OECD AI Transparency Report

Organization: SoftBank Corp. (JP) **Reporting Period:** Q2 2025 **Published:** April 16, 2025

Section 1 – Risk identification and evaluation

a. How does your organization define and/or classify different types of risks related to AI, such as unreasonable risks?

We adopt a risk-based approach to AI governance. We define risk levels in four categories: Prohibited, High, Mid, and Low.
Our risk definitions take into account the social and business impacts when risks manifest. Additionally, they are formulated to allow employees to easily and clearly classify them.
Specifically, we evaluate risks based on two axes: firstly, categories delineated by the EU AIAct and high-risk sectors in Japan (government, finance, energy, transportation, traffic, telecommunications, broadcasting, and healthcare), and secondly, our own additional criteria, including the scale of users and revenue.
*High-risk sectors in Japan are identified based on the 'AI Guidelines for Business' and other codes of conduct published by the Japanese government, which ensure compliance and promote the use of AI, as outlined in the draft document: https://www8.cao.go.jp/cstp/ai/ai_senryaku/6kai/13rikoukakuho.pdf

b. What practices does your organization use to identify and evaluate risks such as vulnerabilities, incidents, emerging risks and misuse, throughout the AI lifecycle?

We perform distinct risk assessment and governance tasks for risk levels classified as 1-a, as follows:

・During the planning stage, we discuss and identify potential risks using a checklist.
・We evaluate the severity and impact of risks using a risk assessment framework.
・Before release, we conduct vulnerability assessments, submit checklists, and hold review meetings involving departments related to data, legal, intellectual property, and AI ethics/governance.
・For high-risk cases, in addition to the checklist, we conduct review meetings including executives.
・Regular reviews and monitoring (especially for high-risk levels, review meetings are held every few months).

c. Describe how your organization conducts testing (e.g., red-teaming) to evaluate the model's/system's fitness for moving beyond the development stage?

When we commercially provide services, we implement a rule that mandates undergoing quality assurance processes. If the services do not meet the prescribed standards, we decide not to release them. When conducting AI red-teaming, we establish a lifecycle that incorporates the evaluation results of the AI red-teaming into post-learning processes.

d. Does your organization use incident reports, including reports shared by other organizations, to help identify risks?

Yes

e. Are quantitative and/or qualitative risk evaluation metrics used and if yes, with what caveats? Does your organization make vulnerability and incident reporting mechanisms accessible to a diverse set of stakeholders? Does your organization have incentive programs for the responsible disclosure of risks, incidents and vulnerabilities?

We evaluate risks using two axes: the EU AI Act and high-risk domestic sectors (such as government, finance, energy, transportation, telecommunications, broadcasting, and healthcare), as well as our own additional criteria, and factors like user numbers and revenue scales. Important points in our evaluations include considering future projections rather than current figures, presenting guidelines to field personnel for unified risk assessment, and having certain risks re-evaluated by a central organization. High-risk items require confirmation from executive-level personnel and, if necessary, expert opinions from lawyers.

Regarding vulnerability and incident reporting mechanisms, we communicate company's incident impact assessment criteria, reporting rules, and flows to internal stakeholders and both internal and external executives, ensuring accessibility.

Additionally, we do not have an incentive program in place.

f. Is external independent expertise leveraged for the identification, assessment, and evaluation of risks and if yes, how? Does your organization have mechanisms to receive reports of risks, incidents or vulnerabilities by third parties?

In April 2024, we established an AI Ethics Committee consisting of internal members and external experts to create a mechanism for receiving insights and information on risks, incidents, and vulnerabilities from third parties. Although most judgments can be made internally, we consult specialists, such as lawyers, when specialized expertise is required.

For risk management evaluation, we conduct verifications in line with the ISO31000 (JISQ31000) risk management system and its processes. For AI risk checks, we collaborate with external organizations to develop checklists and use them for risk assessments. Additionally, we establish mechanisms to receive reports on risks, incidents, or vulnerabilities from third parties by collaborating with threat intelligence vendors and industry voluntary organizations. When external parties discover vulnerabilities, the SoftBank CSIRT accepts the information, and we handle the response with the relevant departments and stakeholders.

[Reference: https://www.softbank.jp/en/corp/aboutus/governance/security/cooperation/]

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

We regularly output efforts towards the development of international technical standards and best practices to relevant ministries and organizations, and these efforts are published as AI Guidelines for Business in Japan, a unified guideline for AI governance. In generative AI development, we ensure comprehensive risk considerations in line with research by Wang et al. (EACL Findings2023) and Röttger et al. (NAACL2024), forming a foundation for safety evaluation development.
[References: Wang et al. (EACL Findings2023): https://arxiv.org/abs/2308.13387, Röttger et al. (NAACL2024): https://arxiv.org/abs/2308.01263]

h. How does your organization collaborate with relevant stakeholders across sectors to assess and adopt risk mitigation measures to address risks, in particular systemic risks?

