J-AISI Business Demonstration WG

Data Quality sub-working group

IT Promotion Agency, Japan(IPA)
Digital Infrastructure Center
Teppei Sakamoto

Jan. 15, 2026

Hiroshima Global Forum for Trustworthy AI

AISI Japan AI Safety Institute

Who we are

Digital Infrastructure Center,
Information-technology Promotion Agency,
Japan(IPA) | Japan AI Safety Institute(J-AISI)

Teppei Sakamoto

2018: NTT DATA Group Corporation
- AI/ML, Data Science, Statistics, etc.
- [XAI (Explainable AI)] (RIC TELECOM, 2021) co-author,

2025: Researcher, IPA [Current]

Japan AI Safety Institute (J-AISI)

AISI Steering Committee

Business Demonstration WG

...

Data Quality SWG

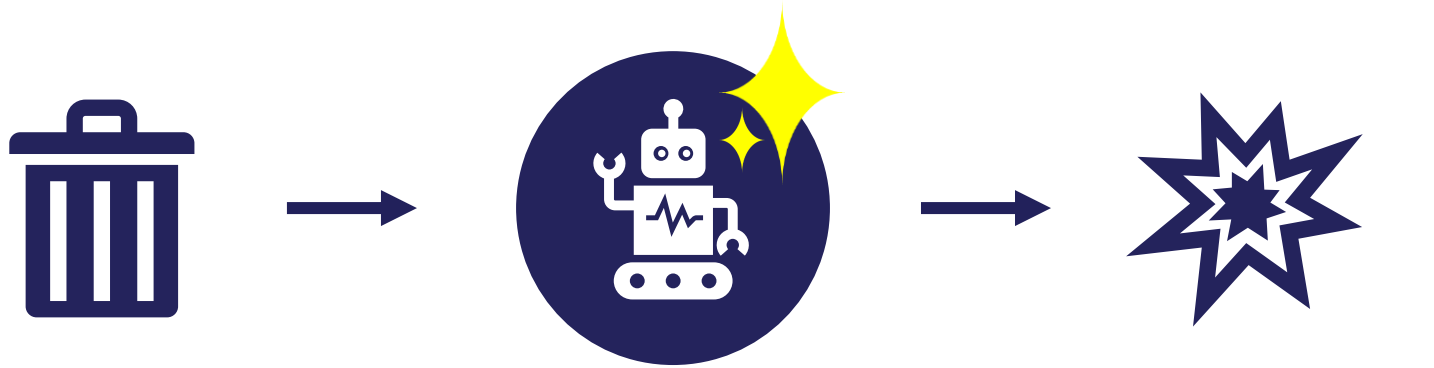
As of Jan. 2026

- **Information-technology Promotion Agency, Japan (Leader)**
- Fujitsu Limited
- Kanematsu Corporation
- National Institute of Advanced Industrial Science and Technology
- National Printing Bureau
- NEC Corporation
- NTT DATA Japan Corporation
- NTT DATA Value Engineer Corporation
- Preferred Networks, Inc.
- PricewaterhouseCoopers Japan LLC
- SAS Institute Japan Ltd.
- WingArc1st Inc.

Data Quality: The Foundation of AI Safety

AISI Japan AI Safety Institute

"Garbage in, Garbage out"



- Data is the foundation—models and training depend on it
- Poor data quality leads to unreliable and unsafe AI outputs
- Understanding data quality challenges is the first step

Typical Data Quality Challenges in the AI Era

AISI Japan AI Safety Institute

Typical data quality challenges in the AI era— making AI smarter won't fix these.

Operational layer

- **Outdated or poorly maintained data**
  - AI outputs become unreliable when data is not updated or curated.
- **Lack of data provenance and transparency**
  - Users cannot judge whether results are reliable if the origin and processing of data are unclear.
- **Bias embedded in data**
  - Biased data leads to biased AI outputs, even when models are well designed.
- **Low machine readability**
  - Data stored in formats that AI cannot properly interpret causes incorrect or misleading results.

