

OECD AI Transparency Report

Organization: Fujitsu (JPN) Reporting Period: Q3 2026 Published: October 2, 2026

Section 0 – Your role in the AI system lifecycle

Responding as (please select all that apply):

- Model developer (/provider)
- Application developer (/provider)
- Deployer

If any of the fields above are selected, please indicate your organization size:

Large enterprise: 250 or more employees

Section 1 – Risk Identification and Evaluation

1. How does your organization define and/or classify different types of risks related to AI?¹

- Maintain an internal taxonomy, classification system, and/or set of risk area definitions
- Align with specific external frameworks, standards, reports, and/or guidance documents such as taxonomies, classification systems, and/or sets of definitions for risk areas (e.g. the NIST AI Risk Management Framework, the ISO/IEC 42001, the OECD AI Principles, the EU General-Purpose AI Code of Practice, the International AI Safety Report). (please specify): OWASP Top 10 for Large Language Model Applications; MITRE ATLAS Risk categories under AI regulatory laws in various jurisdictions, including ISO/IEC 42001, the OECD AI Principles, and the EU AI Act
- Consult with external stakeholders (e.g., from academia, industry groups, civil society, or standard development organisations) when developing or updating risk taxonomies

How does your organization define thresholds at which severe risks, unless adequately mitigated, may be considered "unreasonable"? Please describe the framework(s) most relevant to your organization's role and context. If your organisation does not use formal thresholds, you may describe instead the policies, decision criteria, or governance processes you apply when assessing severe AI risks. You may also describe: • Specific examples of contexts in which your models or systems could, in your assessment, pose unreasonable or intolerable risk; • How mitigations are graduated by capability or propensity level; • How your organisation considers the specificities of different AI technologies, such as generative AI and agentic AI, including any specific risk assessment or governance mechanisms applied to agentic AI capabilities.

Rather than managing risks by a single threshold, the company comprehensively evaluates risks from the perspectives of applicability under AI regulatory laws in each jurisdiction, including prohibited AI

systems under the EU AI Act, AI ethics, and AI safety. If the relevant risk is not resolved, the company operates a process under which the business promotion division stops business promotion through a review committee conducted at the completion of the sales opportunity or planning phase.

2. Both before and after deployment, what practices does your organization use to identify and evaluate vulnerabilities, incidents, emerging risks and misuse throughout the AI system lifecycle?

- Adversarial testing to detect technical vulnerabilities
- Internal benchmarks, standard metrics (e.g., frequency, severity scores)
- Structured techniques to explore the behavior or performance of a model and/or application, such as model elicitation or bottleneck assessments
- Incident reports, including reports shared by other organizations
- Specific risk identification practices for agentic AI capabilities, e.g., autonomous action-taking, tool use, multi-step planning (please specify): Agentic AI Scanner, etc.
- Multilingual and cross-cultural evaluation, including collaboration with local ecosystems (e.g., national AI institutes, language-specific evaluation consortia) for under-represented languages and cultural contexts (please specify): Publicly released multilingual trust benchmark (MAPS) [https://huggingface.co/datasets/Fujitsu-FRE/MAPS_Verified]
- Other(s) (please specify): Fujitsu Security Science Laboratory is conducting research and development on vulnerability evaluation, monitoring, and defense technologies for risks arising in systems where multiple AI agents collaborate, including agent-to-agent interactions, external tool use, external interfaces such as MCP/A2A, and information propagation through RAG and memory. Specifically, the Agentic AI Scanner detects vulnerabilities stemming from known attack methods against agent systems and combinations of such attacks, while agent monitoring technologies analyze and detect anomalous events from execution logs of multi-agent systems.

Please briefly describe your organization's decision-making process for deploying an AI model or system including: who has decision authority; what evidence and criteria are required (e.g., performance, safety/security, robustness, fairness, privacy, compliance, independent external testing, evaluation, validation and verification (TEVV) or assurance); contextual caveats or limitations required in interpreting risk evaluation metric results; what approvals/sign-offs are needed; and what safeguards (e.g., mitigations, monitoring, or phased release) must be in place to move into deployment.

When deploying an AI model or system, the company operates a process for registering the target project in the AI Risk Management System, which provides a mechanism for continuously addressing AI risks throughout the lifecycle. The process includes: (1) checking applicability of AI regulatory laws in the market country and addressing obligations where applicable; (2) determining whether business promotion is acceptable from an AI ethics perspective; and (3) implementing risk reduction from the perspective of AI safety, including safety, fairness, and transparency.

Following deployment, how do you use incident reports — including those shared by other organizations — to identify emerging issues, update risk assessments, and trigger corrective actions.

Fujitsu continuously collects and reviews internal and external incident information and reflects it in AI risk management rules and systems as necessary. For newly identified risks that require action, the

company notifies relevant projects and operates a mechanism to prompt appropriate confirmation and corrective action.

3. Does your organization have mechanisms to receive and/or share reports of risks, incidents or vulnerabilities by/with third parties?

Yes, receiving only

If yes, please indicate which stakeholders reporting mechanisms may be accessible to, recognising specifics will depend on the nature of the report and risk level. For each, indicate whether access is on a standing basis (Always), activated based on context (Ad hoc/as relevant), or not currently accessible.