We convene review meetings involving departments related to security, data, legal, intellectual property, and AI ethics/governance for all AI development and service offerings. Additionally, in April 2024, we established an AI Ethics Committee that includes external experts to integrate diverse perspectives, ensuring objectivity in our internal rules and the handling of sampled cases.

The Risk Management Committee consists of the President, Vice President, auditors, and heads of relevant departments from various sectors. This committee determines the significance of risks, assigns responsibility for managing them, and issues directives for risk mitigation measures.

Any further comments and for implementation documentation

No answer provided

Section 2 – Risk management and information security

a. What steps does your organization take to address risks and vulnerabilities across the AI lifecycle?

In our AI development, we provide development guidelines and checklists to allow development teams to autonomously address risks and vulnerabilities. For projects involving AI red-teaming, we conduct adversarial manual testing and automated benchmark evaluations from multiple perspectives, including harmfulness, bias, and privacy infringement. Our safety team analyzes this data and conducts training to ensure safety.

b. How do testing measures inform actions to address identified risks?

In developing generative AI, we prepare fixed adversarial questions for predefined risk categories, evaluating and analyzing the response answers to enhance safety in problematic areas in further AI red-teaming and training data creation.

c. When does testing take place in secure environments, if at all, and if it does, how?

In accordance with our internal standards, we construct a secure test environment. Additionally, we have a rule to primarily use dummy data during testing.

d. How does your organization promote data quality and mitigate risks of harmful bias, including in training and data collection processes?

We collect data from diverse sources and strive to reduce bias in the data. Additionally, if bias is detected during testing, we tune the model as necessary.

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

The SoftBank Code of Conduct, which all executives, employees and group companies are required to abide by, declares the following: “We recognize the importance of intellectual property rights, and we will respect the intellectual property rights of others while promoting the appropriate protection and utilization of our own intellectual property rights.” Our intellectual property strategy lays out the core principles for how we intend to enhance corporate value and contribute to the industrial development of society as a whole by striving to

create, protect and utilize intellectual property while at the same time respecting the intellectual property rights of others
[Reference:
https://www.softbank.jp/en/corp/set/data/sustainability/documents/reports/pdf/sbkk_sustainability_report_2024_en.pdf
]

f. How does your organization protect privacy? How does your organization guard against systems divulging confidential or sensitive data?

We set up systems and safety measures according to our Privacy Policy and Information Security Policy to handle privacy and confidential data appropriately, which is disclosed externally on our website.
[References:
Privacy Center: https://www.softbank.jp/en/privacy/,
Information Security Policy: https://www.softbank.jp/en/corp/security/,
Personal Data Protection Policy: https://www.softbank.jp/en/privacy/personaldata/data-protection/data-security/,
SoftBank Data Handling Policy: https://www.softbank.jp/en/privacy/personaldata/utilization/data-anonymization/]

**g. How does your organization implement AI-specific information security practices pertaining to operational and cyber/physical security?
i. How does your organization assess cybersecurity risks and implement policies to enhance the cybersecurity of advanced AI systems?
ii. How does your organization protect against security risks the most valuable IP and trade secrets, for example by limiting access to proprietary and unreleased model weights? What measures are in place to ensure the storage of and work with model weights, algorithms, servers, datasets, or other relevant elements are managed in an appropriately secure environment, with limited access controls in place?
iii. 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? iv. How often are security measures reviewed?
v. Does your organization have an insider threat detection program?
**

We build systems and safety measures according to our Information Security Policy, considering AI-specific security measures appropriately. [Reference: https://www.softbank.jp/en/corp/security/]

i.We set ‘Information Security Policy’ to address information leakage risks, protecting and handling information assets properly. We strengthen information security management systems by allocating a Chief Information Security Officer, enhancing internal regulations, establishing audit systems, implementing system security, educating employees, and strengthening management of outsourcing agents. All information assets and involved employees and contractors adhere to this policy.
[Reference: https://www.softbank.jp/en/corp/security/]

ii. For highly confidential information, we set minimal access permissions and implement measures such as multi-factor authentication and encryption as needed.

iii.We investigate vulnerabilities in collaboration with government agencies and industry organizations, prioritize them, and address them accordingly. Directives for handling vulnerabilities are issued based on the level of urgency.

iv. We conduct these activities regularly. However, if we receive high-urgency vulnerability information from government agencies or industry organizations, we may immediately revise our security measures.

v.We have one.

h. How does your organization address vulnerabilities, incidents, emerging risks?