lead → Unsafe Output

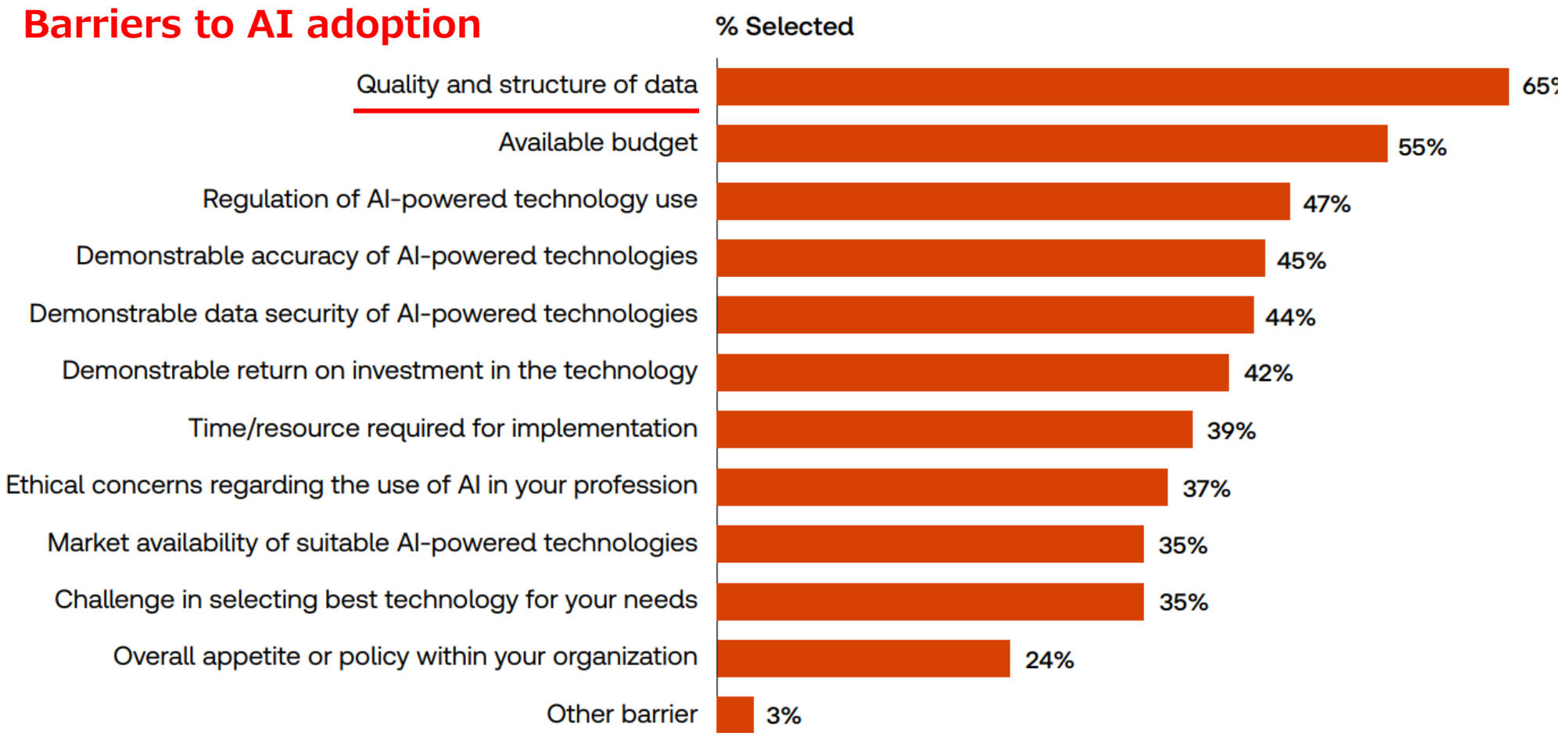
Management & governance layer

- **Gap between standards and practice**
  - While many data quality standards exist, practitioners struggle to apply them in real projects.
- **Immature data quality management in the AI era**
  - International consensus takes time, but AI deployment—especially generative AI—moves much faster.
- **Low organizational prioritization of data quality**
  - Compared to AI applications, data quality is often seen as a supporting function and receives insufficient investment