	Always	Ad hoc / as relevant	Not currently accessible
Relevant employees (security, legal/compliance, etc.)	☐		
Public authorities (e.g., public reporting offices for security incidents or data breaches)	☐		
Supervisory authorities (e.g., data protection authorities)	☐		
Industry partners or industry fora	☐		
Users (end users, enterprise customers, developers using APIs)	☐		
Business customers / deploying organizations (if different from “users”)	☐		
Supply-chain partners (cloud, data providers, etc.): <i>Please refer to the following Q.6.example of a free-text response.</i>	☐		
Academia / researchers	☐		
Independent auditors	☐		
Standards bodies / multi-stakeholder initiatives (OECD.AI, ISO/IEC, IEEE, etc.)	☐		
Specific civil society organizations	☐		
Media / general public (people or communities indirectly impacted)	☐		
Other			

Please describe how your organisation determines when and how reporting mechanisms are made accessible to different stakeholders (e.g. based on legal or regulatory requirements, internal policies, or risk level). Please also provide public URLs or details on how to submit reports. If none, write “Not applicable”

An online inquiry form for products and services is available at all times through the [Fujitsu Global Support](#) website. Inquiries from media, investors, and the general public are also handled through the reporting/contact desk.

To regulatory authorities, the company reports information on the incident summary, causes, scope of impact, corrective measures, and recurrence prevention measures at an appropriate level of detail based on regulatory requirements.

4. If your organization has programs to incentivize risk(s) disclosure, incidents and systemic vulnerabilities, please share what type of programs you use?

- Not applicable

5. Does your organization use or contribute to the development of international technical standards, best practices or tools for the identification, assessment, and evaluation of risks?

Contributes and uses

If contributes or uses, please select all that apply:

- ISO/IEC standards (e.g., ISO/IEC 27001, ISO/IEC 42001, ISO/IEC 23894, ISO/IEC 42005)
- NIST AI Risk Management Framework (AI RMF)
- Industry-led best practices or codes of conduct
- General Purpose AI Code of Practice
- Other international or regional standards (please specify): - Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment (European Commission related departments and expert group) - Machine Learning Quality Management Guideline, 3rd edition (AIST related departments and expert group) - Machine Learning Systems Security Guideline Version 2.00 (Machine Learning Systems Engineering research group, MLSE) - Guide to AI Safety Evaluation Perspectives, Version 1.10 (AISi) The above are standards and frameworks used through AI safety evaluation and AI regulatory compliance in AI risk management. Fujitsu also contributes to international standard-setting through activities such as the following: - Two members have served on the AI4People Scientific Committee since FY23 and contributed to the drafting and publication of white papers. - One member serves as editor in ISO/IEC JTC 1/SC 42 and contributes to the development of standards such as ISO 24030 (AI use cases) and ISO/IEC 42001 (AI management system). - OWASP ☐ At AppSec Global EU, an annual global event held in Europe, Fujitsu announced its differentiating technologies such as LLM vulnerability scanners/guardrail technology, Agentic AI Scanner technology, and policy alignment technology, contributing to enhancing our presence and expanding business opportunities.
(<https://owasp2025globalappseceui.sched.com/event/1xA4X/genai-security-insights-and-current-gaps-in-llm-vulnerability-scanners-and-guardrails>)
(<https://owaspglobalappseceui.sched.com/event/2JpRH/scanning-agentic-ai-systems-beyond-traditional-llm-red-teaming>)
(<https://owaspglobalappseceui.sched.com/event/2JRRk/from-safety-to-policy-enforcing-organizational-rules-in-llms-and-ai-agents>) ☐ Participated in and announced the development of "OWASP Securing Agentic Applications Guide 1.0," a guideline for enhancing the security of AI

agent applications. This contributed to enhancing our presence regarding our expertise in agent security reinforcement. (<https://genai.owasp.org/resource/securing-agentic-applications-guide-1-0/>)

☐ Participated in the online event "GenAI Application Security & Risk Virtual Conference" and presented technologies related to multilingual trust evaluation methods and benchmarks. (<https://cybersecuritysummit.com/summit/owasp-appsec/>)

☐ Continuously gathered information on the latest trends and attack techniques/methods in AI agent security through participation in various general meetings and technical conferences.

☐ As an OWASP Gold Corporate Supporter, we engaged in various external speaking engagements, contributing to enhancing our reputation for contributions to the open community. From FY26, we will be active as a Gold Sponsor in the OWASP GenAI Security Project, which focuses more on generative AI security. (<https://genai.owasp.org/>)

-MITRE Fujitsu participated as a Funding Research Participant in Secure AI, a research project at MITRE that investigates AI-specific security threats and conducts technical studies.

6. Does your organization collaborate with relevant stakeholders across sectors to assess and adopt risk mitigation measures to address risks that could have broad and potentially cascading socio-economic impacts? Please indicate which stakeholders you collaborate with, and whether collaboration is on a standing basis (Always-on), activated based on context (Ad hoc/as relevant), or there is no current collaboration.

	Always-on	Ad hoc / as relevant	No current collaboration
Relevant employees (security, legal/compliance, etc.)	☐		
Public authorities / regulators / AI safety institutes / other public sector bodies and inter-governmental coordination structures	☐		
Industry partners or industry fora	☐		
Users (end users, enterprise customers, developers using APIs)	☐		
Business customers / deploying organizations (if different from "users")	☐		
Supply-chain partners (cloud, data providers, etc.): <i>Please refer to the following example of a free-text response.</i>	☐		
Academia / researchers		☐	
Independent auditors		☐	
Standards bodies / multi-stakeholder initiatives (OECD.AI; ISO/IEC; IEEE; AIID etc.)	☐		
Independent civil society organizations (e.g., for evaluations, audits, or policy review)		☐	

	Always-on	Ad hoc / as relevant	No current collaboration
Media / general public (people or communities indirectly impacted)		☐	
Other(s) (please add any other relevant stakeholder groups):			
Not applicable			

Please describe examples of how your organization collaborates. Examples may include: participating in multi-stakeholder or industry initiatives; partnering with academia, civil society, or the public sector; establishing internal or joint governance structures with external experts or in partnership with other efforts to share knowledge and jointly address risks.

