ADDRESSING THE PRINCIPLES-TO-PRACTICE GAP IN AI ETHICS AND REGULATION: DESIGNING AN AI GOVERNANCE REFERENCE MODEL

Online Appendix

AI Governance Task List

Version 2.0

Key premises of the AI Governance Task List

The AI Governance Task List is a practice-oriented framework for implementing responsible AI. With it, organizations can adopt a controlled, human-centric approach that covers the entire process of AI system development and operations.

The AI Governance Task List supports compliance with the European AI regulation (AI Act). This makes it relevant for organizations that perform in-house AI system development, particularly in high-risk application areas. Furthermore, the framework serves as a practical guide for any organization taking steps towards more responsible AI.

The AI Governance Task List provides a template for decision-makers to address the key questions on governing AI systems. Organizations may use it to design and implement practices and derive organization-specific frameworks for governing their AI systems.

Version history

Version	Notes	Date
0.1	Initial version.	06/2021
0.2	Revision based on feedback from design partners. Added “AI algorithm” focus area (renamed “AI model” in version 2.0) and respective tasks. Edited descriptions of tasks based on design partner feedback.	08/2021
0.3	Revision based on evaluation with selected AI systems of design partners.	02/2022
0.4	Condensed version based on interviews and legal analysis.	09/2022
1.0	First published version. Cross-checked and consolidated task descriptions for publication.	02/2023
2.0	Second published version. Merged relevant parts of the “Development and operations” focus area under “Accountability and ownership”. Reordered the focus areas for clarity. Refined terminology to “AI models” rather than “AI algorithms” to increase precision and reflect current terminology. Revised the structure of tasks, merged similar tasks, and added a matrix table to make the set of AI governance controls more explicit. Significantly revised the task descriptions and clarified organizational roles.	12/2024

List of AI governance focus areas and tasks

Focus area	Task ID	Task description	Governance controls
A. Accountability and ownership	A1	Organization-level accountability	
	A2	AI system-level accountability	
	A3	AI development, operations, and governance integration	
B. AI system	B1	AI system registry and AI ID	Pre-planning
	B2	AI system version control	
	B3	AI system planning and scoping	Design & planning
	B4	AI system environment and architecture	
	B5	AI system metrics	
	B6	AI system verification and validation	Assessment & verification
	B7	AI system approval	
	B8	AI system performance monitoring	Continuous monitoring
	B9	AI system health checks	Periodic health checks
C. AI model	C1	Model registry and model ID	Pre-planning
	C2	Model version control	
	C3	Model planning and scoping	Design & planning
	C4	Model technical environment design	
	C5	Model metrics	
	C6	Model verification and validation	Assessment & verification
	C7	Model approval	
	C8	Model performance monitoring	Continuous monitoring
	C9	Model health checks	Periodic health checks
D. AI data	D1	Data source documentation	Pre-planning
	D2	Data preprocessing	Design & planning
	D3	Data quality metrics	
	D4	Data ontologies, inferences, and proxies	Assessment & verification
	D5	Data quality assurance	
	D6	Data quality monitoring	Continuous monitoring
	D7	Data health checks	Periodic health checks
E. Risk and impact management	E1	AI system harms and impacts pre-assessment	Pre-planning
	E2	Model risk assessment	
	E3	AI system impact metrics design	Design & planning
	E4	AI system impact assessment	Assessment & verification
	E5	AI system non-discrimination assurance	
	E6	AI system negative impact minimization	
	E7	AI system impact monitoring	Continuous monitoring
	E8	AI system impact health checks	Periodic health checks
F. Transparency, explainability, and contestability (TEC)	F1	TEC expectation assessment	Pre-planning
	F2	TEC design	Design & planning
	F3	TEC validation and assurance	Assessment & verification
	F4	TEC monitoring	Continuous monitoring
	F5	TEC health checks	Periodic health checks
G. Regulatory compliance	G1	Regulatory environment assessment	Pre-planning
	G2	Regulatory risks, constraints, and design parameter analysis	Design & planning
	G3	Regulatory design consultation	
	G4	Compliance assessment	Assessment & verification
	G5	Compliance monitoring	Continuous monitoring
	G6	Compliance health checks	Periodic health checks