Barriers to AI adoption perceived by CxOs worldwide

AISI Japan AI Safety Institute

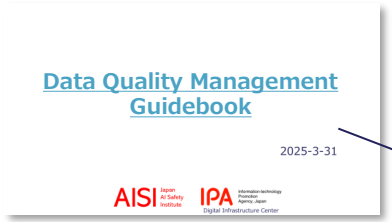
Many companies are adopting AI,
but **the most significant challenge is data.**



(Ref) Thomson Reuters Institute　2025 C-Suite Survey

Data Quality Management Guidebook

- Published **AI data quality management** guidebook
- By viewing data for AI from three perspectives—**Lifecycle**, **Gateway**, and **Governance**—we can reassess data quality from multiple angles.
- This guide consolidates a wide range of **data-quality standards into a practical**, easy-to-apply framework.

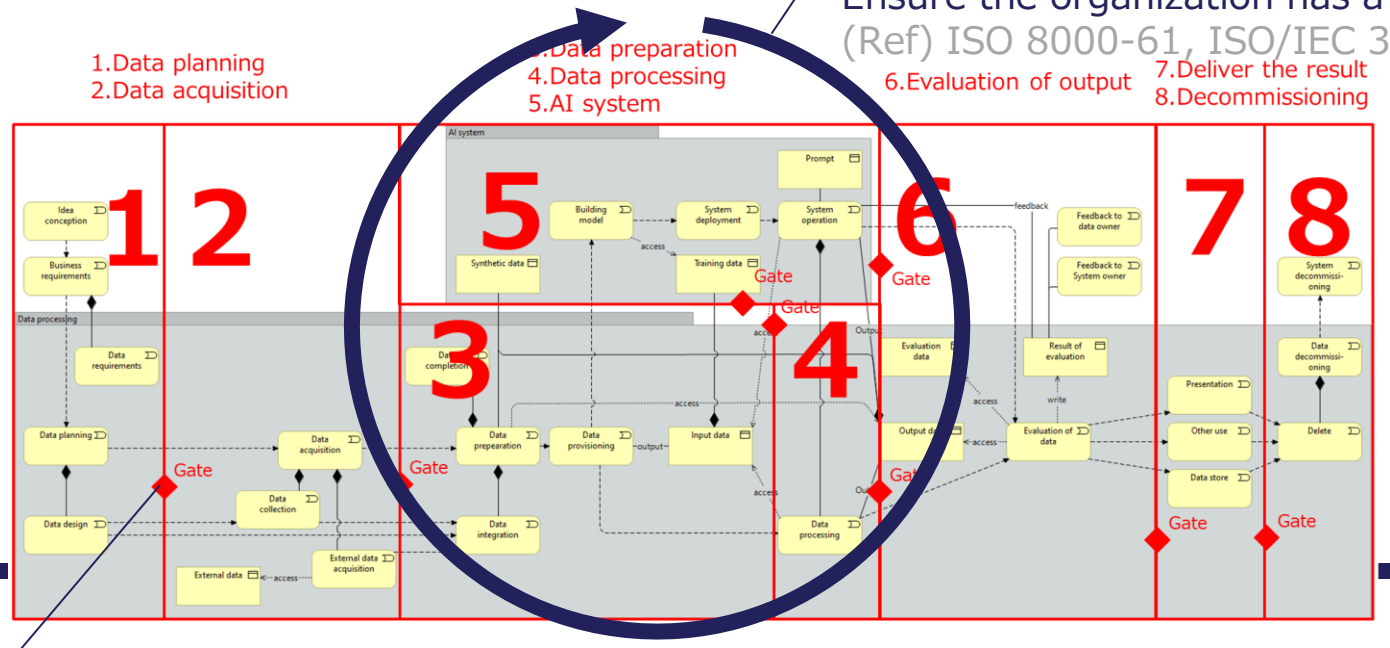


Governance cycle view

Ensure the organization has a sustainable structure.
(Ref) ISO 8000-61, ISO/IEC 38505-1, etc.

Lifecycle view

Track data quality from collection to AI deployment and decommissioning
(Ref) ISO/IEC 25024, ISO/IEC 8183, ISO/IEC 5259-1, etc.