- Risk response based on the risk chain including upstream actors, downstream actors, and customers.
- Establishment of processes for business-owning divisions to confirm risk response status when procuring third-party products.
- Incorporation of external knowledge through reference to public guidelines, participation in private-sector industry groups such as AIGA, exchanges with other vendors, and the AI Ethics External Advisory Committee.
- Project review and discussion in cooperation with relevant divisions such as HR, Legal, and R&D, reflecting the latest risk countermeasures.

Any further comments and implementation documentation.

No answer provided

Section 2 – Risk Management and Information Security

7. Across the AI system lifecycle, what steps does your organization take to manage risks and vulnerabilities? Please indicate which specific practices, tools or metrics your organization uses at each lifecycle stage. You may refer to tools or metrics present in the OECD.AI Catalogue of Tools and Metrics or suggest additions. Respondents may complete only the lifecycle stages relevant to their organisation's role. Stages that do not apply may be left blank.

Plan and design

Collect and process data

Build, integrate and interpret model

Verify and validate

Deploy

Operate and monitor

Please provide a consolidated description of your organisation's approach to risk management across the AI system lifecycle, including how the practices identified above operate together and any cross-cutting governance, review or escalation processes. Please also indicate how often these practices are reviewed and what types of events trigger reviews and updates. Where relevant, please distinguish between different types of AI systems (e.g. generative or agentic AI).

Fujitsu's AI risk management rules and AI Risk Management System (AI-RMS) provide a company-wide framework for managing AI risks throughout the AI system lifecycle. The framework integrates AI regulatory compliance, AI ethics reviews, AI safety assessments, and project governance into a unified risk management process covering planning, design, development, validation, deployment, and operation. Projects identified as having elevated risks are subject to additional review, risk mitigation measures, and escalation processes involving relevant business, legal, technical, and governance stakeholders.

Relevant URL:

[AI Guidelines for Business Appendix Ver1.2 March 31, 2026, Column 8: Fujitsu Group's efforts on AI governance](#)

The following events are used as inputs to periodic internal and external environmental analysis and are reflected in AI risk management rules and systems as needed: new or amended laws and regulations, publication of public guidelines and international frameworks, industry and sector trends, AI technology trends, and incidents.

Please describe any measures used to ensure data quality and fairness across the data collection and curation stage (e.g., dataset audits, filtering training data to remove harmful content or personal information, fine-tuning on curated data, use of moderation tools, assessment of limitations to model generalisability).

Through AI safety evaluation, the division advises measures using [Fujitsu AI Ethics for Fairness](#), which facilitates verification and improvement of fairness in AI models.

[Fujitsu AI Ethics for Fairness](#)

8. What security measures are used in testing AI systems?

- Isolated and secure test environments
- Network isolation from external access
- Locally hosted testing environments
- Restrictions on outbound Internet or API access, including a list of permitted domains or APIs
- Isolation of staging environments (from production / operational environments)
- Controlled environments for red-teaming and adversarial testing
- Containment measures for prompt-injection and jailbreak attempts
- Controlled security vulnerability testing
- Rate limits and resource caps
- Human-in-the-loop controls for higher-risk test cases
- Access controls, including roles-based permissions

- Disabling, restricting or simulating high-impact capabilities during testing
- Audit logging and monitoring
- Use of “AI-safe” test-data, as appropriate
- Use of synthetic, anonymized, or dummy data, as appropriate

8.A. Please indicate, if relevant, the team or role responsible for overseeing secure testing and whether this responsibility is shared across multiple teams and/or functions.

For projects with high AI security risk, the company plans to prepare options to conduct red teaming in response to project needs.

9. How does your organization protect intellectual property, including copyright-protected content?

9.A. Safeguards that promote lawful data use and transparency in AI training:

- Maintain a copyright policy that addresses data sourcing, content provenance, safeguards, and organizational roles and responsibilities
- Publishing “model cards” or contextual information on training data and model limitations
- Apply provenance controls to guard against training data being sourced from websites or repositories that have been determined to engage in large-scale piracy
- Refrain from the unlawful circumvention of technological protection measures in data sourcing
- Maintain policies or practices to respect copyright holders’ preferences through, where appropriate/applicable, mechanisms like Robots.txt “do not train” tags or other machine-readable protocols developed through voluntary, industry-led standards processes
- Conduct data cleaning, qualification, and/or tokenization activities as part of a transformative data preparation process

9.B. Safeguards that reduce the risk of dissemination of outputs that may infringe copyright:

- Restricting outputs that raise ethical, human rights, or legal concerns—such as non-consensual intimate imagery, unauthorized deepfakes, or outputs that are substantially similar to copyrighted works.
- Filtering results using keyword or image-based restrictions to reduce risks of misuse

10. How do you protect privacy and prevent your systems from disclosing confidential or sensitive data across the AI lifecycle?

