

OECD AI Transparency Report

Organization: Rakuten Group, Inc. (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)

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): Rakuten uses an internal AI risk taxonomy and aligns its risk identification and evaluation practices with external frameworks and requirements, including the NIST AI Risk Management Framework, the OECD AI Principles, Japan's AI Guidelines for Business, and applicable laws and regulations.

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.

Rakuten does not currently use a single documented "severe risk" threshold for advanced AI systems. Instead, AI-related risks are evaluated through established governance processes, starting with a structured internal consultation with AI governance and safety specialists, followed by a risk assessment, an AI Guardrail review, and other required compliance reviews as applicable. If risks

cannot be sufficiently addressed through technical, operational, or organizational safeguards, the system may be required to implement additional controls, undergo redesign, delay deployment, or not proceed to release until the identified concerns are addressed.

Rakuten's AI safety principles identify categories of outputs and use that are considered unreasonable or intolerable in any context, such as providing instructions or material support for chemical, biological, radiological, or nuclear (CBRN) weapons, generating child sexual abuse material, or facilitating attacks on critical infrastructure or third-party systems; these are subject to unconditional refusal. Other categories, such as professional advice in regulated domains (e.g., medical, legal, or financial) or adult content, are addressed on a graduated basis depending on context, such as the assessed severity of potential harm and the specificity of the request, ranging from qualified responses with appropriate disclaimers to refusal.

Rakuten applies a risk-based governance approach to agentic AI. Relevant use cases are assessed through a structured risk assessment and AI Guardrail processes, with controls including restrictions on authorized actions and access, human oversight or interruptibility, and logging, as appropriate.

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?

- Internal red teaming³
- Adversarial testing to detect technical vulnerabilities
- Publicly available benchmarks and measurement tools
- Internal benchmarks, standard metrics (e.g., frequency, severity scores)
- Qualitative assessments (e.g., expert judgment, narrative scoring)
- Impact assessments such as personal data protection assessment and fundamental rights impact assessments
- Vulnerability reports
- 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): Assessed through a structured risk assessment process, including definition of authorized action scope and required extent of human review for agentic AI 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.

Deployment decisions are based on a multi-disciplinary review process coordinated through a structured internal consultation with AI governance and safety specialists. Depending on the use case, reviews may include AI Risk Assessment, privacy, data control, legal/IP, vendor management, procurement, software review, and security architecture review. AI systems and AI services subject to AI Guardrail requirements must undergo mandatory review, risk assessment, registration, and approval by designated representatives before release.

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.

Post-deployment incident and vulnerability information is used by relevant teams for investigation, root-cause analysis, updates to risk assessments, corrective and preventive actions, and broader remediation where appropriate. Corrective action is triggered based on defined criteria such as the severity, scope, and potential impact of the reported issue; issues meeting these criteria are escalated to senior management, the Group CISO, and other relevant governance bodies for follow-up.

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

Yes, both receiving and sharing

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>Cloud infrastructure providers and AI model/component providers.</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)		☐	

	Always	Ad hoc / as relevant	Not currently accessible
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”

Reports concerning vulnerabilities, incidents, and related issues are received through public or designated reporting channels and escalated to the appropriate specialist teams based on their nature and severity. Public reports may be submitted through Rakuten's Vulnerability Disclosure Program at <https://bugcrowd.com/engagements/rakuten-vdp-pro>

How does your organization share relevant information as appropriate?

- Information-sharing partnerships (with confidentiality agreements as appropriate)
- Reporting to regulatory authorities as relevant

Please describe what type of information is shared.

Shared information may include vulnerabilities, security and privacy incidents, impacts, mitigation measures, and corrective actions. The confidentiality and necessity of the information are considered before sharing.

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

- Vulnerability Disclosure Programs (VDPs) with safe harbor / liability protection

Please provide public URLs and/or details on your incentive program(s). If none, write “Not applicable.”

<https://bugcrowd.com/engagements/rakuten-vdp-pro>

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?

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)

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.)			☐
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)			☐
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.

Rakuten collaborates on the assessment and adoption of risk mitigation measures by participating in industry forums such as the AI Governance Association and engaging with public authorities and policy stakeholders through frameworks such as the HAIP.

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).

Rakuten applies a lifecycle-based AI risk management process. During planning and design, AI use cases undergo a structured internal consultation with AI governance and safety specialists, and, where required, an AI Risk Assessment is conducted and recorded in a centralized risk register. Applicable privacy, data-control, legal/IP, vendor-management, and security reviews are conducted based on the use case. Before release, systems subject to the AI Guardrail requirements are registered, required AI Guardrail controls are implemented, and their implementation status and supporting evidence are recorded or submitted through the AI Risk Assessment process, as applicable. Following deployment, incidents, vulnerabilities, or material changes may trigger investigation, reassessment, remediation, escalation, and updates to controls. Agentic AI use cases additionally document authorized actions, access to systems and data, and the extent of human review. The framework is updated in response to incidents, material changes in use cases or technologies, and legal or regulatory developments.

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).