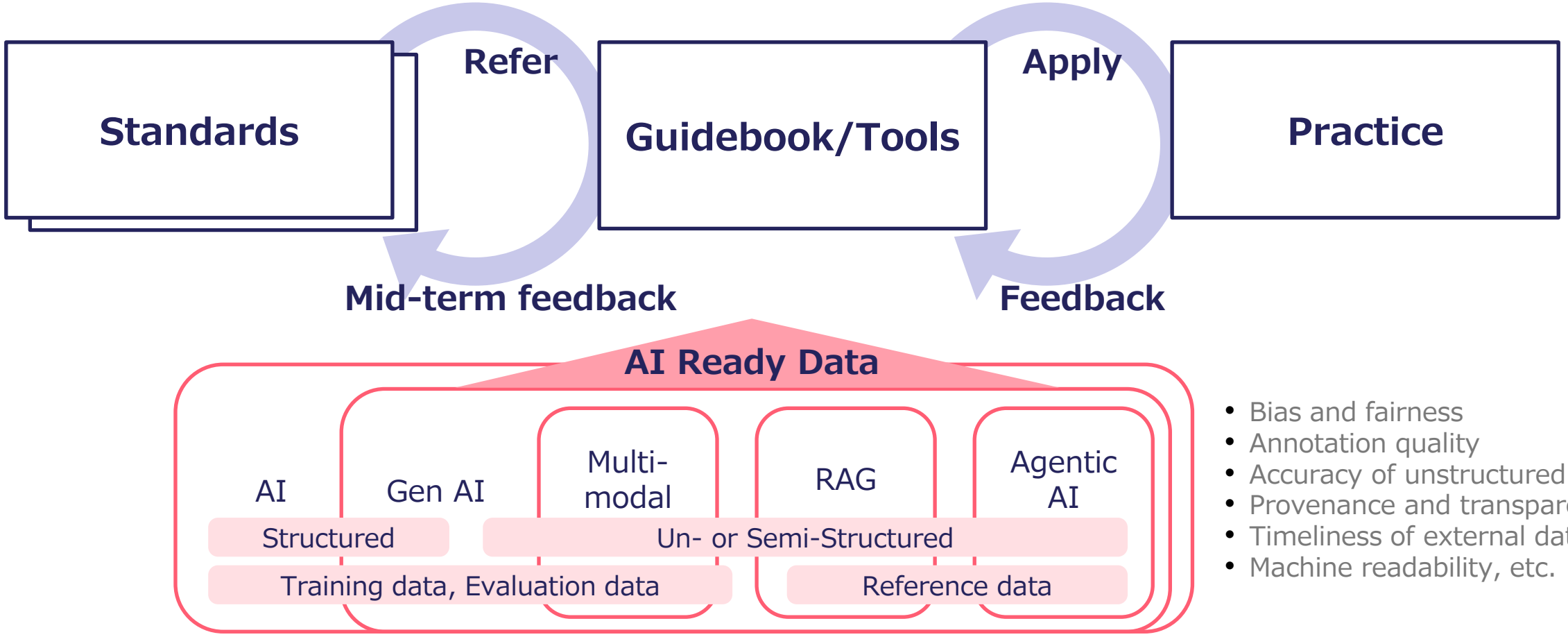
Gateway view (Characteristics)

Define quality criteria at key decision points
(Ref) ISO/IEC 25012, ISO/IEC 5259, etc.

Data quality SWG's Approach: From Standards to Practice

AISI Japan AI Safety Institute

We complement existing data quality standards with practice-oriented guidance and tools, **bridging standards and real-world practice.**



- Bias and fairness
- Annotation quality
- Accuracy of unstructured data
- Provenance and transparency
- Timeliness of external data
- Machine readability, etc.

Planned Deliverables

AISI Japan AI Safety Institute

We are incrementally updating the guidebook and developing supporting tools based on real-world use.

released

Data Quality Management guidebook v1.0

- Created to cover multiple standards
- Created internally by IPA

Application and feedback from WG participants

Data Quality Management guidebook (revised)

Make a mini-revision to improve the quality of the document(**v1.0.1 released Dec. 2025**)
To be used as a master, to be checked and revised for coverage.

Simplified guidebook

An **easy-to-use**, simplified guide focusing on essential cross-use-case controls and key controls for specific use cases.