AI governance task matrix

Governance focus areas	Governance controls				
	Pre-planning	**Design & planning**	**Assessment & verification**	**Continuous monitoring**	**Periodic health checks**
A. Accountability and ownership A1 Organization-level accountability A2 AI system-level accountability A3 AI development, operations, and governance integration					
B. AI system	B1 AI system registry and AI ID B2 AI system version control	B3 AI system planning and scoping B4 AI system environment and architecture B5 AI system metrics	B6 AI system verification and validation B7 AI system approval	B8 AI system performance monitoring	B9 AI system health checks
C. AI model	C1 Model registry and model ID C2 Model version control	C3 Model planning and scoping C4 Model technical environment design C5 Model metrics	C6 Model verification and validation C7 Model approval	C8 Model performance monitoring	C9 Model health checks
D. AI data	D1 Data source documentation	D2 Data preprocessing D3 Data quality metrics	D4 Data ontologies, inferences, and proxies D5 Data quality assurance	D6 Data quality monitoring	D7 Data health checks
E. Risk and impact management	E1 AI system harms and impacts pre-assessment E2 Model risk assessment	E3 AI system impact metrics design	E4 AI system impact assessment E5 AI system non-discrimination assurance E6 AI system negative impact minimization	E7 AI system impact monitoring	E8 AI system impact health checks
F. Transparency, explainability, and contestability (TEC)	F1 TEC expectation assessment	F2 TEC design	F3 TEC validation and assurance	F4 TEC monitoring	F5 TEC health checks
G. Regulatory compliance	G1 Regulatory environment assessment	G2 Regulatory risks, constraints, and design parameter analysis G3 Regulatory design consultation	G4 Compliance assessment	G5 Compliance monitoring	G6 Compliance health checks

A Accountability and ownership

The Accountability and ownership focus area entails putting in place the decision rights and responsibilities to govern AI systems and AI models and integrating AI governance with organizational AI development, operations, and governance.

1 Organization-level accountability

The organization should establish an organizational role accountable for overseeing AI system development and operations in the organization. We refer to this organizational role as Head of AI.

The Head of AI should have sufficient knowledge and understanding of the organization's AI system operations to make informed risk-reward decisions.

The Head of AI should have adequate authority and resources to implement and entrench coordinated AI operations.

2 AI system-level accountability

Each AI system should have an organizational owner accountable for its development, deployment, operations, and ultimately, retirement in cooperation with the Head of AI, as appropriate in each case. We refer to this organizational role as AI System Owner.

The AI System Owner should have sufficient knowledge and understanding of the AI system to make informed risk-reward decisions.

The AI System Owner should have adequate authority and resources to ensure that the AI system's development and use align with the organization's objectives and risk tolerance.

In addition, when appropriate, organizations may appoint an AI Model Owner for each AI system. The AI Model Owner is accountable for the model's development, deployment, operations, and retirement in cooperation with the AI System Owners and the Head of AI, as appropriate in each case. AI models can be developed in-house and/or sourced from vendors.

3 AI development, operations, and governance integration

The Head of AI should ensure that AI governance is integrated with other organizational governance processes in a way appropriate for the organization.

The Head of AI should ensure that the organization designs and implements

1) appropriate workflows and processes for its AI-related data acquisition, permitting, and analytics operations, including approvals and sign-offs,
2) appropriate workflows and processes for its model development and AI system development operations, including approvals and sign-offs,
3) appropriate workflows and processes for AI system operations, including approvals and sign-offs, and
4) appropriate workflows and processes for AI retirement, including approvals and sign-offs.

The workflows and processes should align with the organization's objectives and risk tolerance.

The Head of AI should establish clear decision rights between the organizational and AI system levels, including authority limits, escalation triggers, and coordination mechanisms. This includes defining when AI System Owners must escalate decisions to the Head of AI, such as when risks exceed organizational thresholds or when decisions affect multiple systems.

B AI system

AI systems are information systems that generate outputs, such as predictions, recommendations, decisions, or content, for a given set of human-defined objectives (see ISO/IEC 22989:2022 Artificial intelligence concepts and terminology). The AI system focus area entails documentation, performance metrics, and approval processes for AI systems as well as establishing and maintaining an organization-wide AI system registry providing a complete view of the organizations' AI systems.

1 AI system registry and AI ID

Organizations are responsible for maintaining appropriate awareness and oversight of AI systems they develop, operate, or use, regardless of whether these systems are developed in-house or procured from external vendors.

The Head of AI should ensure that an AI system registry is established and that appropriate governance processes are defined.

The registry should

1) identify all AI systems the organization develops, uses, or has retired,
2) assign them a unique identifier,
3) identify the organizational owner accountable for the AI system, and
4) contain the relevant documentation on the AI system.

2 AI system version control

An AI system will likely undergo several redesigns and update cycles during its lifecycle. Designing and implementing an effective version control system is an important element of the AI governance framework.

The AI System Owner should ensure that the organization defines, documents, and establishes

1) version control and approval workflows and
2) appropriate quality control processes for new versions and updates.

The AI System Owner should ensure that the AI system version control design aligns with the organization's objectives and risk tolerance.

