Report

Organization: **Data Privacy and AI**

www.dataprivacyandai.com

Publication date: Apr 22, 2025, 09:00 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?

Organization identify and differentiate the risks of AI systems based on the risk classification pursuant EU-AI-Act. The first step is to find out if the AI Systems provide a forbidden practice while using and interaction with us humans. Second step is to identify if it´s a high-risk AI system pursuant Art. 6 EU-AI-Act. If no, then identify if its an AI system with or without a systemic risk pursuant Art. 50 EU-AI-Act. Based on this results derivate the regulatory requirements the AI-System has to follow and to fulfill. Additional to this we have to identify the possible user group of the AI-System and to find out specific risks we have to handle and to mitigate with this group, especially if its a vulnerable user group. In this context we have to check, the way of access to the AI-System. The other risks are from the use case, the purpose, physical product in which the AI-System will work, possible interactions with other AI systems and the area of operation, e.g. autonomous driving, health case. Finally we have to check, what will happen, when the AI-System doesn´t work, when data processings are interrupted or the automated decision making make mistakes - what would be the consequences for provider, deployer and user groups. Based on this results we can create a risk register including probability, risk score and extent of damage.

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 are using technical tools to identify and evaluate risks.

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?

My organization is a consulting company to guide through EU-AI-Act, ISO/IEC42001 and additional norms and standards from ISO/IEC, IEEE etc. For testing we refer to tools like "Orthrus" from AI & Partners.

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?

At the moment only if a data breach pursuant GDPR is happened. Currently we are in the beginning with the understanding of EU-AI-Act and how we can implement it. For future it would be great opportunity to help other organizations with incident reporting and to offer programs how responsible can solve problems and mitigate risks of their AI-Systems.

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?

Yes, we are in the Network of Governance Experts from AI & Partners, IEEE.org. A global risk register would be helpful in the future to share experiences of AI Systems and their risks and tools/workflows to mitigate the risks.

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?

Using of this standards, e.g. IEEE 7000-2021, ISO/IEC42001, ISO/IEC5259-1, NIST 2.0 Cybersecurity Framework

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?

Particular, I recommend to my clients to install an ethic-board or an expert of AI-Ethics/AI-Risks in other advisory boards or steering committees and to create a risk register to have an overwiew about the mitigation status.

Any further comments and for implementation documentation

Creating of policy for risk management in context with AI based on HLC of ISO/IEC42001 and additional Norm ISO/IEC23894:2023 AI Guidance on risk management, Implementing a protection level concept including a threshold to trigger an approval procedure, if the risk level is to high. Implementing real-time-monitoring features as part of post-marked monitoring, Implementing Iteration to review and mitigate the risks asap.

Section 2 - Risk management and information security

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

I recommend to my clients to implement an internal auditing procedure (audit management) as a part of internal control system. Additional to record all AI-Assets in a digital inventory to know about AI-Tools an their risks. For Developers I recommend to have a procedure in the design and training process to eliminate forbidden practice during design process (supported by IEEE7000-2021). For purchasing departments I recommend to ask for information, if companies are purchasing AI as a Service - additional in the process of AI-Purchaising an check to eliminate forbidden practice is needed. Finally I recommend to implement ISO/IEC42001 - AI-Managementsystem to handle risks along whole AI-Lifecycle.

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

Testing is supported by technical tools they can address the risks and give developer some recommendations to mitigate the risks and to improve development.

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

not applicable, I´m not a developer. In general Testing should take place in the end of training procedures and when AI-System is leaving the test environment it needs a second test in real time environment with feedback loops.

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

I recommend to my clients to define the needed quality of data for corpora. Additional I recommend to have a clear process for training data including legal basis pursuant GDPR and fulfillment of all human rights. Furthermore it should be clear who will interact in real life with this AI. Based on this findings the set up for the corpora could be selected.

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

It´s part of "process for training data". In this procedure is to check, if intellectual property including copy-rights are fulfilled.