Rakuten applies safeguards during data collection and curation to support data quality, fairness, privacy, and lawful use. Depending on the model and dataset, these measures may include reviewing data sources and usage rights, filtering or curating training data to reduce personal data and harmful content, minimizing sensitive or regulated data, and assessing data quality, representativeness, and potential bias. Relevant model-development activities are also subject to applicable copyright and

legal/IP review requirements, including consideration of data-sourcing legality, content provenance, and the avoidance of unlawful circumvention of technical protection measures.

For relevant Rakuten-developed models, Rakuten also uses curated data in additional fine-tuning rounds informed by a Constitutional AI approach, in which model responses are critiqued against a documented set of safety principles and the resulting critiques are used as training signals. Rakuten assesses limitations to model generalizability by evaluating the trade-off between improved safety behavior and general capability, sometimes referred to as the "alignment tax," across general-purpose benchmarks.

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

- Isolated and secure test environments
- Restrictions on outbound Internet or API access, including a list of permitted domains or APIs
- Controlled environments for red-teaming and adversarial testing
- Access controls, including roles-based permissions
- Audit logging and monitoring
- 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.

Oversight of secure testing is with a dedicated cybersecurity function under the Group CISO, working in coordination with AI development, product security, and quality assurance.

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
- Refrain from the unlawful circumvention of technological protection measures in data sourcing
- Conduct data cleaning, qualification, and/or tokenization activities as part of a transformative data preparation process
- Engage in AI training that comprises a non-consumptive use of any underlying data
- Consult and/or retain an intellectual property attorney to protect, enforce, and comply with, intellectual property rights.

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.

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)

Describe key privacy measures and data-protection standards applied.

Rakuten conducts privacy and legal review of the handling of personal data, user inputs, and AI outputs, and makes information about its privacy practices and controls accessible to a diverse set of stakeholders. Rakuten applies security controls including encryption at rest and in transit, and, for applicable products, provides organizational customers with controls to configure privacy choices such as data retention, logging, and access settings.

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
- Integrate cybersecurity AI risks into broader cybersecurity risk management activities
- 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
- 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
- Implement insider-risk escalation or response procedures

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

In addition to Rakuten's standard AI security controls—including AI-specific risk assessment and threat modeling, security policies aligned with recognized standards, incident response planning, staff training, multi-factor authentication, encryption, activity monitoring and logging, secure development practices, and AI-specific security testing—Rakuten protects high-value AI assets through role-based and need-to-know access controls, least-privilege principles, authentication controls, secure access-controlled environments, logging, monitoring and alerting, and safeguards designed to prevent data leakage and unauthorized access. AI systems subject to AI Guardrail requirements undergo registration, risk assessment, and implementation of mandatory security controls before deployment. The available evidence does not establish general use of hardware security modules or trusted execution environments for these assets.

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
- Recognition of external finders of confirmed risks or vulnerabilities
- 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.

- Constraints on the agent's action space (e.g., tool-use permissions, environment sandboxing, allowlists/blocklists of resources, scope limits per task)
- Human oversight or interruptibility mechanisms (approval requirements for sensitive actions, escalation paths, interruption interfaces)
- Logging of agent actions

Please provide any further comments and implementation documentation.

Rakuten's AI Risk Assessment process, completed prior to deployment, requires developers of agentic AI systems to define the scope of actions the system is authorized to perform on IT systems and data (e.g., system configuration changes, data manipulation, and performing actions on behalf of customers such as placing orders or bookings), and to specify the extent of human review applied to AI-driven actions and outputs. Findings from this process have led to recommendations such as adding a confirmation step before higher-risk autonomous actions (e.g., checkout or payment execution) prior to external deployment. Input/output logging requirements applicable to AI systems generally also extend to agentic AI systems.

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.

<https://huggingface.co/Rakuten/RakutenAI-3.0>

https://global.rakuten.com/corp/news/press/2026/0317_01.html

<https://ai.rakuten.com/>

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

At or before major product releases

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?

- Not applicable

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.

Not applicable.

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.

<https://privacy.rakuten.co.jp/english/> (Rakuten Group Privacy Policy)

<https://privacy.rakuten.co.jp/> (Rakuten Privacy Center, Japanese).

These policies address the collection, use, and processing of personal data, including personal data processed in connection with AI-enabled services, but do not separately or explicitly reference user prompts or AI-generated outputs as distinct categories.

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?

- Not applicable

Please provide links to any publicly available information.

Not applicable.

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

No

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?

- Executive oversight
- Senior management oversight
- Governance council or committee

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).

AI risk management is integrated into Rakuten Group's governance framework as part of its overall information security and AI governance structure. Advisory and assessment functions related to AI risks are provided by centralized governance, security, and risk management teams operating under the oversight of the Group Chief Information Security Officer (Group CISO); outcomes of these assessments, including proposals and decisions, are reported to the Group AI & Data Committee, which serves as the organization's governance council for AI-related matters. Decision-making authority for AI risk-related matters rests with senior management within each business and technical unit, in accordance with their respective responsibility and risk ownership. Final accountability for the establishment and operation of the group-wide AI risk management and assurance framework is maintained through a shared governance structure, in which the Group CISO is responsible for AI security assurance, while overall AI governance is led by the Group CAIDO. Policies and governance arrangements related to AI risk management are reviewed periodically and are updated as necessary in response to regulatory developments or significant changes in AI use cases, technologies, or associated risks.