The AI System Owner should ensure that the organization implements the planned AI system version control processes.

If the AI system version control processes indicate non-conformance with version control practices or a misalignment with organizational objectives or risk tolerance, the AI System Owner should initiate appropriate measures to address the non-conformance or restore alignment.

3 AI system planning and scoping

Once an organization initiates an AI system development project, it should perform a preliminary **pre-design** of the system. The AI System Owner should ensure that the organization

1) enters the AI system into the AI system registry (B1),
2) assesses whether the AI system can align with the organization's objectives and risk tolerance,
3) initiates the development processes and assigns roles and responsibilities, and
4) identifies and documents the planned AI system's key features and design constraints.

Identifying and understanding the intended **use case** of an AI system and its other possible uses is key to sustainable AI development and use. The use case affects the system's regulatory environment and may have significant reputational risk implications.

The AI System Owner should ensure that the organization defines and documents

1) the intended use case of the AI system and
2) the possible other uses of the AI system.

The AI System Owner should ensure that the use case definition aligns with the organization's objectives and risk tolerance.

The AI System Owner should ensure that the organization takes adequate measures to prevent inappropriate AI system use.

People in organizations use AI systems. Some AI systems make decisions that directly or indirectly affect humans and their rights and obligations (affected persons).

The organization should define and document

1) the intended AI system user organizations and human users,
2) the intended affected persons, and
3) possible other users and affected persons.

The AI System Owner (A2) should ensure that the AI system user definitions align with the organization's objectives and risk tolerance.

4 AI system environment and architecture

AI systems are embedded in the business and organizational environment. This environment typically consists of technological and social elements. The operating environment is a key driver of AI system impacts.

The organization should define and document

1) the intended business or operational model and environment of the AI system,
2) the intended IT environment the AI system is embedded in and interacts with, and
3) the other intended AI systems and other information systems the AI system interacts with.

AI systems are typically part of information systems. Information systems contain various data, computing infrastructures and resources, system architecture, and interfaces.

The organization should

1) define and document the position of the AI system in the organization's IT architecture, and
2) document and manage the AI system's interactions with the organization's other IT systems.

5 AI system metrics

Establishing appropriate performance metrics is a key means to ensure the AI system performs in a desired way.

The AI System Owner should ensure that the organization defines and documents pre-deployment performance metrics that the AI system must meet before system deployment or subsequent update releases.

To evaluate the AI system's performance against the organization's objectives and risk tolerance, the organization needs to establish appropriate performance metrics.

The AI System Owner should ensure that the organization defines and documents the performance metrics for assessing AI system performance.

The AI System Owner should ensure that the performance metrics align with the organization's objectives and risk tolerance.

6 AI system verification and validation

Verifying and validating AI system performance is a crucial aspect of AI system development and quality control.

In AI systems with machine learning components, verification will require thorough validation testing and both theoretical and analytical verification. In many cases, validation necessitates that the developing organization constructs a simulation environment to explore model performance using extensive samples of real-world, non-training data inputs.

The AI System Owner should ensure that the organization develops appropriate verification and validation methods for adequate AI system performance.

7 AI system approval

Decisions on approving AI systems and AI system versions to be taken into operational use should be informed and preceded by a careful review of the AI system documentation.

Before deciding to deploy an AI system or AI system version, the AI System Owner should review all documentation on the AI system and ensure that the AI system impacts are acceptable and the system meets the performance targets for deployment.

8 AI system performance monitoring

Monitoring AI system performance is crucial to ensuring that the system sustains the desired level of performance. The monitoring must be systematic and metrics-based to achieve consistency over time.

The AI System Owner should ensure that the organization defines, documents, and establishes

1) workflows and technical interfaces to facilitate the monitoring of AI system performance, including, for example
    a) the automated or manual production and reporting of performance metrics data,
    b) alarm thresholds, and
    c) workflows that allocate monitoring responsibilities.
2) workflows and technical interfaces to facilitate the detection of system malfunctions and other anomalous events, and
3) workflows to address issues detected during regular monitoring and health checks.

The AI System Owner should ensure that the AI system performance monitoring process aligns with the organization's objectives and risk tolerance.

The AI System Owner should ensure that the organization implements the planned AI system performance monitoring processes.

If the AI system performance monitoring processes indicate non-conformance with performance standards or a misalignment with organizational objectives or risk tolerance, the AI System Owner should initiate appropriate measures to address the non-conformance or restore alignment.

9 AI system health checks

AI systems may be subject to performance deterioration over the medium and long term. In addition, the business, operational, IT, and regulatory environments and stakeholder pressures will change over time. These processes may jeopardize system performance or lead to unacceptable risks.