- Make information about data privacy, practices, and controls accessible to a diverse set of stakeholders
- Use security controls, including encryption at rest and in transit
- Provide organizational customers with controls to configure and implement privacy choices (e.g., data retention, logging, access, or processing settings)
- Apply Privacy-Enhancing Technologies (PETs) where applicable (e.g. differential privacy, federated learning, synthetic data generation, homomorphic encryption)

Describe key privacy measures and data-protection standards applied.

Privacy measures are implemented according to the nature of each individual project, mainly using the measures checked in Q10. The Fujitsu Group Information Security Measures Standard, based on external standards such as NIST CSF and the ISO/IEC 27000 series, has been established and data protection measures are implemented.

11. How does your organization assess cybersecurity risks and implement policies to enhance the cybersecurity of advanced AI systems across the AI system lifecycle? Respondents may complete the checkbox subsections in 11.A–11.D or describe their overall approach under 11.E.

11.A. Governance, policy and assessments

- Conduct regular cybersecurity risk and vulnerability assessments tailored to AI models or systems
- Perform threat modeling that includes AI-specific risks (e.g., data poisoning, prompt injection)
- Maintain or update security policies aligned with recognized standards (e.g., NIST CSF, ISO/IEC 27001, SOC 2, CIS Critical Security Controls), and review them regularly in response to emerging risks or regulatory guidance
- Maintain vulnerability management processes (scanning, channels for external reporting, patching, remediation)
- Implement incident response, disaster recovery, or business continuity plans, including escalation procedures for high-severity incidents
- Train staff on cybersecurity best practices and AI-specific threats, including how to recognize and respond to misuse or unintended outcomes of advanced AI systems
- Regularly review security measures and update them in response to emerging risks, incidents, or major releases

11.B. Technical and operational controls

- Apply multi-factor authentication (MFA) and role-based access controls for advanced AI systems
- Encrypt AI assets (datasets, models, configurations) in storage and in transit as appropriate
- Restrict physical access to servers, GPUs, or devices hosting AI models, systems, and/or applications
- Monitor system activity and log access to detect anomalies or misuse
- Conduct secure AI development and/or deployment practices
- Conduct security testing focused on AI-specific risks (e.g., red-teaming, adversarial testing, stress tests)

11.C. Protection of sensitive or high value AI assets

- Limit access to proprietary AI assets based on role or need-to-know
- Store model weights, datasets, and algorithms in secure, access-controlled environments
- Track and log access to sensitive AI resources
- Require NDAs or confidentiality agreements for staff and contractors with access

11.D. Insider risk protection detections

- Monitor for unauthorized internal access or suspicious activity
- Provide staff training on cybersecurity best practices and AI-specific threats

11.E. Please describe the main security measures your organization applies and indicate which are specific to particularly sensitive AI assets, where relevant.

AI safety evaluation: The company presents threat detection and countermeasures for evasion attacks, poisoning attacks, model extraction, and similar threats, and presents prompt-injection countermeasures for systems using generative AI.

12. What is your organization's vulnerability management process? Does your organization take actions to address identified risks and vulnerabilities, including in collaboration with other stakeholders?

- Internal vulnerability disclosure receipt and tracking processes
- Investigation of reported or identified risks and vulnerabilities
- Development of patches or other mitigations for confirmed risks or vulnerabilities
- Testing of patches or other mitigations
- Scanning for and mitigation of similar risks or vulnerabilities
- Collaborative vulnerability sharing (e.g., with partners, industry groups)
- Vulnerability disclosure program that is aligned with internationally recognized standards, e.g. ISO/IEC 30111 and 29147.
- Reporting to regulators / authorities when relevant
- Defined escalation criteria and follow-on actions for identified vulnerabilities (e.g., for external reporting or information sharing)

13. What post-deployment practices does your organization use to monitor and respond to vulnerabilities, incidents, emerging risks and misuse of advanced AI systems?

- Conduct post-deployment monitoring
- Perform root-cause analysis and apply lessons learned to improve future development and/or deployment practices

14. Do you develop or deploy agentic AI systems (systems that can autonomously plan/act or use tools)?⁴

Yes

14.A. If yes, please indicate which categories of controls your organization applies. Please select all.

- Human oversight or interruptibility mechanisms (approval requirements for sensitive actions, escalation paths, interruption interfaces)
- Other (specify): In AI ethics reviews, Fujitsu provides guidance and recommendations to route AI actions through appropriate approval processes and to involve human judgment

Please provide any further comments and implementation documentation.

Section 3 – Transparency Reporting on Advanced AI Systems

15. Does your organization publish public reports and/or technical documentation related to the capabilities, limitations, and domains of appropriate and inappropriate use of advanced AI systems? Are these clear and understandable by non-technical audiences?

Yes

15.A. If yes, which of the following information is included in your organization's publicly available documentation for any of your AI products?

- Capabilities, behaviours, and limitations, including domains of appropriate and inappropriate use or indicative thresholds associated with systemic risks.
- Technical documentation and instructions for use if relevant.
- Plain-language or non-technical summaries

Please provide links or references showing where this information is made public.

Relevant URLs:

[Fujitsu's AI Ethics and Governance](#)

[Fujitsu's AI Ethics Research](#)

Guidance is provided on appropriate content and frequency of information disclosure in light of the nature of each individual project.

15.B. If yes, how often are such reports usually updated?

Other (e.g., triggered by material changes, incidents, or new risks)

16. How does your organization share information with a diverse set of stakeholders (other organizations, governments, civil society and academia, etc.) regarding the outcome of evaluations of risks and impacts related to an advanced AI system?