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).

AI risk awareness and responsible AI use are promoted primarily through internal training programs and the communication of enterprise-wide policies and guidelines. Training and awareness activities are provided to relevant employees through a combination of e-learning courses, workshops, and internal seminars that cover topics such as generative AI basics, information security, privacy, and appropriate AI use. In addition, group-wide policies, operational standards, and guidance on AI use and risk management are disseminated through internal portals and awareness initiatives. These activities are designed to raise understanding of AI-related risks and to support the safe and responsible use of AI across the organization.

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.

Reported incidents and the steps taken to address them are documented and maintained internally in accordance with Rakuten Group's incident management framework. Records are captured through internal reporting and tracking mechanisms, including logs and related documentation that describe the nature of the matter, the response actions taken, and the remediation or preventive measures implemented. Where appropriate, root-cause analysis is conducted, and findings are used to inform follow-up actions and improvements to controls or governance practices. Escalation thresholds are defined based on the severity, scope, and potential impact of the matter. Items that meet predefined criteria are escalated to senior management, the Group CISO, and other relevant governance bodies, and may be further surfaced to external parties or authorities in accordance with applicable laws, regulations, and internal policies.

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).

Rakuten Group shares knowledge, lessons learned, and best practices related to AI risk management primarily through participation in relevant industry, regulatory, and multi-stakeholder forums, and through collaboration with trusted partners where appropriate. Such sharing may occur in the context of vulnerability handling, governance discussions, or risk management exchanges, and typically focuses on practices for identifying, assessing, and mitigating AI-related risks. Information is shared in a manner that is proportionate and aligned with confidentiality, legal obligations, and the nature of the stakeholders involved, and does not involve the disclosure of sensitive or proprietary information.

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

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

Rakuten Group uses internationally recognized standards and frameworks to inform its AI risk management and governance policies. In particular, AI risk identification, assessment, and management practices are aligned with ISO/IEC 42001 for AI management systems and the NIST AI Risk Management Framework (AI RMF). These frameworks serve as reference points for designing and maintaining internal AI governance processes, risk assessments, and operational controls, helping ensure that AI-related risks are addressed in a structured, risk-based, and proportionate manner across the organization.

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?

- Documentation or FAQs explaining when and how AI is used in the system
- Access to provenance information (e.g., model version, content authenticity signals)

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).

Rakuten Group provides public documentation explaining how its advanced AI systems are developed and used, including model cards and related technical materials (e.g., for RakutenAI-3.0) that describe the model's capabilities, limitations, and intended use. These materials also disclose provenance-related information such as model version and base model lineage. At this time, dedicated persistent visual or audio indicators, distinct AI-generated content labeling, and disclosure mechanisms specific to autonomous AI actions have not yet been implemented across Rakuten Group services; such mechanisms may be introduced as relevant products and use cases evolve.

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

No

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
- 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.

Rakuten advances research and investment across multiple areas relevant to the safety and trustworthiness of its AI systems. This includes safety and security testing (e.g., red-teaming and adversarial evaluation of Rakuten LLMs), alignment research grounded in Rakuten's AI safety principles (e.g., Constitutional AI-based fine-tuning), evaluation of data quality and fairness in training data, publication of technical documentation and model cards to support transparency, and evaluation of model robustness and generalizability. Related outputs are shared through open-weight model releases, public model cards and technical documentation, and, where appropriate, academic or industry publications; some outputs remain internal where they concern proprietary methods or unreleased capabilities.

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)

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

Rakuten invests in AI adoption and skills development through initiatives including Rakuten AI University for marketplace merchants, internal AI training programmes, and the release of Japanese-language open-weight models. These initiatives support productivity, workforce skills development, and broader access to AI tools.

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.

Rakuten supports digital literacy and user understanding of AI through several initiatives. Rakuten AI University provides training programs and resources for merchants on Rakuten's online marketplace (e.g., Rakuten Ichiba) to help them understand and effectively utilize AI technologies, including their capabilities and limitations. Rakuten also provides workshops, webinars, and seminars covering AI and digital transformation topics for a broader audience. These initiatives are primarily aimed at merchants and business users rather than being specifically targeted at underserved or underrepresented communities.

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.

Rakuten pursues AI projects intended to deliver societal and public benefits. Rakuten has released Japanese-language open-weight AI models (e.g., Rakuten AI 7B and Rakuten AI 2.0) to the open-source community, enabling public and private organizations, researchers, and community groups to access and build on these technologies free of charge, supporting broader innovation and digital-literacy efforts in Japan's AI ecosystem. Rakuten also supports AI-related education through collaboration with universities and by using AI to improve logistics and resource efficiency (e.g., route optimization and inventory management), contributing to broader sustainability and productivity goals.

Any further comments and implementation documentation.

No answer provided