

## Crosswalk 1 – Terminology

### NIST AI Risk Management Framework (NIST AI RMF) and Japan AI Guidelines for Business (AI GfB)

NIST AI RMF 1.0 - Characteristics of Trustworthy AI Systems	Japan AI GfB - Common Guiding Principles
<p><b>Valid &amp; Reliable –</b> <i>(Includes accuracy and robustness)</i></p> <p><b>Validation:</b> “confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled”<sup>1</sup></p> <p><b>Reliability:</b> “ability of an item to perform as required, without failure, for a given time interval, under given conditions”<sup>2</sup></p> <p><b>Accuracy:</b> “closeness of results of observations, computations, or estimates to the true values or the values accepted as being true”<sup>2</sup></p> <p><b>Robustness:</b> “ability of a system to maintain its level of performance under a variety of circumstances”<sup>2</sup></p> <p><sup>1</sup> ISO 9000:2015 <sup>2</sup> ISO/IEC TS 5723:2022</p>	<p><b>Validation:</b> <i>(There is no definition for validation. Instead, as an element of transparency, the AI GfB indicates the importance of ensuring the verifiability of the AI systems and services as necessary and technically possible.)</i></p> <p><b>Reliability:</b> The AI works satisfactorily for the requirements, including the accuracy of its output</p> <p><b>Accuracy:</b> The AI works satisfactorily for the requirements</p> <p><b>Robustness:</b> Maintaining performance levels under a variety of conditions and avoiding significantly incorrect decisions regarding unrelated events</p> <p><u>AI GfB Context</u> 2) Safety <i>(Includes accuracy, reliability, and robustness)</i> (1) Consideration for human life, body, property and mind as well as the environment (3) Proper training 6) Transparency (1) Ensuring verifiability</p>

<p><b>Safe –</b></p> <p><b>Safe:</b> AI systems should “not under defined conditions, lead to a state in which human life, health, property, or the environment is endangered”<sup>1</sup></p> <p><sup>1</sup> ISO/IEC TS 5723:2022</p>	<p><b>Safety:</b> Through the development, provision, and use of AI systems and services, each AI business actor should not harm the lives, bodies, and property of stakeholders involved with AI.</p> <p>Note: As a complimentary concept to safety, the AI GfB proposes <b>Human-centric</b> (see definition in ‘Other Terminology’ listed below).</p> <p><u>AI GfB Context</u></p> <p>1) Human-centric  (3) Measures against disinformation, etc.</p> <p>2) Safety  (1) Consideration for human life, body property and mind as well as the environment  (2) Proper use</p>
<p><b>Secure &amp; Resilient –</b></p> <p><b>Secure:</b> AI systems that can maintain confidentiality, integrity, and availability through protection mechanisms that prevent unauthorized access and use may be said to be secure.</p> <p><b>Resilient:</b> AI systems, as well as the ecosystems in which they are deployed, may be said to be resilient if they can withstand unexpected adverse events or unexpected changes in their environment or use – or if they can maintain their functions and structure in the face of internal and external change and degrade safely and gracefully when this is necessary.<sup>1</sup></p> <p><sup>1</sup>Adapted from: ISO/IEC TS 5723:2022</p>	<p><b>Security:</b> To maintain confidentiality, integrity and availability of AI systems and services, and always ensure the safe use of AI, to take reasonable measures in light of current technological level.</p> <p>Note: Recognize that it is not possible to completely eliminate vulnerabilities in AI systems and services.</p> <p><b>Resilient:</b> <i>(There is no clear mention of resilience in the AI GfB. Rather, the AI GfB focuses on security, while also recognizing the vulnerabilities in AI systems and services.)</i></p> <p><u>AI GfB Context</u></p> <p>5) Ensuring security  (1) Security measures affecting AI systems and services  (2) Attention to the latest trends</p>

**Accountable & Transparent –**

**Transparency:** reflects the extent to which information about an AI system and its outputs is available to individuals interacting with such a system – regardless of whether they are even aware that they are doing so

**Accountability:** In the development, provision, and use of AI systems and services, it is important for each AI business actor to fulfill accountability to its stakeholders to a reasonable extent with regard to ensuring traceability and compliance with “Common Guiding Principles”.

Note: Accountability is sometimes defined as explainability. But in this document, information disclosure is handled by transparency, and accountability refers to the assumption of de facto and legal responsibility for AI and to the prerequisites for assuming that responsibility.

**Transparency:** In the development, provision, and use of AI systems and services, it is important for AI business actor to provide appropriate information to stakeholders within a reasonable scope considering the social context in which the AI systems and services are used and ensuring the verifiability of the AI systems and services as necessary and technically possible.

Note: Matters related to information disclosure are broadly referred to as ‘transparency’.

