Report

Organization: **KDDI Corporation**

https://www.kddi.com

Publication date: Apr 22, 2025, 08:59 AM PDT

Reporting period: 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?

KDDI has published the "AI R&D and Utilization Principles for KDDI Group" [1]. Based on the values outlined in these principles, KDDI categorizes risks on a use-case basis, while considering both domestic and international guidelines and regulations, including:

* AI Guidelines for Business (MIC and MEXT, Japan)[2]

* AI Act (EU)[3]

* AI Risk Management Framework (NIST)[4]

[1] AI R&D and Utilization Principles for KDDI Group

https://www.kddi.com/english/corporate/kddi/public/ai_principles/

[2] AI Guidelines for Business

https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/02ryutsu20_04000019.html

[3] EU AI Act

https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

[4] NIST AI Risk Management Framework

https://www.nist.gov/itl/ai-risk-management-framework

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

KDDI has developed a checklist regarding the risks mentioned in the response to *section 1.a*. Specifically, this checklist addresses concrete examples of AI-related risks and assesses whether those risks exist. In this context, we consciously adopt a "by design" approach to utilize this checklist during the planning phase of AI systems and to conduct evaluations.

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?

At KDDI, internal auditors conduct interviews to assess the appropriateness of AI systems by utilizing the responses from the checklist referenced in *section 1.b* during the planning phase. Additionally, KDDI convenes an Advisory Board meeting [1] as needed to solicit insights on appropriate data utilization, including AI, by inviting external experts.

[1] KDDI Advisory Board on Appropriate Data Utilization for the Realization of Society 5.0 (Japanese only)

https://www.kddi.com/corporate/kddi/public/privacy-portal/advisory-board/

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?

* KDDI utilizes both quantitative and qualitative risk assessment indicators, considering risk impacts and other relevant factors. Its internal regulations establish quantitative and qualitative criteria, and risk assessments are conducted based on the significance of the systems [1].

* All stakeholders can report vulnerabilities or incidents through KDDI's inquiry contact form [2].

* Currently, there is no incentive program for the responsible disclosure of risks, incidents, or vulnerabilities.

[1] Risk Management and Internal Controls

https://www.kddi.com/english/corporate/sustainability/risk-management/

[2] Contact

https://www.kddi.com/english/contact/

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 evaluating risks, KDDI utilizes Advisory Board meetings with external experts, as mentioned in the response to *section 1.c*, to gather feedback. Additionally, as indicated in the response to *section 1.e*, KDDI has established an inquiry contact form for third-party inquiries, including reports.

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?

KDDI adheres guidelines and best practices in accordance with the response to *section 1. a*.

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?

KDDI engages in discussions regarding risk mitigation measures through various channels, including:

* Advisory Board meetings with external experts, as mentioned in the response to *section 1.c*.

* Participation in the AI Governance Association [1].

[1] AI Governance Association (Japanese only).

https://www.ai-governance.jp/resources/home

Section 2 - Risk management and information security

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

To address the risks referenced in the response to *section 1.a*, KDDI has developed a checklist. We intentionally utilize this checklist during the planning phase of AI systems to ensure effective risk management.

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

KDDI has provided examples of relevant testing methods corresponding to each item in the checklist mentioned in the response to *section 2.a*, specifically concerning approaches to address risks and vulnerabilities.

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

KDDI conducts tests in a validation environment that is isolated from the commercial environment, both prior to commercial deployment and as necessary.

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

ELYZA, KDDI's subsidiary that develops AI models, is focused on the following objectives:

- Automatically excluding content that contains inappropriate expressions (offensive, discriminatory, or sexual content) from the data utilized.
- Additionally, manual removal of harmful content is conducted in accordance with the classification recommended by MLCommons' taxonomy of hazards. Guidelines are established to ensure that responses are sensitive to gender and minority considerations.

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

- KDDI strives to protect rights in compliance with applicable laws and regulations in the respective countries and regions.
- In particular, KDDI's subsidiary ELYZA enhances rights protection through copyright management, the conclusion of licensing agreements, and legal measures in the event of infringement. Furthermore, it adheres to robots.txt, refraining from utilizing data if crawling is not permitted.

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

KDDI discloses and adheres to its privacy protection policy, which includes safety management measures to prevent data leaks, referred to as "KDDI's Privacy Policy" [1]. To ensure compliance, the responsible department conducts privacy impact assessments prior to service implementation [2].

[1] KDDI's Privacy Policy

https://www.kddi.com/english/corporate/kddi/public/privacy/

[2] Initiatives Related to Privacy Governance at KDDI (Japanese only)

https://www.soumu.go.jp/main_content/000810946.pdf

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?**

KDDI has investigated and analyzed external trends regarding AI and reflects these findings in its internal security regulations and guidelines for system development.

