist is a summary of the main part
- With the checklist, **an overview of the status of their own initiatives** can be obtained

- Utilize the **"worksheet for specific approaches"** to examine specific practices (see next page for how to utilize the worksheet).

16

Appendix 7. Worksheet for specific approaches Utilization

- Describes important matters when considering specific approaches to the guideline content.
- Presumes that each business will customize and utilize this worksheet according to its own business activities and circumstances.

Japanese only

別添７Ｃ．具体的なアプローチ検討のためのワークシート (共通の指針関連)

利用上の留意点

- Before using the worksheet, **decide who will be responsible for its implementation**.
- Each business will **customize its own approach** according to the nature of its business, its circumstances, etc.
- **Periodically update** by considering and writing the "review date".
- The worksheet contains important information when considering specific approaches and can be used as **a reference for businesses to customize** their own approaches.

Appendix 8. Attachment 8. Hypothetical case study across AI business actors (Overview)

- Attachment 8. provides a "hypothetical case study across AI business actors" in which AI developers, AI providers, and AI business users consider important matters in accordance with the guideline.
- Enables each actor to recall specific cases in which they actually incorporated the content of the guideline and clarify the points where cooperation is essential.

Japanese only

Case	採用AI				

機械学習モデル：XGBoost（エントリーシートの文章で、応募者に対して合否を判断する）
https://ifi.u-tokyo.ac.jp/wp/wp-content/uploads/2022/10/RCModel_Case01_Recruitment-AI_JP.pdf

			AI開発部門	人材採用部門（採用AIチーム）	人材採用担当者
			AI開発者	AI提供者	AI利用者
No	分類	共通の指針/各主体に関する事項	本UCにおいて主体が実施している活動	本UCにおいて主体が実施している活動	本UCにおいて主体が実施している活動
1) 人間中心					
各主体は、AIシステム・サービスの開発・提供・利用において、後述する各事項を含む全ての取り組むべき事項が導出される土台として、少なくとも憲法が保障する又は国際的に認められた人権を侵すことがないようにすべきである。また、AIが人々の能力を拡張し、多様…					
①人間の尊厳及び個人の自律					
1	共通	AIが活用される際の社会的文脈を踏まえ、人間の尊厳及び個人の自律を尊重する	AIシステムの開発において、学習データの収集やラベリング、モデルの性能評価等は、AI開発者だけで完結せず、AI提供者側で	AIサービスの提供において、AI利用者が最終判断(応募者の合否)を行えるようになっている(Human-in-the-loop)	AIサービスの提供において、AI利用者が最終判断(応募者の合否)を行えるようになっている(Human-in-the-loop)
2	共通	特に、AIを人間の脳・身体と連携させる場合には、その周辺技術に関する情報を踏まえつつ、諸外国及び研究機関における生命	脳・身体と連携するケースではないため対象外	脳・身体と連携するケースではないため対象外	脳・身体と連携するケースではないため対象外
3	共通	個人の権利・利益に重要な影響を及ぼす可能性のある分野においてAIを利用したプロファイリングを行う場合、個人の尊厳を尊重し、アウトプットの正確性を可能な限り維持させつつ、AI の予測、推奨、判断等の限界を理解して利用し、かつ生じうる不利益等を慎重に検討した上で、不適切な目的に利用しない	AIシステムの開発において、実際の予測結果を学習データに用いる際には個人情報の取扱いに関わる誓約書の締結やアクセス権管理等を実施している。※公平性とプライバシーについては、「３）公平性」「４）プライバシー保護」を参照	AIシステムの開発において、実際の予測結果を学習データに用いる際には個人情報の取扱いに関わる誓約書の締結やアクセス権管理等を実施している。※公平性とプライバシーについては、「３）公平性」「４）プライバシー保護」を参照	AIシステムの開発において、実際の予測結果を学習データに用いる際には個人情報の取扱いに関わる誓約書の締結やアクセス権管理等を実施している。※公平性とプライバシーについては、「３）公平性」「４）プライバシー保護」を参照
②AIによる意思決定・感情の操作等への留意					
1	共通	人間の意思決定、認知等、感情を不当に操作することを目的とした、又は意識的に知覚できないレベルでの操作を前提としたAIシステム・サービスの開発・提供・利用は行わない	本ケースに関しては、2)①-3と同じ論点になる	本ケースに関しては、2)①-3と同じ論点になる	本ケースに関しては、2)①-3と同じ論点になる
2	共通	AIシステムの開発・提供・利用において、自動化バイアス等のAIに過度に依存するリスクに注意を払い、必要な対策を講じる	本ケースに関しては、2)①-3と同じ論点になる	本ケースに関しては、2)①-3と同じ論点になる	本ケースに関しては、2)①-3と同じ論点になる

Appendix 9. Comparison with international guidelines (Overview)

- Appendix 9. summarizes the correspondence between the elements of the guideline and other guidelines, etc.
- This is to be used to clarify the correspondence between the guideline and other (domestic and international) guidelines, etc., when the AI business actor is considering more detailed content.

Primary guidelines, etc. used as references

Japanese only

- Ethics guidelines for trustworthy AI（April 2019, EU）
- Recommendation of the Council on Artificial Intelligence（May 2019, OECD Cabinet Meeting）
- Recommendation on the Ethics of Artificial Intelligence（November 2021, UNESCO）
- Blueprint for an AI Bill of Rights（October 2022, THE WHITE HOUSE）
- Artificial Intelligence Risk Management Framework（AI RMF 1.0）（January 2023, NIST）
- Advancing accountability in AI（February 2023, OECD）
- Proposal for a Regulation laying down harmonized rules on artificial intelligence（AI Act）（April 2021, EU）
- AI Product Quality Assurance Guidelines
  （April 2024, Consortium of Quality Assurance for Artificial-Intelligence-based Products and Services）
- Guidelines for Machine Learning Quality Management, 4th Edition
  （December 2023, National Institute of Advanced Industrial Science and Technology）

Guidelines, etc. cited or integrated into the guideline

- Human-centered AI social principles（March 2019, Integrated Innovation Strategy）
- The Draft AI R&D GUIDELINES for International Discussions（July 2017, The Conference toward AI Network Society）
- AI Utilization Guidelines -Practical Reference for AI utilization-（August 2019, The Conference toward AI Network Society ）
- Governance Guidelines for Implementation of AI Principles Ver. 1.1
  （January 2022, Expert Group on How AI Principles Should be Implemented）
- Contract Guidance On Utilization Of AI And Data（March 2018, Ministry of Economy, Trade and Industry）
- The Hiroshima AI Process Comprehensive Policy Framework（December 2023, G7）