AI GfB Context

- 1) Human-centric
  - (5) User support
- 2) Safety
  - (1) Consideration for human life, body property and mind as well as the environment
  - (3) Proper training
- 6) Transparency
  - (1) Ensuring verifiability
  - (2) Providing information to related stakeholders
  - (3) Reasonable and honest response
- 7) Accountability
  - (1) Improving traceability

	<ul style="list-style-type: none"> <li>(2) Explanation of compliance status of “Common Guiding Principles”</li> <li>(3) Clarification of responsible person</li> <li>(4) Distribution of responsibilities between parties</li> <li>(5) Specific responses to stakeholders</li> </ul>
<p><b>Explainable &amp; Interpretable –</b></p> <p><b>Explainability:</b> refers to a representation of the mechanisms underlying AI systems’ operation</p> <p><b>Interpretability:</b> refers to the meaning of AI systems’ output in the context of their designed functional purposes</p>	<p><b>Explainability:</b> <i>(There is no definition for explainability.)</i></p> <p><b>Interpretability:</b> <i>(There is no definition for interpretability.)</i></p> <p>Note: Improving explainability and interpretability to related stakeholders is associated with the action to analyze and understand what kind of explanation is required and take necessary measures for the purpose of gaining a sense of satisfaction and security feeling from related stakeholders, as well as presenting evidence of AI operations.</p> <p><u>AI GfB Context</u></p> <ul style="list-style-type: none"> <li>6) Transparency <ul style="list-style-type: none"> <li>(1) Ensuring verifiability</li> <li>(2) Providing information to related stakeholders</li> <li>(4) Improving explainability and interpretability to related stakeholders</li> </ul> </li> <li>7) Accountability <ul style="list-style-type: none"> <li>(6) Documentation</li> </ul> </li> </ul>

<p><b>Privacy-enhanced –</b></p> <p><b>Privacy:</b> refers generally to the norms and practices that help to safeguard human autonomy, identity, and dignity. These norms and practices typically address freedom from intrusion, limiting observation, or individuals’ agency to consent to disclosure or control of facets of their identities (e.g., body, data, reputation).</p>	<p><b>Privacy:</b> <i>(There is no definition for privacy.)</i></p> <p>Note: As a complimentary concept to privacy, the AI GfB proposes <b>Human-centric</b> (see definition in ‘Other Terminology’ listed below).</p> <p><u>AI GfB Context</u></p> <p>1) Human-centric  (1) Human dignity and individual autonomy</p> <p>4) Privacy protection  (1) Privacy protection in general AI systems and services</p>
<p><b>Fair with harmful bias managed –</b></p> <p><b>Fair:</b> Fairness in AI includes concerns for equality and equity by addressing issues such as harmful bias and discrimination. Standards of fairness can be complex and difficult to define because perceptions of fairness differ among cultures and may shift depending on application.</p> <p><b>Harmful Bias Managed:</b> Bias exists in many forms and can become ingrained in the automated systems that help make decisions about our lives. While bias is not always a negative phenomenon, AI systems can potentially increase the speed and scale of biases and perpetuate and amplify harms to individuals, groups, communities, organizations, and society. NIST identifies three major categories of AI bias to be managed (below), all of which can occur without prejudice, partiality, or discriminatory intent.</p> <p><b>Systemic bias:</b> can be present in AI datasets, the organizational norms, practices, and processes across the AI lifecycle, and the broader society that uses AI systems.</p> <p><b>Computational and statistical biases:</b> can be present in AI datasets and algorithmic processes, and often stem from systematic errors due to non-</p>	<p><b>Fairness:</b> In the development, provision, and use of AI systems and services, it is important for AI business actors to make efforts to do away with prejudice and discrimination not to impose unfair or harmful actions against specific individuals or groups based on diverse backgrounds such as race, gender, nationality, age, political opinion, religion, etc. In addition, it is important to recognize that some biases cannot be avoided, and evaluate whether these unavoidable biases are acceptable from the perspective of respecting human rights and diverse cultures. It is important to develop, provide, and use AI systems and services based on this understanding.</p> <p>Note: There are no categories of biases, though the AI GfB presupposes a variety of biases.</p> <p><u>AI GfB Context</u></p> <p>1) Human-centric</p>

<p>representative samples.</p> <p><b>Human-cognitive biases:</b> relate to how an individual or group perceives AI system information to make a decision or fill in missing information, or how humans think about purposes and functions of an AI system. Human-cognitive biases are omnipresent in decision-making processes across the AI lifecycle and system use, including the design, implementation, operation, and maintenance of AI.</p>	<ul style="list-style-type: none"> <li>(2) Attention to decision-making and emotional manipulation by AI</li> <li>(3) Measures against disinformation, etc.</li> <li>(4) Ensuring diversity and inclusion</li> </ul> <p>3) Fairness</p> <ul style="list-style-type: none"> <li>(1) Consideration of bias included in each component technology of AI model</li> </ul>
	<p><b>Other Terminology –</b> <i>(The following terms are not defined in the NIST AI RMF.)</i></p> <p><b>Human-centric:</b> In the development, provision, and use of AI systems and services, each AI business actor must not at least violate the human rights guaranteed by the Constitution or internationally recognized, as the basis for deriving all matters to be addressed, including the matters described below.</p> <p><b>Education/Literacy:</b> Each AI business actor is expected to provide the necessary education so that those involved in AI within each AI business actor can have the knowledge, literacy, and ethical sense to properly understand AI and use it socially. Furthermore, each AI business actor is expected to educate stakeholders, taking into consideration the complexity of AI, its characteristics such as misinformation, and the possibility of intentional misuse.</p> <p><b>Ensuring fair competition:</b> Each AI business actor is expected to strive to maintain a fair competitive environment surrounding AI so that new business services that utilize AI are created and sustainable economic growth is maintained, and solutions to social issues are presented.</p> <p><b>Innovation:</b> Each AI business actor is expected to strive to contribute to promoting innovation in society as a whole.</p>