- Regulatory reporting as relevant
- Other (specify): Internal risk evaluation results are visualized and shared through the AI Risk Management System.

Describe how evaluation results are shared and with which stakeholders (e.g. other organizations, governments, civil society and academia, etc). You may also indicate whether non-technical or plain-language summaries are provided for broader audiences.

Fujitsu reviews risks and impacts associated with the development, provision, and use of AI, including advanced AI systems, through internal AI ethics checks and consultation/review mechanisms, and continuously reviews its operations and AI ethics governance by incorporating perspectives from external experts. Specifically, the company has established an AI Ethics External Advisory Committee composed of experts in diverse fields, including not only AI but also law, biomedicine, ecology, SDGs,

and consumer administration, and receives objective evaluations and recommendations on AI ethics policies, internal governance, and implementation status.

17. Does your organization publish public privacy policies addressing the use of personal data, user prompts, and/or the outputs of advanced AI systems?

Yes

If yes, provide links to privacy policies or notices covering personal data, user inputs, and AI-generated outputs.

Relevant URL:

[Fujitsu's Global Privacy Policy](#)

18. How does your organization provide information about the sources of data used for the training of advanced AI systems, as appropriate, including information related to the sourcing of data annotation and enrichment?

- Provide information or tools to downstream actors on data quality
- Regulatory reporting as relevant

Please provide links to any publicly available information.

Information on the sources of training data for advanced AI systems incorporated into the company's business, as well as information related to the procurement of data annotation and data enrichment, is provided to the extent deemed appropriate, taking into account the nature of each individual AI project, the company's role, the types of data used, contracts with customers, partners and suppliers, and applicable legal and regulatory requirements. Such information is not published on the web in a common format.

19. Does your organization demonstrate transparency related to advanced AI systems through any other methods?

Yes

If yes, describe additional method(s), including where relevant transparency measures related to AI agents

Fujitsu demonstrates the Fujitsu Group AI Commitment and AI governance framework through white papers, websites, and similar materials. For individual AI systems, the company works to ensure transparency by providing the necessary information to the necessary parties within the necessary scope, depending on the company's role and degree of involvement, without relying on a single prescribed method.

20. Does your organization publish statistical insights derived from anonymised, aggregated usage data to advance understanding of AI adoption and its economic impacts in areas including occupations and tasks by region, sector and country? To this end, do you contribute to international cooperation efforts?

No

Please estimate the total staff time and annual expenses associated with conducting compliance activities and include any further comments and implementation documentation (optional).

No answer provided

Section 4 – Organizational Governance, Incident Management, and Transparency

21. How has AI risk management been embedded in your organization's governance framework?

- Board of Directors oversight
- Executive oversight
- Senior management oversight
- Governance council or committee
- Other (e.g., distributed / multi-institutional governance arrangement) (please specify): Business divisions are responsible for control and execution in the field, while the company-wide lead division, the AI Integrity Center, provides policies, rules, and mechanisms.

Please describe how AI risk management is integrated into governance structures, including: which roles or bodies have advisory versus decision-making authority; which roles or bodies have final accountability for AI risk-related decisions; and how often policies are reviewed or updated (e.g., following major incidents, regulatory changes, or material updates to systems).

Advisory authority is held by the AI Integrity Center, while decision-making authority and final accountability for AI risk-related decisions reside with the relevant business divisions.

Policies, rules, and governance processes are reviewed and updated in response to new or amended laws and regulations, publication of public guidelines and international frameworks, industry and sector trends, AI technology developments, serious incidents, and similar events.

22. Are relevant staff trained on your organization's governance policies and risk management practices?

Yes

If yes, describe training methods (e.g., onboarding modules, regular sessions, and confidential channels to report concerns).

Relevant staff are interpreted as Fujitsu Group employees. Staff receive training on the Fujitsu Group AI governance policy and AI risk management practices through the following methods:

- E-learning on AI risk management in AI business.
- Publication of explanatory videos accompanying the launch of AI risk management operations.
- Publication of use-case-specific demonstration videos for the AI Risk Management System.
- E-learning on AI ethics.

- AI ethics training for new employees.

23. Are steps taken to address reported incidents documented and maintained internally?

Yes

If yes, describe how incident records are kept and used for follow-up (e.g., incident logs, root-cause analysis reports, internal tracking systems). You may indicate whether and how you define escalation thresholds for when incidents must be surfaced to senior leadership or external parties.

Incident reports and records are created in the company-wide risk management and compliance management system, and external reporting is conducted when necessary based on escalation rules.

24. Does your organization share research and best practices on addressing or managing risk?

Yes

If yes, describe how research or good practices are shared (e.g., publications, conferences, collaborations with other sectors).

The control division regularly reports and shares practices for addressing and managing risks in AI safety evaluation with relevant divisions, including research and development divisions. This shares the latest information on risk response across the organization and raises the knowledge baseline. Based on the reported and shared information, information is also shared with business divisions. This supports the development of risk management structures that reflect business strategies and regional characteristics.

25. Does your organization use international technical standards or best practices for AI risk management and governance policies?

Yes

If yes, please indicate which standards or frameworks apply:

- ISO/IEC standards (e.g., ISO/IEC 27001, ISO/IEC 42001)
- NIST AI Risk Management Framework (AI RMF)
- OECD AI Principles
- Industry-led best practices or codes of conduct

If yes, describe which international technical standards or best practices your organization uses.

- Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment (European Commission related departments and expert group).
- Machine Learning Quality Management Guideline, 3rd edition (AIST related departments and expert group).
- Machine Learning Systems Security Guideline Version 2.00 (Machine Learning Systems Engineering research group, MLSE).
- Guide to AI Safety Evaluation Perspectives, Version 1.10 (AISi).

- AI Guidelines for Business (Ministry of Internal Affairs and Communications; Ministry of Economy, Trade and Industry).

Any further comments and implementation documentation.

No answer provided

Section 5 – Content Authentication & Provenance Mechanisms

26. What mechanisms, if any, does your organization put in place to allow end-users, where possible and appropriate, to know when they are interacting with an advanced AI system developed by your organization?

- User disclosure (e.g., text notice, verbal statement, or system label)
- Distinct branding or labeling for AI-generated content (e.g., “AI-generated” tag, metadata label)
- Documentation or FAQs explaining when and how AI is used in the system
- Disclosure to end-users when an AI system is acting autonomously or executing actions on their behalf (e.g., making purchases, sending communications, modifying files)
- Other (specify): From the perspective of cognitive security, Fujitsu laboratory is conducting research and development on technologies to support users’ recognition, judgment, and verification of AI-generated and AI-mediated information, including clear indication of AI use, presentation of output limitations and supporting information, warnings regarding disinformation, misinformation, deepfakes and similar issues, and support technologies for judging the reliability of information.

[<https://documents.research.global.fujitsu.com/trustable-internet/>]

Provide any additional details or context on how or when users are informed they are interacting with advanced AI systems (e.g. if such mechanisms are planned, in testing, or applied only to certain products).

Fujitsu designs and implements methods and timings for notifying users that they are interacting with an advanced AI system, depending on the individual AI project, the nature of the product or service, how the AI is used, and the level of impact on the user. As needed, we indicate the fact that AI is being used, the scope of processing handled by the AI, the limitations of its output, and the necessity of human verification or judgment, through interfaces, explanations at the start of use, terms of service, manuals, and explanatory materials.

27. Does your organization use content provenance detection, labeling or watermarking mechanisms that enable users to identify content generated by advanced AI systems?

Yes

27.A. If yes, how?

- Latent disclosure (e.g., hidden provenance information)
- Manifest disclosure (e.g., visible label)

- Other (specify): As a countermeasure against misinformation and disinformation, Fujitsu is researching technologies that analyze text, images, audio, video, and other media to determine whether information was created by generative AI or similar means and whether multiple pieces of information are consistent without contradiction, thereby helping users judge the reliability of information. Fujitsu's Trustable Internet technologies collect supporting evidence for information of uncertain authenticity and manage relationships as evidence graphs and endorsement graphs, taking an approach that verifies the correctness of information and the consistency of evidence. This can contribute to automated fact-checking, rapid decision-making support, and suppression of misinformation and disinformation dissemination for diverse media such as social media posts, news articles and video. [<https://en-documents.research.global.fujitsu.com/trustable-internet/>] Internationally, technical standards related to content provenance and authenticity, such as C2PA, are being discussed as important mechanisms for verifying content origin and authenticity in the era of AI-generated content. C2PA is an open standard that attaches cryptographically signed metadata to media files and makes it possible to verify content origin and editing history, and it can also support identification of AI-generated or AI-modified content. Fujitsu is pursuing research and development and external collaboration informed by international trends in misinformation and disinformation countermeasures, evidence management, media analysis, provenance, and authenticity. Fujitsu has also founded the international consortium Frontria to address misinformation and disinformation and emerging AI risks, bringing together cutting-edge technologies and expertise from around the world to create a digital society that combines trust and safety. [<https://global.fujitsu/en-global/pr/news/2025/12/02-01>]

27.B. If yes, does your organization use international technical standards or best practices when developing or implementing content provenance?

No

27.C. Does your organization collaborate on or invest in research to advance content authentication and provenance?

Yes

Describe how provenance mechanisms are implemented and which standards or tools are used, and any collaborations or investments to improve content authentication (e.g., C2PA credentials, joint R&D projects, open-source initiatives, standard-setting groups). You may also indicate whether these mechanisms are tested for robustness against removal, tampering, or circumvention.

Fujitsu recognizes that the spread of generative AI and synthetic content is increasing social risks such as misinformation and disinformation, deepfakes, impersonation, and information tampering. Based on this recognition, Fujitsu is promoting R&D and external collaborations to ensure [content trustworthiness](#). In particular, for [misinformation and disinformation countermeasures](#), Fujitsu is researching technologies that support users in judging the credibility of information by analyzing whether information across multiple media, including text, images, audio, and video, was created by generative AI or similar technologies, and whether multiple pieces of information are consistently aligned without contradictions.

Specifically, Fujitsu is advancing the construction of a [misinformation and disinformation countermeasures platform](#) through joint R&D with nine academic and industrial partners. This

platform integrates misinformation and disinformation detection, collection and integrated management of evidential information, analysis of consistency and contradictions between pieces of evidence, and social impact assessment. In this initiative, Fujitsu, as the prime contractor, is responsible for promoting and leading R&D, technology integration, and building the overall platform, aiming to socially implement research outcomes through the creation of use cases with public institutions and private companies.

Furthermore, Fujitsu's Trustable Internet related technologies adopt an approach to verify the correctness of information and the consistency of evidence by collecting evidential information for unverified information and managing their relationships as evidence graphs/endorsement graphs. This can contribute to automating fact-checking, supporting rapid decision-making, and suppressing the spread of misinformation and disinformation across diverse media such as social media posts, news articles, and videos.