The specific details are not disclosed, but we implement checks through guidelines and checklists, as well as manual adversarial testing via AI red-teaming to prevent vulnerabilities and incidents. We gather and act on new risk information as necessary.

Any further comments and for implementation documentation

No answer provided

Section 3 – Transparency reporting on advanced AI systems

**a. Does your organization publish clear and understandable reports and/or technical documentation related to the capabilities, limitations, and domains of appropriate and inappropriate use of advanced AI systems?
i. How often are such reports usually updated?
ii. How are new significant releases reflected in such reports?
iii. Which of the following information is included in your organization's publicly available documentation: details and results of the evaluations conducted for potential safety, security, and societal risks including risks to the enjoyment of human rights; assessments of the model's or system's effects and risks to safety and society (such as those related to harmful bias, discrimination, threats to protection of privacy or personal data, fairness); results of red-teaming or other testing conducted to evaluate the model's/system's fitness for moving beyond the development stage; capacities of a model/system and significant limitations in performance with implications for appropriate use domains; other technical documentation and instructions for use if relevant.
**

In generative AI development, we disclose training environments, model structures, and limitations in model cards for publicly available models.

i. Models are updated irregularly. Updates may occur whenever the model is revised.
ii. Our new model cards reflect these updates.
iii.
・details and results of the evaluations conducted for potential safety
・capacities of a model/system and significant limitations in performance with implications for appropriate use domains
・other technical documentation and instructions for use if relevant.

b. 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?

We publish information on our homepage, technical blog, and the respective site when the model is released. Furthermore, we established an AI Ethics Committee in April 2024 to incorporate diverse perspectives, including external experts, enhancing AI governance and addressing ethical issues in AI system development and operation, such as bias, fairness, explainability, and transparency proactively, ensuring objectivity in our internal rules and responses to sampled cases.

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

Yes, we do.
[Reference: https://www.softbank.jp/en/privacy/]

d. 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?

We disclose part of the safety evaluation data and frameworks.
[Reference: https://www.sbintuitions.co.jp/blog/entry/2025/03/19/120122]

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

We contribute AI development technologies in conferences and publications.

Any further comments and for implementation documentation

No answer provided

Section 4 – Organizational governance, incident management and transparency

a. How has AI risk management been embedded in your organization governance framework? When and under what circumstances are policies updated?

We integrate risk checks into the workflow by conducting review meetings involving departments related to security, data, legal, intellectual property, and AI ethics/governance before releasing any services.

Our governance framework includes policies, internal regulations, guidelines, checklists, and other tools to create an environment where autonomous checks can be performed. We commit to revising and flexibly adapting our policies as necessary, taking into account AI-related guidelines from various countries and regions, changes in people's lifestyles and environments, industry case studies, technological advancements, and dialogue with various stakeholders.

b. Are relevant staff trained on your organization's governance policies and risk management practices? If so, how?

We share information and conduct training in collaboration with group companies, consulting firms, related ministries, and other companies. We also accumulate knowledge through advice in regularly held AI Ethics Committees with external experts.

c. Does your organization communicate its risk management policies and practices with users and/or the public? If so, how?

Externally, we disclose the risk management structure and methods (https://www.softbank.jp/en/corp/aboutus/governance/riskmanagement/) and have published the 'SoftBank AI Ethics Policy', aiming for the proper use of AI for people's happiness (https://www.softbank.jp/en/corp/aboutus/governance/ai-ethics/).
Internally, we provide mandatory learning content for all employees, conduct study sessions twice a year, and send monthly newsletters for deeper understanding of AI risks and countermeasures.

d. Are steps taken to address reported incidents documented and maintained internally? If so, how?

We have not had any incidents so far, but we will handle incidents according to our incident response flow already defined within our organization. The incident reporting procedures are documented, managed, and stored in designated access-controlled locations.

e. How does your organization share relevant information about vulnerabilities, incidents, emerging risks, and misuse with others?

We obtain vulnerability and risk information from government agencies and industry organizations, and we share our own information as necessary.

Internally, we collaborate with relevant departments on incidents and new risks, share risk trend analyses, provide mandatory learning content for all employees, conduct biannual study sessions, and distribute a monthly newsletter.

f. Does your organization share information, as appropriate, with relevant other stakeholders regarding advanced AI system incidents? If so, how? Does your organization share and report incident-related information publicly?

We have not had any incidents, so there is no record of sharing or publishing incident information.