Tools

Tools to assist in quality management

Qualitative assessment tool

Self-checking tool

Quantitative assessment tool

Automated evaluation

Case studies
(good practice/bad practice)

Trial report

All tools are under continuous development

Feedback from Practitioners on the Guidebook

AISI Japan AI Safety Institute

The guidebook was **well received conceptually**, but practitioners found it **difficult to apply directly in daily operations**.

✓ What was positively evaluated

- **Clear core message**
  - Emphasizes *"Garbage in, Garbage out,"* highlighting that AI success fundamentally depends on data quality.
- **Alignment with international standards**
  - Grounded in ISO and other international standards, ensuring credibility and legitimacy.
- **Comprehensive coverage**
  - Provides a broad, high-level view of data quality across governance, processes, and characteristics.

→ What needs improvement

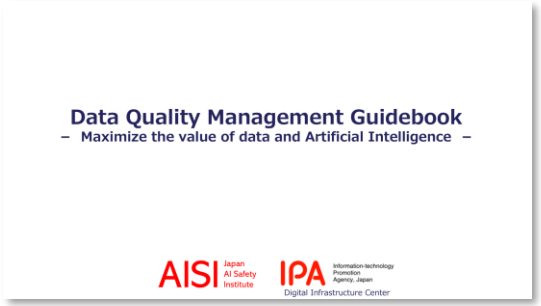
- **Gap between standards and practice**
  - Still insufficient practical guidance for real-world projects.
- **Clarity of value and use cases**
  - Need clearer ROI of data quality and concrete, use-case-specific examples.
- **AI-era specific challenges**
  - Need guidance on data–AI version management and modality-specific issues (text, image, video, etc.).
- **Document usability**
  - Improve navigation, terminology, and overall readability for non-experts and busy practitioners.

Developing a Practical Checklist from the Guidebook

AISI Japan AI Safety Institute

Based on the guidebook, we are developing a **practical checklist** for real-world use with data quality SWG and domain-specific SWGs.

Guidebook



Checklist

Quality improvement in practice

- Data Quality SWG
- Domain-specific SWGs

Feedback

First round of pilot application is currently underway

- Spreadsheet-based checklist
- 70+ items covering key data quality considerations
- Not intended for full compliance, but to raise awareness and support practitioners' decision-making.

A Case Study Using Japan's Official Gazette Data

AISI Japan AI Safety Institute

Using Japan's Official Gazette as an example of high-value public data, we observe **how data quality levels affect AI response and safety**.

Which laws were amended? User

AI response AI

1 2 3 4

Quality Level

1	2	3	4
Raw	**Digitized**	**Structured**	**Semantically Grounded**
Unstructured, multi-column layout, OCR errors included	Text is digitized, but not organized	Noise removed, limited to legal content	Enables entity resolution and precise identification

What this examination aims to clarify

- Accuracy and risk of AI responses depending on data quality level*(Effectiveness & Safety)*
- Trade-off between accuracy and preparation cost*(Cost–effectiveness)*

This examination does not validate the effectiveness of the checklist itself. Rather, it illustrates one of the key issues addressed by the checklist— machine readability and data structuring—through a concrete example. This case study is for working-group level discussion.

Future directions: From Implementation to Standards

AISI Japan AI Safety Institute

We aim to contribute **Japan-originated**, **practice-driven** frameworks to global AI discussions, reflecting **Japan's culture of quality**.

Short-term Initiatives (FY 2025)	Mid-term Initiatives (FY 2026–2027)	Long-term Initiatives (Future Vision)
• Revise guidebook • Develop and test checklists • Validate applicability through pilot use	• Enhance assessment tool capabilities • Cross-domain models • Address new data and usage types (e.g., multimodal, multi-agent)	• Establish continuous data quality models • Update guidebook and tools • Contribute Japanese frameworks to international discussions

Key Messages

- **Data quality is the foundation of AI safety.**
- We are developing a **data quality management guidebook**.
- We are collecting **practitioner feedback** for real-world use.
- We are building an actionable **checklist** to support practice.
- We are illustrating the approach through **case study using Official Gazette** data.

How you can engage

- Access our resource, share your cases and discuss with us.
  https://aisi.go.jp/output/output_framework/data_quality_management_guidebook/

Thank you!

AISI

Japan AI Safety Institute