The AI System Owner should ensure that the organization conducts regular comprehensive reviews of the AI system (health checks) to ensure that the AI system aligns with the organization's objectives and risk tolerance.

The AI System Owner should ensure that the organization defines, documents, and establishes workflows and technical interfaces for reviewing

1) the AI system use case,
2) the AI system users,
3) the AI system operational environment,
4) the AI system technical environment,
5) the AI system metrics,
6) the AI system version control practices,
7) the AI system monitoring practices, and
8) the AI system health check practices.

The AI System Owner should ensure that the organization performs the regular planned health checks.

The health checks should assess whether the AI system aligns with the organization's objectives and risk tolerance. If a health check discloses a misalignment, the AI System Owner should initiate appropriate measures to restore alignment.

C AI model

AI models are mathematical constructs that generate inferences or predictions based on input data (see ISO/IEC 22989:2022 Artificial intelligence concepts and terminology). An AI system (see Focus Area B) contains one or more AI models. The AI model focus area entails documentation, performance metrics, and approval processes for AI models as well as establishing and maintaining an organization-wide AI model registry.

1 Model registry and model ID

Organizations are responsible for maintaining appropriate awareness and oversight of the AI models they develop, operate, or use, regardless of whether these models are developed in-house or procured from external vendors.

The Head of AI should ensure that an AI model registry is established and that appropriate governance processes are defined.

The model registry should

1) identify, to the extent possible, all models the organization develops, uses, or has retired,
2) assign them a unique identifier, and
3) contain the relevant documentation on the model.

2 Model version control

AI models will likely undergo several redesigns and update cycles during their lifetime. Some models may learn continually. Designing and implementing an effective version control system is an important part of the AI governance framework.

The AI System Owner should ensure that the organization defines, documents, and establishes

1) version control and approval workflows and
2) appropriate quality control processes for new versions and updates.

The AI System Owner should ensure that the AI model version control design and practices align with the organization's objectives and risk tolerance.

The AI System Owner should ensure that the organization implements the planned model version control processes.

If the model version control processes indicate non-conformance with version control practices or a misalignment with organizational objectives or risk tolerance, the AI System Owner should initiate appropriate measures to address the non-conformance or restore alignment.

3 Model planning and scoping

Once an organization initiates a model development project, it should perform a preliminary pre-design of the model. The AI System Owner should ensure that the organization

1) enters the model into the model registry (C1),
2) assesses whether the model can align with the organization's objectives and risk tolerance,
3) initiates the development processes and assigns roles and responsibilities, and
4) identifies and documents the key features and design constraints for the planned model.

Understanding the intended uses of a model, together with its possible misuses, is key to sustainable AI development and use.

For each model in its model registry, the organization should define and document, to the extent possible,

1) the intended uses of the model and
2) the possible foreseeable unintended uses and misuses of the model, if relevant.

The use case definition should guide the development processes and build on the risk and impact pre-design and assessment outcomes.

The AI System Owner should ensure that the intended use case aligns with the organization's objectives and risk tolerance. The AI System Owner should ensure that the organization takes adequate measures to prevent inappropriate model use.

4 Model technical environment design

When operational, models are typically part of AI systems. The AI system architecture and its connections to the organization's other IT systems affect the AI system's impacts.

The organization should

1) define and document the position of the model in the AI systems it is a part of, and
2) document and manage interactions with the organization's other AI systems and IT systems.

The AI System Owner should ensure that the model's technical environment aligns with the organization's objectives and risk tolerance.

5 Model metrics

The organization can only ensure desired model performance by designing appropriate metrics to evaluate it.

The AI System Owner should ensure that the organization defines and documents **pre-deployment performance metrics** that the model must meet prior to deployment or updates.

The AI System Owner should ensure that the organization defines and documents **operational performance metrics** for assessing model performance during operational use.

The AI System Owner should ensure that the performance metrics align with the organization's objectives and risk tolerance.

6 Model verification and validation

Verifying and validating model performance is a crucial aspect of AI system development and quality control.

In AI systems with machine learning components, verification will require comprehensive validation testing in addition to theoretical and analytical verification. In many cases, validation will require that the developer organization builds a simulation environment where it can explore model performance using comprehensive samples of real-world, non-training data inputs. Further, validation may require developing post hoc interpretability tools to gain insight into model logic.

The AI System Owner should ensure that the organization develops appropriate verification and validation methods to ensure adequate model performance.

7 Model approval

Decisions approving models and model versions to be taken into operational use should be informed and preceded by an appropriate review of the model.

Prior to deciding to deploy a model or a model version, the AI System Owner should review all documentation on the model and the associated AI system and ensure that the model meets the performance targets for deployment.