As for industry examples, we share them through mandatory learning content for all employees, study sessions, and monthly newsletters. Should a significant incident occur, we plan to disclose information promptly through press releases, announcements, and notifications to customers as needed.

g. How does your organization share research and best practices on addressing or managing risk?

We share best practices and mandatory learning content, study sessions, and newsletters in the checklists for each item.
We also hold consultation meetings to examine details and provide advice on cases.

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

We use them.

Any further comments and for implementation documentation

No answer provided

Section 5 – Content authentication & provenance mechanisms

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

According to our internal guidelines, for services utilizing advanced AI systems, we make it a principle to explain to users that AI is being used.

b. Does your organization use content provenance detection, labeling or watermarking mechanisms that enable users to identify content generated by advanced AI systems? If yes, how? Does your organization use international technical standards or best practices when developing or implementing content provenance?

We research content provenance detection, labeling or watermarking mechanisms, pursuing optimal approaches to help users readily identify generated content, referencing international technical standards and best practices.

Any further comments and for implementation documentation

No answer provided

Section 6 – Research & investment to advance AI safety & mitigate societal risks

a. How does your organization advance research and investment related to the following: security, safety, bias and disinformation, fairness, explainability and interpretability, transparency, robustness, and/or trustworthiness of advanced AI systems?

1. Focus on AI Ethics and Governance:
We established an AI Ethics Committee in April 2024 to incorporate diverse perspectives, including external experts, enhancing AI governance and addressing ethical issues in AI system development and operation, such as bias, fairness, explainability, and transparency proactively, ensuring objectivity in our internal rules and responses to sampled cases.
We have also developed the 'SoftBank AI Ethics Policy,' clearly stating human-centered principles, fairness, transparency, safety, privacy protection, and basic principles for responsible AI development and utilization.
We research issues such as bias generated during automatic evaluations using LLM, safety evaluation benchmarks in Japanese, bias in corpus for pre-training, and fingerprinting technology to claim the ownership of our models.

2. Strengthening Security Measures:
We have established specialized organizations like the Information Security Committee (ISC) and SoftBank CSIRT, constructing a strict management system based on our information security policy. This initiative aims

to protect our information assets, including AI systems, from threats such as cyber attacks and information leaks.

Furthermore, we enhance education and training for security personnel, providing opportunities to acquire advanced security skills and knowledge. We've also created a highly specialized white-hat hacker team, the "Red Team", which simulates attacks on our internal systems to strengthen our security measures.

3. Building Robust and Reliable Infrastructure:
We are advancing the construction of next-generation societal infrastructure to support future digital services, including dispersed AI data centers. This effort not only provides a foundation for stable operations of AI systems but also aims to improve the reliability of AI systems by enhancing resilience against disasters.

Moreover, we promote the spread of 5G/6G networks, which feature high speed, large capacity, low latency, and multiple simultaneous connections, creating an infrastructure environment that maximizes AI system performance. This leads to improved user experience and increased reliability of AI systems.

[Reference:
https://www.softbank.jp/en/corp/set/data/sustainability/documents/reports/pdf/sbkk_sustainability_report_2024_en.pdf
]

b. How does your organization collaborate on and invest in research to advance the state of content authentication and provenance?

We are still in the research and development stage but are conducting studies on content authentication and provenance in cooperation with various companies and academic institutions.

c. Does your organization participate in projects, collaborations, and investments in research that support the advancement of AI safety, security, and trustworthiness, as well as risk evaluation and mitigation tools?

We presented a study on the bias in LLM at the NeurIPS 2024 Safe Generative AI Workshop, and also conducted joint research with the National Institute for Japanese Language on safety benchmarks, which was presented in March 2025.
[Reference:
https://www.softbank.jp/en/corp/set/data/sustainability/documents/reports/pdf/sbkk_sustainability_report_2024_en.pdf
]

d. What research or investment is your organization pursuing to minimize socio-economic and/or environmental risks from AI?

1. AI Ethics and Governance:
We established an AI Ethics Committee in April 2024 to incorporate diverse perspectives, including external experts, enhancing AI governance and addressing ethical issues in AI system development and operation, such as bias, fairness, explainability, and transparency proactively, ensuring objectivity in our internal rules and responses to sampled cases.
We have also developed the 'SoftBank AI Ethics Policy,' clearly stating human-centered principles, fairness, transparency, safety, privacy protection, and basic principles for responsible AI development and utilization.

2. Reducing Environmental Impact:

We are investing in decentralized AI data centers that use renewable energy, addressing the increased electricity consumption associated with AI proliferation. This approach not only reduces environmental impact by decentralizing power consumption but also contributes to the stabilization of energy supply.