- i. KDDI conducts risk assessments based on internal regulations and guidelines formulated in accordance with "(KDDI's) Security Policy" [1], evaluating risks according to the significance of the systems. KDDI has established regulations and guidelines to enhance the security of advanced AI systems.
- ii. KDDI has obtained ISO/IEC 27001 (ISMS) certification and implements security management in compliance with this standard. Specifically, as outlined in "Initiatives for Providing Safe and Secure Services" [2], KDDI classifies information assets based on the internal regulations regarding information security, established in consideration of "(KDDI's) Security Policy" [1], and defines the handling of these assets according to their significance. KDDI takes security measures based on the importance of individual information assets and ensures thorough management.
- iii. As indicated in "Initiatives for Providing Safe and Secure Services" [2], KDDI collects vulnerability information and evaluates its impact on systems. KDDI collaborates within information-sharing frameworks with public institutions and industry associations to address risks and vulnerabilities.
- iv. KDDI reviews and implements security measures at least once a year.
- v. KDDI detects threats through security monitoring.

[1] (KDDI's) Security Policy

https://www.kddi.com/english/corporate/kddi/public/security/

[2] Initiatives for Providing Safe and Secure Services (Japanese only)

https://www.kddi.com/english/corporate/kddi/public/security-portal/safety

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

KDDI has established internal regulations based on "(KDDI's) Security Policy" [1] to address vulnerabilities, incidents, and risks throughout the system lifecycle. This policy applies to all systems, including AI systems. Specifically, this information is documented in the "Cybersecurity Annual Report 2024" [2].

[1] (KDDI's) Security Policy

https://www.kddi.com/english/corporate/kddi/public/security/

[2] Cybersecurity Annual Report 2024

https://www.kddi.com/extlib/files/english/corporate/kddi/public/security-portal/cybersecurity-annual-report/pdf/csar_2024_e.pdf

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.

For advanced AI systems, while KDDI shares terms of use with actual users, it does not publish them publicly. Instead, for services available to the public, KDDI discloses these terms in the form of publicly accessible documents, such as the following:

https://www.kddi.com/extlib/files/corporate/kddi/kokai/keiyaku_yakkan/pdf/au_support.pdf

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?

KDDI conducts information sharing in various formats, including:

- Sharing information through discussions with the AI Governance Association [1], in which we participate.
- Participating in national projects related to the systematization of AI security research and disseminating results through the “AI Security Portal” [2].

[1] AI Governance Association (Japanese only)

https://www.ai-governance.jp/resources/home

[2] AI Security Portal (Currently, only the Japanese version is available, with the English version scheduled for release soon.)

https://aisecurity-portal.org/

---

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

KDDI discloses relevant information on its website under the title “KDDI's Privacy Policy” [1].

[1] KDDI's Privacy Policy

https://www.kddi.com/english/corporate/kddi/public/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?

KDDI provides data only when necessary. However, in accordance with its Security and Privacy Policy [1,2], KDDI excludes information pertaining to trade secrets and privacy-related data.

[1] (KDDI's) Security Policy

https://www.kddi.com/english/corporate/kddi/public/security/

[2] KDDI's Privacy Policy

https://www.kddi.com/english/corporate/kddi/public/privacy/

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

There is nothing particularly additional to mention.

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?

KDDI has established the “AI R&D and Utilization Principles for KDDI Group” [1] as part of its governance framework and is committed to adhering to them. As noted at the bottom of these principles, they will be continuously reviewed and flexibly revised as necessary based on technological innovations in AI systems and services, as well as on international trends.

[1] AI R&D and Utilization Principles for KDDI Group

https://www.kddi.com/english/corporate/kddi/public/ai_principles/

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

KDDI conducts training for appropriate staff through courses within its internal human resource development program, KDDI DX University [1].

[1] AI Use Cases and Initiatives for AI Governance at KDDI (Japanese only)

https://www.soumu.go.jp/main_content/000826718.pdf

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

KDDI discloses its “(KDDI's) Security Policy” [1] and “KDDI's Privacy Policy” [2] on its website. Furthermore, KDDI has established dedicated portals [3,4] to explain these policies and facilitate communication with users.

[1] (KDDI's) Security Policy

https://www.kddi.com/english/corporate/kddi/public/security/

[2] KDDI's Privacy Policy

https://www.kddi.com/english/corporate/kddi/public/privacy/

[3] Security Portal

https://www.kddi.com/english/corporate/kddi/public/security-portal/

[4] Privacy Portal

https://www.kddi.com/english/corporate/kddi/public/privacy-portal/

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

KDDI has established a risk management system to conduct these processes.

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

KDDI engages in information sharing within established frameworks with public institutions and industry associations [1].

[1] Cybersecurity Annual Report 2024

https://www.kddi.com/extlib/files/english/corporate/kddi/public/security-portal/cybersecurity-annual-report/pdf/csar_2024_e.pdf

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?

KDDI implements the practices outlined in the response to *section 4.e*, regardless of whether they pertain to advanced AI systems.

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

Through KDDI's subsidiary, the KDDI Research Institute, it publicly shares the results of research related to risk management and response through seminars and portals [1].

[1] AI Security Portal (Currently, only the Japanese version is available, with the English version scheduled for release soon.)

https://aisecurity-portal.org/

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

KDDI refers to Japan's “AI Guidelines for Business”, which are based on international technical standards and best practices regarding AI risk management and governance, as mentioned in *section 1.a*.