At times, the organization may have limited access to the models in its AI system. In these cases, the approval process should review all available documentation and make a decision on whether deploying the model creates risks that exceed the organization's risk tolerance or breach its legal obligations.

8 Model performance monitoring

Monitoring model performance is crucial to ensure that the organization sustains the desired level of operational performance. The monitoring must be systematic and metrics-based to achieve consistency over time.

The AI System Owner should ensure that the organization defines, documents, and establishes

1) workflows and technical interfaces to facilitate the monitoring of model performance, including, for example,
    a) automated or manual production and reporting of performance metrics data,
    b) alarm thresholds, and
    c) workflows that allocate monitoring responsibilities.
2) workflows to address issues detected during regular monitoring and health checks.

The AI System Owner should ensure that the model performance monitoring design process aligns with the organization’s objectives and risk tolerance.

The AI System Owner should ensure that the organization implements the planned model performance monitoring processes.

If the model performance monitoring processes indicate non-conformance with performance standards or a misalignment with organizational objectives or risk tolerance, the AI System Owner should initiate appropriate measures to address the non-conformance or restore alignment.

9 Model health checks

Models may be subject to performance deterioration over the medium and long term. In addition, the business, operational, IT, and regulatory environments and stakeholder pressures will change over time. These processes may jeopardize model performance or lead to the emergence of unacceptable risks.

The AI System Owner should ensure that the organization conducts regular comprehensive reviews of the model (model health checks) to ensure that the model aligns with the organization’s objectives and risk tolerance.

The AI System Owner should ensure that the organization defines, documents, and establishes workflows and technical interfaces to review

1) the model use case,
2) the model users,
3) the model operational environment,
4) the model technical environment,
5) the model deployment metrics,
6) the model operational use metrics,
7) the model version control practices,
8) the model performance monitoring practices, and
9) the model health check practices.

The AI System Owner should ensure that the organization performs the regular planned health checks.

The health checks should assess whether the model aligns with the organization’s objectives and risk tolerance. If a health check indicates a misalignment, the AI System Owner should initiate appropriate measures to restore alignment.

D AI data

The AI data focus area entails data quality management, data sourcing, data quality metrics, and control points.

1 Data source documentation

The AI System Owner should ensure that the organization defines and documents AI system data sources. The AI System Owner should ensure that the organization has the right to process the data.

The AI System Owner should ensure that the organization defines and documents training, validation, and operational data sources. The AI System Owner should ensure that the organization has the right to process the model training, validation, and operational data.

2 Data preprocessing

Training, validation, and operational data undergo preprocessing in many AI systems.

The AI System Owner should ensure that the organization designs and implements appropriate workflows and technical interfaces for effective and appropriate data preprocessing.

As training data, validation data, and operational data often differ qualitatively, the AI System Owner should ensure that the organization understands the differences and designs and implements appropriate workflows and interfaces for preprocessing each data category.

The AI System Owner should ensure that the data preprocessing process aligns with the organization's objectives and risk tolerance.

3 Data quality metrics

The organization can only ensure desired AI system performance by designing appropriate metrics to evaluate data quality.

The AI System Owner should ensure that the organization defines and documents data quality metrics for assessing the quality of data used by the AI system.

The AI System Owner should ensure that the data quality metrics align with the organization's objectives and risk tolerance.

4 Data ontologies, inferences, and proxies

Data resources contain various categories of data. The categories reflect explicit or implicit data ontologies. Data ontologies consist of entity taxonomies (what entities are assumed to exist) and models of entity-relationalities and causality (how the entities relate to each other). Data ontologies may have significant implications on how models and AI systems function and what risks they create and to whom.

The AI System Owner should ensure that the organization adequately understands

1) the AI system and model data ontology (the entities and categories in the data),
2) the risks related to possible inconclusive evidence, system bias, and discrimination risk,
3) what inferences are drawn from the data and what proxies (indirect measures for unobservable variables) are created when the organization uses a machine learning approach to develop a model.

The AI System Owner should ensure that the organization develops and implements measures to minimize and mitigate possible data-related risks.

In particular, if the AI system makes decisions that affect natural persons, the AI System Owner should ensure that the organization conducts an appropriate assessment of the AI system's discrimination and misidentification risks (including false positives, false negatives, and incorrect categorization of individuals).

The AI System Owner should ensure that the residual risks are acceptable and align with the organization's objectives and risk tolerance.

5 Data quality assurance

Adequate data quality is a crucial precondition to all AI system operations.

The AI System Owner should ensure that the organization designs and establishes appropriate workflows and technical arrangements for

1) gathering and producing information on data quality, and
2) ensuring that the data (including the training, validation, and testing data) is of adequate quality and sufficiently relevant, complete, and representative.

Data quality analyses should also include an analysis of additional data needs.