Moreover, we are investing in the research and development of energy-efficient technologies utilizing AI, aiming to improve power efficiency in base stations and data centers, and to optimize energy management in smart buildings.

3. Human Resource Development:
We offer internal education programs such as "SoftBank University Tech" and "AI Campus from SBU Tech" to cultivate AI engineers capable of promoting AI development. To enhance understanding of AI risks and countermeasures, we also provide mandatory learning content for all employees, conduct biannual study sessions, and distribute a monthly newsletter.

[Reference:
https://www.softbank.jp/en/corp/set/data/sustainability/documents/reports/pdf/sbkk_sustainability_report_2024_en.pdf
]

Any further comments and for implementation documentation

No answer provided

Section 7 – Advancing human and global interests

a. What research or investment is your organization pursuing to maximize socio-economic and environmental benefits from AI? Please provide examples.

1. AI Ethics and Governance:
We established an AI Ethics Committee in April 2024 to incorporate diverse perspectives, including external experts, enhancing AI governance and addressing ethical issues in AI system development and operation, such as bias, fairness, explainability, and transparency proactively, ensuring objectivity in our internal rules and responses to sampled cases.
We have also developed the 'SoftBank AI Ethics Policy,' clearly stating human-centered principles, fairness, transparency, safety, privacy protection, and basic principles for responsible AI development and utilization.

2. Reducing Environmental Impact:
To address the increased electricity consumption associated with the proliferation of AI, we are constructing decentralized AI data centers that utilize renewable energy. This approach not only reduces environmental impact by decentralizing power consumption but also contributes to the stabilization of energy supply.

Moreover, we are investing in the research and development of energy-efficient technologies utilizing AI, aiming to improve power efficiency in base stations and data centers, and to optimize energy management in smart buildings.

3. Human Resource Development:
We offer internal education programs such as "SoftBank University Tech" and "AI Campus from SBU Tech" to cultivate AI engineers capable of promoting AI development. To enhance understanding of AI risks and

countermeasures, we also provide mandatory learning content for all employees, conduct biannual study sessions, and distribute a monthly newsletter.

[Reference:
https://www.softbank.jp/en/corp/set/data/sustainability/documents/reports/pdf/sbkk_sustainability_report_2024_en.pdf
]

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

We offer internal education programs such as "SoftBank University Tech" and "AI Campus from SBU Tech" to cultivate AI engineers capable of promoting AI development. To enhance understanding of AI risks and countermeasures, we also provide mandatory learning content for all employees, conduct biannual study sessions, and distribute a monthly newsletter.

c. Does your organization prioritize AI projects for responsible stewardship of trustworthy and human-centric AI in support of the UN Sustainable Development Goals? Please provide examples.

We are addressing the challenges of increased electricity demand and centralized data processing due to the proliferation of AI by building next-generation societal infrastructure that is compatible with a sustainable society, such as decentralized AI data centers.

We established an AI Ethics Committee in April 2024 to incorporate diverse perspectives, including external experts, enhancing AI governance and addressing ethical issues in AI system development and operation, such as bias, fairness, explainability, and transparency proactively, ensuring objectivity in our internal rules and responses to sampled cases.
We have also developed the 'SoftBank AI Ethics Policy', clearly stating human-centered principles, fairness, transparency, safety, privacy protection, and basic principles for responsible AI development and utilization.

Furthermore, we have formulated the 'SoftBank AI Ethics Policy', which sets guidelines for the ethical use of AI, clearly defining essential principles for responsible AI development and use, including human-centered principles, fairness, transparency, safety, and privacy protection.

[Reference:
https://www.softbank.jp/en/corp/set/data/sustainability/documents/reports/pdf/sbkk_sustainability_report_2024_en.pdf
]

d. Does your organization collaborate with civil society and community groups to identify and develop AI solutions in support of the UN Sustainable Development Goals and to address the world's greatest challenges? Please provide examples.

In response to the increased electricity consumption accompanying the proliferation of AI, we are constructing decentralized AI data centers utilizing renewable energy. This disperses electricity consumption, reducing environmental impact and contributing to energy supply stability.
We have positioned 'Regional Revitalization' as an important theme and are providing AI solutions and digital services to address regional issues in collaboration with local governments. Additionally, we collaborate with over 1,000 NPOs as part of our CSR activities and will continue to deepen relationships with NPO/NGO organizations to promote efforts addressing further social issues.

[Reference: https://www.softbank.jp/en/corp/sustainability/]

Any further comments and for implementation documentation

No answer provided