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?

One of the principles KDDI adheres to in the "AI R&D and Utilization Principles for KDDI Group" [1] is “accountability”, which encompasses the important aspect of disclosing the use of AI.

[1] AI R&D and Utilization Principles for KDDI Group

https://www.kddi.com/english/corporate/kddi/public/ai_principles/

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?

KDDI is advancing the research and development referenced in the response to *section 6.b*; however, actual operational deployment has not yet taken place.

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?

Through our subsidiary, KDDI Research Inc., we are conducting research and development [1] on security, safety, transparency, misinformation, and robustness.

[1] AI Use Cases and Initiatives for AI Governance at KDDI (Japanese only)

https://www.soumu.go.jp/main_content/000826718.pdf

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

KDDI, through its subsidiary KDDI Research Inc., conducts research on topics such as:

- Regarding text: Researching mechanisms to detect machine-translated texts generated by AI [1].
- Regarding images: Researching electronic watermarking techniques to identify sources.

[1] Machine Translated Text Detection Through Text Similarity with Round-Trip Translation

https://aclanthology.org/2021.naacl-main.462/

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?

Through our subsidiary, KDDI Research Inc., KDDI participates in national projects in Japan related to the systematization of AI security research; as a result, KDDI has launched a portal [1].

[1] AI Security Portal (Currently, only the Japanese version is available, with the English version scheduled for release soon.)

https://aisecurity-portal.org/

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

As one of the environmental risks, the substantial electricity consumption associated with AI utilization is noted. KDDI is addressing this issue by developing immersion cooling technology for servers used in data centers [1]. Additionally, KDDI is engaged in systematic research on the negative impacts of AI misuse and degradation on society, including the organization of countermeasures.

[1] Introduction to Initiatives for Achieving a Decarbonized Society and Participation in Related Organizations and Initiatives (Japanese only)

https://www.kddi.com/extlib/corporate/sustainability/efforts-environment/carbon/pdf/homepage_CN.pdf

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.

As stated in KDDI's corporate vision, the company aims to create new values and address societal issues through the evolution of the power to connect using AI, and it is investing accordingly [1]. For instance, it is investing in emergency response utilizing AI drones and the implementation of sleep functions for AI-powered base stations. Additionally, KDDI is actively investing in AI startups [2, 3] and is engaged in business co-creation with other companies [4].

[1] KDDI Integrated Sustainability and Financial Report

https://www.kddi.com/extlib/files/english/corporate/ir/ir-library/sustainability-integrated-report/pdf/kddi_sir2024_e.pdf

[2] Japan’s KDDI to Take Control of Generative AI Startup Elyza

https://japannews.yomiuri.co.jp/business/companies/20240319-175463/

[3] KDDI Open Innovation Fund III Invests in Sakana AI

https://www.kddi.com/english/open-innovation-program/news/20240116/

[4] Announcement Regarding KDDI-MUFG Collaboration: Next Step (Co-creation 2.0) (Japanese only)

https://newsroom.kddi.com/news/detail/kddi_nr_s-19_3593.html

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.

KDDI conducts internal education for all employees [1]. Conversely, externally, it promotes the societal implementation of the activities mentioned in the questions through its participation in the AI Governance Association [2].

[1] AI Use Cases and Initiatives for AI Governance at KDDI (Japanese only)

https://www.soumu.go.jp/main_content/000826718.pdf

[2] AI Governance Association (Japanese only)

https://www.ai-governance.jp/resources/home

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.

KDDI recognizes respect for human rights as one of its material issues (materialities) and has identified the critical concern outlined in the "KDDI Group Human Rights Policy" [1] as the "avoidance of human rights violations arising from the advancement of technology" (including AI). Furthermore, while this is not limited to AI, KDDI regards respect for human rights as a crucial indicator in the implementation of its projects. To enhance this further, KDDI is engaging in dialogue with external experts [2].

[1] KDDI Group Human Rights Policy

https://www.kddi.com/english/corporate/kddi/philosophy/human-rights/

[2] KDDI Integrated Sustainability and Financial Report 2024

https://www.kddi.com/extlib/files/english/corporate/ir/ir-library/sustainability-integrated-report/pdf/kddi_sir2024_e.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.

KDDI presents its policies for addressing issues as "Social Contribution Activities" [1]. The company also discloses the activities undertaken in conjunction with this on its website.

In addition, in collaboration with Amazon Web Services Japan, KDDI provides comprehensive support for the social implementation of generative AI for businesses and local governments. This collaboration aims to address issues such as labor shortages and operational efficiency by offering support for utilizing generative AI developed by startups and open-source generative AI, along with tailored AI solutions to meet the specific challenges faced by each entity [2].

[1] Social Contribution Activities

https://www.kddi.com/english/corporate/sustainability/contribution/

[2] KDDI to Collaborate with AWS to Bring Generative AI to Society

https://newsroom.kddi.com/english/news/detail/kddi_pr-991.html