6 Data quality monitoring

Monitoring the data quality of an AI system is crucial in ensuring it maintains the desired level of operational performance. Data quality monitoring must be systematic and metrics-based to achieve consistency over time.

The AI System Owner should ensure that the organization defines, documents, and establishes workflows and technical interfaces to facilitate the monitoring of data quality. In particular, the workflows should enable the identification of anomalous data entries and data drift, and should include

1) automated or manual production and reporting of data quality indicators and alarm thresholds,
2) workflows that allocate monitoring responsibilities, and
3) workflows to address issues detected during regular monitoring and health checks.

The AI System Owner should ensure that the data quality monitoring design process aligns with the organization's objectives and risk tolerance.

The AI System Owner should ensure that the organization implements the planned data quality monitoring processes.

If the data quality monitoring processes indicate non-conformance with data quality standards, data drift, or a misalignment with organizational objectives or risk tolerance, the AI System Owner should initiate appropriate measures to address the non-conformance or restore alignment.

7 Data health checks

Data resources may be subject to deterioration over the medium and long term. In addition, the business, operational, IT, and regulatory environments and stakeholder pressures will change over time. These processes may jeopardize data access or data quality and lead to unacceptable risks.

The AI System Owner should ensure that the organization designs processes for regular, comprehensive reviews of the AI system's data resources (data health checks) to ensure that the data-related risks are acceptable and align with the organization's objectives and risk tolerance.

The AI System Owner should ensure that the organization defines, documents, and establishes workflows and technical interfaces to review

1) AI system and model data sources (D1),
2) data preprocessing practices (D2),
3) data quality (D3, D5, and D6), and
4) data ontology, inferences, and proxies (D4).

The AI System Owner should ensure that the organization performs the regular planned health checks.

The data health checks should assess whether the AI system's data resources and data-related processes align with the organization's objectives and risk tolerance. If a health check indicates a misalignment, the AI System Owner should initiate appropriate measures to restore alignment.

E Risk and impact management

The risk and impact management focus area entails identifying, managing, and monitoring potential risks and impacts caused by the AI system. Risks may be related to, e.g., discrimination, violations of fundamental rights, and safety.

1 AI system harms and impacts pre-assessment

The intensity of potential harms and impacts will vary significantly across different AI systems. Industrial AI systems with no direct effects on individuals or the environment will likely create limited harms and impacts. An AI system that makes irreversible decisions that affect the rights and obligations of individuals will have profound impacts and may generate significant harms.

The AI System Owner should ensure that the organization conducts a harms and impacts pre-assessment at the outset of AI system development. The pre-assessment outcomes should be documented.

The pre-assessment should cover a wide range of potential harm and impact creation pathways. The pre-assessment should consider the harms the AI system may create and the impact it may have on its users, possible decision-making subjects, other affected parties, society at large, and the environment.

AI system risk assessments often focus on the direct harms the systems may create. The harms and impacts pre-assessment should, however, aim at also identifying the potential system-level harms and impacts. These include the social action affordances the system may create or modify, its potential wealth and power distribution implications, and its effects on equality.

The ethical advisory function should be involved in the pre-assessment of the harms and impacts if it is likely that the AI system will create a non-negligible risk of harm to individuals or the environment.

The AI System Owner should ensure that a harms and impacts pre-assessment is conducted when the design parameters of the AI system undergo significant changes. The organization should define thresholds for changes that trigger a harms and impacts pre-assessment.

2 Model risk assessment

AI systems consist of AI models. Hence, AI system performance is driven by model performance. Possible AI system biases and unfair outcomes often emanate from model design. If the organization has access to the models in the AI systems, identifying the possible model risks and assessing their gravity is key to sustainable AI system development and operation.

Model risk assessment should cover, to the extent possible, a wide range of model-related risk sources and causes.

The AI System Owner should at least ensure that the organization

1) explores and documents how the model affects the operations of the entire AI system,
2) explores and documents the possible risk of biased and, in particular, discriminatory outcomes, and
3) explores and documents the risk of unfair outcomes and harms the model may generate.

As identifying biases and unfairness is often complex and contentious, the reviews should involve ethical and legal experts, particularly if the organization intends to use the model in a

high-risk use case. In machine learning models, testing model outputs may be necessary for identifying biases and discriminatory outcomes.

In addition, if a machine learning model incorporates inferences made from training data, the risk assessment should review and assess

1) the risk of detecting non-existent patterns and correlations in the data,
2) the degree of model inscrutability and explainability.

3 AI system impact metrics design

Acceptable AI system impacts can only be ensured by deploying appropriate metrics to measure them.

The AI System Owner should ensure that the organization defines and documents metrics for monitoring the AI system's impacts during its operational use.