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

It does exist "terms of use" for AI-Systems. Data Privacy Teams has to check AI-Tools in view of data privacy topics before AI-Tools are in use.

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?**

I recommend to my clients to use ISO/IEC27001 ff. as standard and additional to implement an internal process to check the cyber security risk and to review them in a periodical routine.

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

I recommend to my clients to install a reporting point including reporting procedure. Part of this should be a workflow to analyze vulnerabilities, incidents and risks including solving procedure and due date.

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´m not a developer of AI, but in my recommendations I suggest the approach of information and to install an information procedure to all stakeholders along the whole AI-Life-Cycle. That points are good point to integrate it in the information procedure as minimum standard.

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?

website, social media, newsletter of information

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

yes

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?

yes, additional it´s part of the advice to other organizations to provide information about sources of data used for training, it´s part of ISO/IEC42001 we are use as basis

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

yes,

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?

It´s embedded in Managementsystems for AI based on ISO/IEC42001, Risk Management is mandatory part of ISO/IEC42001 and requirement of EU-AI-Act. The updating follows the common routine for Management Systems, in circumstances of new recognitions and risk with impact a shorter up-date is needed, then it´s to do following of standardized review and approval procedure.

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

Yes, we created a webinar to understand the content and the need of the policies for governance and risk management. ISO/IEC42001 requires the understanding of the policies.

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

internal risk management communication

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

in progress, should adapted into other incident response procedures as we know from data privacy or security incidents.

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

I recommend to my clients to inform via website or an instruction guideline, as well as we have to publish based on EU-AI Act in accessible way.

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?

not at the moment, no infrastructure to do this inside the country

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

I recommend to my clients to share it via website, social media or newsletter, speeches ...

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

Yes, I recommend to my clients to use ISO/IEC Standards, as 42001 and 42005, NIST-Framework or additional standards from IEEE, ETSI, CEN...

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?

I´m not a developer, but I advice to my clients to publish Dataprotection Notice pursuant GDPR on website and AI-System should be a kind of label, that´s clear for the users that they are interacting with an AI-Tool or got a response message from AI-Assistant.

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?

no, not yet. How can we find the standards?

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?

knowledge transfer, trainings to the named topics, use of testing-tools and riskmanagement software, implementing feedback loops and iteration procedures based on CIP, self assessments

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

not applicable

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?

curently in children's shoes, should be more

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

Implementing ESG as management approach, implementing human-centered design approach based on ISO/IEC9241-210:2019, training to understand "principles of trustworthy and ethical AI" based on EU-AI-Act recital 27, CIP

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.

Implementing ESG as management approach, implementing human-centered design approach based on ISO/IEC9241-210:2019, training to understand "principles of trustworthy and ethical AI" based on EU-AI-Act recital 27, CIP

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.

Yes, I´m a Lecturer on a university of Applied Science for "digital product management" and I impart knowledge about AI-Ethics, human-centered design approach, ISO42001:2023, conformity checks, requirements of EU-AI Act, GDPR, Data Act etc. for privacy by design, safety by design, compliance by design etc. Additional I´m a trainer on DIHK (Deutsche Industrie und Handelskammer) and train the AI-Managers. I write articles to "AI-Literacy", "Protection of AI-Training data", "AI and AI Regulation – a driver for digital and sustainable change?", "ESG and human-centered design in AI as a new leadership approach – a synergy for more humanity and sustainability with and through AI?" and I´m a contributor by AI&Partners to reflect their reports.

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.

yes, in my advice and support based on requirements of EU-AI Act it´s mandatory for me to priorize named projects, e.g. bliro.io. Further more in my role as auditor it´s unacceptable to focus revers projects.

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.

yes, I´m member of the "Innovethic advisory board", Member in KI-Park Berlin and KImpact Switzerland, Member and Voice in AIGN - AI-Governance Network, Contributor by AI&Partners, Member in working groups for digital transformation by GPM Germany