OECD AI Transparency Report

Organization: ABEJA.inc (JPN) **Reporting Period:** Q4 2025 **Published:** November 18, 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?

Based on factors such as the use case, the user, and the content of the AI's inference, we classify risks according to the entities affected by the risk, and the nature and content of the rights and interests that are impacted.

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

At the planning stage of AI development, before an AI's release, and upon major functional improvements, it is our practice to consult with the team responsible for AI risk management. This team conducts a risk assessment to perform risk identification and evaluation. In identifying risks, we carry out activities such as researching past incidents, exchanging opinions with domain experts, consulting with external specialists, and reviewing relevant documents and guidelines.

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?

We identify appropriate, industry-accepted testing methodologies by researching relevant literature. Subsequently, we conduct mandatory testing prior to release and evaluate the results to determine whether to proceed with deployment.

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 utilize both quantitative and qualitative metrics. We conduct a holistic assessment from various perspectives to evaluate the overall level of risk, noting that these metrics—particularly quantitative ones—do not measure the precise magnitude of the risk itself.

Reporting mechanisms for vulnerabilities and other related issues are available through the repository where the LLM is published.

We do not utilize incentive programs for the discovery of vulnerabilities or other issues.

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?

We sought opinions from an advisory council composed of external experts prior to development.
Regarding the reporting of vulnerabilities and other related issues, the repository where the LLM is published facilitates open discussion, which allows such reports to be made.

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?

Employees from our company are significantly involved in the development of the Japanese government's AI Guidelines for Business, serving as the member of the deliberation committee. Furthermore, one of the employees is serving as the member of the Global Partnership on AI (GPAI) as an expert.

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 exchange opinions with various stakeholders through academia (such as universities), industry associations, and government bodies. Additionally, our organization has established an advisory committee composed of diverse external experts, whom we consult as needed.

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?

Prior to AI development, consultation with the legal team responsible for AI risk management is mandatory. Upon consultation, the legal team conducts an AI risk assessment and establishes risk mitigation measures. Before an AI is released, it must undergo predefined tests, and only those that pass are released. We follow similar steps for large-scale additional training.

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

Based on the test results, we take necessary actions, such as re-training the model.

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

It depends on the circumstances, but testing is always conducted at least prior to release.

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

We define inappropriate data in advance and use filters and other methods to exclude it from the training data.

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

Under Japanese law, the use of copyrighted works for the purpose of information analysis is permitted as an exception to copyright. At the output level, we prevent the generation of content similar to the training data by using an enormous volume of training data.

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

While the definition of privacy can vary, regarding the protection of confidential information—which is central to privacy—we fundamentally use only publicly available information for our training data.

**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?
**

1.We conduct security risk assessments based on standards such as ISO.
2.We manage access controls according to the importance of information assets. Data is stored using highly secure cloud servers.
3. We employ a continuous, risk-based approach based on standards such as ISO, executing a clear lifecycle of vulnerability identification, prioritization, response, re-evaluation, and improvement.
4. Measures are reviewed at least annually, and additionally as needed
5.We have internal regulations regarding the prohibition of insider threat and related activities.

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

We maintain a continuous, ISMS-compliant security posture across the entire AI system. We comprehensively address vulnerabilities, incidents, emerging risks, and misuse through risk assessment from the design phase, strict vulnerability management, and real-time post-deployment monitoring and incident response.

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.
**

i. We disclose information at the time of release. Subsequently, we plan to update the disclosures as necessary when new information is obtained.
ii. Model version upgrades have not yet been implemented.
iii. Information regarding harmful bias, discrimination, threats to privacy or personal data protection, and fairness is described within the procedures for constructing the dataset used for the model.
https://github.com/abeja-inc/Megatron-LM/blob/main/docs/dataset/about_data.md
https://tech-blog.abeja.asia/entry/abeja-nedo-project-part2-202405

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 make this information publicly available through blogs and other formats.

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

We publish a privacy policy.
https://www.abejainc.com/security-policy-and-privacy-policy

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?

This information is published on our blog and elsewhere.

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

We disclose our initiatives for Advanced AI at various presentation opportunities, such as serving as references for the government and through research at universities.

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?

AI risk management is stipulated in various internal regulations, including our risk management policy. We have appointed a head of AI risk management within the company. We plan to revise these policies as necessary

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

Yes. The head of AI risk management conducts company-wide seminars on AI risks.

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

We have formulated an AI policy and made it publicly available on our website.
https://www.abejainc.com/ai-policy

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

Incident reports are documented and managed in accordance with internal regulations.

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

We disclose this information on the system screen for SaaS models, and on the repository for Open Access models

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 publish our Advanced AI on a public repository. We will disclose incident information on this repository.

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

Our AI risk manager is an expert at the Global Partnership on AI (GPAI) and shares information through GPAI projects. Additionally, our employees may serve as members of Japanese government committees on AI risk management, sharing information through such committees. We also belong to the Japan Deep Learning Association (JDLA), one of Japan's largest industrial organizations for AI, and provide information through it.

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

We reference ISO 42001, the (Japanese) AI Guidelines for Business, and the NIST AI RMF

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?

For services such as chatbots, we indicate that it is an AI chatbot on the system's top screen or equivalent location.

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 clearly indicate on the service screen that it is a generative AI service.

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?

Our R&D team researches various cutting-edge AI technologies, including the themes listed. As a company, we also participate in various research groups in academia and other areas to conduct research on these topics.

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

We are currently investigating the usefulness of such content authentication itself.

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 do so through joint projects with universities and the government

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

We conduct various literature reviews and present our findings at government committees and other venues. Additionally, our employees individually publish papers and books.

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.

We are advancing research with the government on the use of AI in the medical field, among other areas.

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 frequently dispatch lecturers to university lectures and to seminars held by industry associations and academic societies.

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 emphasize the perspective of responsible AI and do not develop or accept contracts for AI that poses human rights issues.

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.

We frequently exchange opinions with government bodies, universities, industry associations, and academic societies.

Any further comments and for implementation documentation

No answer provided