The AI System Owner should ensure that the impact metrics align with the organization's objectives and risk tolerance.

4 AI system impact assessment

The impacts of AI systems on human health and safety are key concerns that organizations should address when developing and using AI. Alongside these impacts, the effects on fundamental rights are likely to be the focal points of future regulatory instruments.

If the AI system harms and impacts pre-assessment (E1) indicates that the AI system will likely have non-negligible impacts on the health, safety, and fundamental rights of individuals, the AI System Owner should ensure that the organization undertakes and documents

1) a health impact review to identify the potential health impacts the AI system may have on the physical and psychological well-being of its users, subjects, and other affected parties,
2) a safety impact review to identify the potential safety risks the AI system may impose on individuals' and organizations' tangible assets, and
3) a fundamental rights impact review to identify the potential impacts that the AI system may have on the protection and realization of individuals' fundamental rights. The legal advisory function should be involved in the assessment.

5 AI system non-discrimination assurance

Many jurisdictions have non-discrimination laws and impose equal treatment requirements. Non-compliance with the non-discrimination laws and equal treatment requirements is incompatible with sustainable AI operations and may create significant legal and reputational risks.

The AI System Owner should ensure that the organization conducts and documents a non-discrimination assurance process to ensure that the AI system outputs are compliant with non-discrimination laws and equal treatment requirements. The legal advisory function should be involved in both designing and conducting the assurance.

Ensuring that an AI system creates no discrimination risk is challenging due to the nature of non-discrimination and equal treatment. According to non-discrimination legislation,[1] an AI system could directly discriminate against a person if the system treated a person less favorably than another based on their age, nationality, language, religion, belief, opinion, political activity, trade union activity, family relationships, state of health, disability, sexual orientation, or other personal characteristics. Less favorable treatment is discrimination even if based on an apparently neutral rule. Despite the prima facie ban, differential treatment can be

[1] See, e.g., the Finnish Non-Discrimination Act.

justified if mandated by law or the treatment has an acceptable objective in terms of basic and human rights, and the measures to attain the aim are proportionate.

Conducting a diligent non-discrimination assurance is particularly important for AI systems with models developed using machine learning approaches. Machine learning approaches may result in inadvertent discrimination. As the models are often unexplainable, detecting discriminatory bias may require the use of post-hoc analysis tools and real-world data AI system output testing.

6 AI system negative impact minimization

Minimizing the AI system's negative impacts is an important part of the AI governance framework. Minimizing the negative impacts requires first that the potential impacts are analyzed and appropriate measures are taken to eliminate or reduce adverse impacts where possible. Second, minimization requires that the organization mitigates the effects of the adverse impacts that it cannot eliminate, or, third, manages their consequences.

To arrive at an acceptable AI system impact, the AI System Owner should ensure that the organization

1) conducts a thorough analysis of potential impacts the system may have on its users, subjects, or affected parties, or the environment, and
2) develops and implements a risk minimization plan.

The risk minimization plan should be designed to guarantee that the AI system's impacts are acceptable and aligned with the organization's objectives and risk tolerance.

The risk minimization plan should outline

1) appropriate measures to eliminate adverse impacts to the extent possible,
2) appropriate measures to reduce adverse impacts that cannot be eliminated,
3) appropriate measures to mitigate the effects of the residual adverse impacts, and
4) appropriate measures to manage the adverse impacts that cannot be mitigated.

7 AI system impact monitoring

Monitoring the AI system's impacts is crucial to ensuring that they remain acceptable. The monitoring must be systematic and metrics-based to achieve consistency over time.

The AI System Owner should ensure that the organization defines, documents, and establishes

1) workflows and technical interfaces to facilitate the monitoring of AI system impact, including, for example
   a) automated or manual production and reporting of impact metrics data,
   b) alarm thresholds, and
   c) workflows that allocate monitoring responsibilities, and
2) workflows to address issues detected during regular monitoring and health checks.

The AI System Owner should ensure that the AI system impact monitoring process aligns with the organization's objectives and risk tolerance.

The AI System Owner should ensure that the organization implements the planned AI system impact monitoring processes.

If the AI system impact monitoring processes indicate non-conformance with impact standards or a misalignment with organizational objectives or risk tolerance, the AI System Owner should initiate appropriate measures to address the non-conformance or restore alignment.

8 AI system impact health checks

AI systems may create changing impacts over their operational lifetime. The AI System Owner should ensure that the organization conducts regular comprehensive reviews of the AI system impacts (impact health checks) to ensure that the system's effects on users, subjects, affected parties, and the environment remain acceptable.