In addition, internationally, technical standards related to content provenance and authenticity, such as C2PA, are being discussed as important mechanisms for checking the origin and verifying the authenticity of content in the era of AI-generated content. C2PA is an open standard that appends cryptographically signed metadata to media files, allowing the creation source and editing history of content to be verified, and also contributes to identifying AI-generated and AI-modified content. Fujitsu is advancing R&D and external collaborations based on international standardization trends regarding misinformation and disinformation countermeasures, evidence management, media analysis, and provenance/authenticity. Moreover, Fujitsu has established "[Frontria](#)", an international consortium dedicated to responding to misinformation and disinformation and new AI risks, bringing together cutting-edge technologies and knowledge from around the world to create a digital society that combines trust and safety.

Any further comments and implementation documentation.

No answer provided

Section 6 – Research & Investment to Advance AI Safety & Mitigate Societal Risks

28. Does your organization advance research and investment in the following areas?

- Security
- Safety
- Data quality and representativeness
- Information integrity
- System performance and outcomes (e.g. accuracy, consistency)
- Explainability or interpretability
- Transparency
- Robustness or reliability

If yes, please describe key research or investment initiatives in the areas selected above (e.g., dedicated safety teams, funding of research programs, contributions to open-source safety tools,

publishing research, partnerships and collaborative research initiatives). You may also indicate whether related outputs are publicly shared, shared with partners or within consortia or not publicly disclosed.

Fujitsu is advancing research, development, implementation, and external collaboration on both AI ethics governance and AI trust technology to realize the social implementation of safe, secure, and reliable AI.

In terms of AI ethics governance, Fujitsu formulated and published the "[Fujitsu Group AI Commitment](#)" in 2019, outlining its policy to prioritize value creation through AI, human-centricity, sustainability, respect for human decision-making, transparency, and accountability. They have also established an external AI Ethics Committee, incorporating insights from outside experts in diverse fields such as law, biomedical science, SDGs, and consumer administration, not just AI. Furthermore, the AI Ethics Office leads the Fujitsu Group's AI ethics governance strategy, internal practices, and external dissemination, conducting comprehensive ethical risk reviews for all AI businesses and providing AI ethics education for all employees.

Regarding research and development, Fujitsu is promoting R&D in [AI ethics technology](#) based on four pillars: compliance with laws and guidelines, Ethics-by-Design and AI governance, addressing biases inherent in generative AI, and tackling social issues. Specifically, they offer and make public tools such as [AI Ethics Impact Assessment](#) to identify and evaluate ethical risks before designing AI services, [Fujitsu AI Ethics for Fairness](#) to verify and improve the fairness of AI training data and decisions (including [intersectional fairness](#)), [Fujitsu LLM Bias Diagnosis](#) to identify bias in LLMs, and AI Risk Management Tools to manage ethical, quality, and security risks inherent in each phase of the AI lifecycle.

In the security aspect of generative AI and advanced AI systems, Fujitsu has developed a "[LLM Vulnerability Scanner & Guardrails](#)" that comprehensively evaluates LLM vulnerabilities and automatically applies optimal defense technologies to generative AI. This technology aims to enable even non-experts to safely operate generative AI, addressing the challenges of increasing attacks on generative AI, diversifying attack methods, and a shortage of experts.

For further technical details:

[LLM Vulnerability Scanner & Guardrails \(Fujitsu Research\)](#)

Additionally, in [multi-AI agent security technology](#), Fujitsu is advancing R&D to support proactive measures against new threats to IT systems and generative AI by linking attacking AI agents, defending AI agents, and testing AI agents.

To address misinformation and disinformation and social risks, in response to the societal problem of misinformation and disinformation circulation caused by generative AI and synthetic content, Fujitsu has embarked on technology development for misinformation and disinformation analysis as part of a NEDO public solicitation project under the "[Economic Security Important Technology Development Program](#)" in collaboration with the Cabinet Office and the Ministry of Economy, Trade and Industry. This initiative is advancing R&D for a misinformation and disinformation countermeasure system that combines misinformation and disinformation detection for each medium, evidence and endorsement

management, comprehensive truthfulness judgment support, and misinformation and disinformation impact assessment.

Furthermore, Fujitsu is working on building a misinformation and disinformation countermeasure platform in an "[All-Japan framework involving nine industry-academia partners](#)", which integrates processes from misinformation and disinformation detection to evidence gathering, analysis, and evaluation. In addition, Fujitsu has founded "[Frontria](#)", an international consortium to address misinformation and disinformation and new AI risks, bringing together cutting-edge technologies and knowledge from around the world to create a digital society that combines trust and safety.

Through these efforts, [Fujitsu continuously promotes R&D](#) to enhance the safety, security, fairness, explainability, transparency, robustness, and trustworthiness of AI, linking research outcomes to AI services, AI platforms, and social implementation.

Related reference:

- [Fujitsu Kozuchi AI Platform](#)

[Fujitsu AI Ethics Impact Assessment](#)

[Intersectional Fairness](#)

[Fujitsu LLM Bias Diagnosis](#)

[LLM Vulnerability scanner and guardrails](#)

[Fujitsu AI Ethics for Fairness](#)

29. Does your organization pursue research or investment related to socio-economic or environmental considerations in the development and deployment of AI?

- Socio-economic considerations (e.g., AI adoption and productivity, workforce transitions, skills development, access to AI tools)
- Environmental considerations (e.g., energy use, computational efficiency, resource use)

If yes, please provide examples of relevant research or investment initiatives for the areas selected above.

Fujitsu regards both socio-economic and environmental risks from AI as important issues and is pursuing research, development, and investment to address each. For socio-economic risks, Fujitsu recognizes unfair AI decisions, bias, privacy infringement, misinformation and disinformation, deepfakes, and misuse of generative AI as risks that can undermine trust among individuals, companies, and society. Fujitsu therefore advances research and development in AI ethics governance, AI fairness, AI risk assessment, AI security, and misinformation and disinformation countermeasures. Fujitsu identifies compliance with laws and guidelines, Ethics-by-Design and AI governance, addressing bias in generative AI, and tackling social issues as pillars of its AI ethics technologies.

As concrete measures to reduce socio-economic risks, Fujitsu develops and provides Fujitsu AI Ethics for Fairness to verify and improve fairness in AI training data and decisions, AI ethics impact assessment to identify and evaluate ethical risks before designing AI services, and AI risk management tools. To address the societal challenge posed by the spread of misinformation and disinformation associated with generative AI and synthetic content, Fujitsu is developing a countermeasure platform that integrates misinformation and disinformation detection, evidence collection, analysis, and evaluation, aiming to contribute to a trustworthy digital society.

For environmental risks, Fujitsu recognizes that the expansion of AI technologies, including generative

AI, is increasing data center power consumption and affecting the global environment. Fujitsu is therefore conducting R&D on [power-saving AI computing infrastructure](#). Specifically, Fujitsu is researching and developing AI Computing Broker to improve GPU operating efficiency, [next-generation CPU FUJITSU-MONAKA](#) combining low power consumption and high performance, and water-cooling technologies to address increasing data center heat generation.

Fujitsu has also developed [generative AI reconstruction technology](#) to reduce the size and power consumption of large language models and strengthen its large language model Takane. This technology reduces memory consumption by up to 94% through 1-bit quantization, maintains 89% of accuracy compared with before quantization, and achieves three-times faster processing. This enables large generative AI models that previously required four high-end GPUs to run quickly on one low-end GPU, thereby contributing to power savings in AI operations.

In this way, Fujitsu works to minimize risks while expanding the benefits of AI to society by advancing R&D in AI ethics, fairness, security, and misinformation and disinformation countermeasures for socio-economic risks, and by promoting power-saving, lightweight, and high-efficiency AI computing infrastructure for environmental risks.

Any further comments and implementation documentation.

No answer provided

Section 7 – Advancing Human and Global Interests

30. Does your organization support any digital literacy, education or training initiatives to improve user awareness and/or help people understand the nature, capabilities and limitations of advanced AI systems?

Yes

If yes, describe any digital literacy or training initiatives your organization supports (e.g., user education, public AI-education programs, school or university partnerships, community workshops). You may also indicate whether these efforts include or prioritise underserved or underrepresented communities.

Fujitsu is autonomously building organizational governance as a company that develops and provides AI. Because collaboration with stakeholders is essential to ensuring trustworthiness in AI business, Fujitsu contributes to improving AI literacy across society, including among users, by returning part of these outcomes to user companies and society through education on AI technologies and AI-related risks.

Fujitsu also works to improve AI literacy among a wide range of social groups, including students. For example, in accordance with government guidance such as the Human-Centric AI Social Principles and the AI Guidelines for Business, Fujitsu has released the Fujitsu AI Ethics Karuta, free teaching materials that make AI ethics and risks easy for beginners to learn. Using these materials, Fujitsu holds

workshops and special classes, and discloses examples of countermeasures together with explanations of AI-specific risks such as hallucination and bias for internal and external audiences.

31. Does your organization pursue AI projects that aim to deliver societal or public benefits, including through partnerships with civil society and community groups?

Yes

If yes, please provide examples, including any collaborations with civil society, community groups, or efforts to address broader societal challenges.

Fujitsu has launched the Fujitsu Uvance business model, which aims to create markets based on social issues and generate high value-added businesses. Fujitsu considers partnerships with civil society and community groups necessary to bring together diverse perspectives and knowledge essential to solving global-scale issues and to bridge the literacy gap in technical knowledge and data between AI providers and civil society, which is a challenge for ensuring trustworthiness in AI business.

Related references:

[SDG-related Activities in Fujitsu](#)

[Fujitsu's Sustainability Activities: United Nations Global Compact](#)

As an important element of these partnerships, Fujitsu promotes industry-academia collaboration with educational and research institutions for AI ethics research and development and human-resource development. Specifically, Fujitsu establishes joint courses at educational institutions, gives lectures, and accepts interns. The knowledge gained through these exchanges is applied throughout the value chain to the development of AI solutions, including research and development of AI technologies themselves, development of technologies that improve trustworthiness, and development of measures to improve user literacy.

Fujitsu has also established the international consortium [Frontria](#) to address complex contemporary issues such as the spread of misinformation and disinformation by generative AI, vulnerabilities in AI systems, and responses to laws and regulations. More than 84 international organizations cooperate to develop technologies and create new use cases according to the needs of each industry, focusing on misinformation and disinformation countermeasures, ensuring AI trustworthiness, and strengthening AI security, with the aim of realizing a digital society that combines trustworthiness and safety. Through international cooperation, Frontria aims to improve the reliability of information and provide inclusive and sustainable solutions.

Any further comments and implementation documentation.

No answer provided