The AI System Owner should ensure that the organization defines, documents, and establishes workflows and technical interfaces to review

1) whether the impact metrics (E3) remain relevant and complete,
2) whether new types of impacts have emerged that were not identified in the initial assessment,
3) whether the negative impact minimization plan (E6) remains effective, and
4) whether impact monitoring practices (E7) capture all significant impacts.

The AI System Owner should ensure that the organization performs the regular planned impact health checks.

The health checks should assess whether the impacts of the AI system align with the organization's objectives and risk tolerance. If a health check indicates a misalignment, the AI System Owner should initiate appropriate measures to restore alignment.

F Transparency, explainability, and contestability (TEC)

The transparency, explainability, and contestability focus area entails ensuring AI systems' transparency, explainability, and contestability as well as fulfilling respective requirements coming from regulation and stakeholders.

1 TEC expectation assessment

AI systems can have various stakeholders with varying needs and requirements regarding transparency, explainability, and contestability.

Some of the expectations may be regulatory in origin. GDPR, for example, imposes a right to explanation for automated decisions.

Internal and external stakeholders also impose TEC requirements on the system. For example, recommendation engine users may require information on the system logic to be able to assess the accuracy and appropriateness of the recommendations.

The AI System Owner should ensure that the organization conducts a TEC expectation assessment at the outset of the development process. In the assessment, the organization should identify relevant TEC stakeholders and TEC expectations.

Identifying regulatory transparency and explainability expectations typically requires input from legal experts and is a part of the compliance process. Identifying non-regulatory TEC stakeholder expectations may benefit from stakeholder consultations or co-creation efforts.

2 TEC design

After identifying the appropriate transparency, explainability, and contestability expectations, the AI System Owner should ensure that the organization designs appropriate technical and organizational structures to satisfy the TEC expectations and align the AI system with the organization's objectives and risk tolerance.

The TEC design should address how a sufficient level of transparency, explainability, and contestability is secured by

1) model and AI system design,
2) technical interfaces and other arrangements that allow the end-users and others affected to gain adequate visibility into the AI system,
3) procedures, technical systems, technical interfaces, and other arrangements that allow decision-making subjects to contest the decisions in an appropriate manner and to an appropriate degree, and
4) user instructions and end-user transparency and explainability documents such as explanations of decision-making logic.

3 TEC validation and assurance

The AI System Owner should ensure that the TEC structures and processes function as designed. Explainability outputs should be verified for accuracy and comprehensibility to the relevant stakeholders. Contestability mechanisms should be evaluated to ensure they provide recourse in an appropriate manner and to an appropriate degree.

4 TEC monitoring

Monitoring AI system TEC performance is crucial to ensuring that the organization sustains the desired level of TEC performance. The monitoring must be systematic and metrics-based to achieve consistency over time.

The AI System Owner should ensure that the organization defines, documents, and establishes

1) workflows and technical interfaces to facilitate the monitoring of AI system TEC performance, including, for example,
   a) automated or manual production and reporting of performance metrics data,
   b) alarm thresholds, and
   c) workflows that allocate monitoring responsibilities.
2) workflows to address issues detected during regular monitoring and health checks.

The AI System Owner should ensure that the AI system's TEC performance monitoring process aligns with the organization's objectives and risk tolerance.

The AI System Owner should ensure that the organization implements the planned TEC monitoring processes.

If the TEC monitoring processes indicate non-conformance with TEC standards or a misalignment with organizational objectives or risk tolerance, the AI System Owner should initiate appropriate measures to address the non-conformance or restore alignment.

5 TEC health checks

The AI System Owner should ensure that the organization conducts regular comprehensive reviews of the AI system's transparency, explainability, and contestability (TEC health checks) to ensure that these align with the organization's objectives and risk tolerance.

The AI System Owner should ensure that the organization defines, documents, and establishes workflows and technical interfaces to review how a sufficient level of transparency, explainability, and contestability is secured by

1) model and AI system design,
2) technical interfaces and other arrangements that allow the end-users and others affected to gain adequate visibility into the AI system,
3) procedures, technical systems, technical interfaces, and other arrangements that allow decision-making subjects to contest the decisions in an appropriate manner and to an appropriate degree, and
4) user instructions and end-user transparency and explainability documents such as explanations of decision-making logic.

The AI System Owner should ensure that the organization implements the planned TEC health check measures.

If the health check discloses non-conformance with TEC performance standards or a misalignment with organizational objectives or risk tolerance, the AI System Owner should initiate appropriate measures to address the non-conformance or restore alignment.

G Regulatory compliance

The regulatory compliance focus area entails understanding the regulatory environment of an AI system and ensuring its compliance with the relevant regulations (in the EU, for example, the GDPR, AI Act, national regulations, and sector-specific regulations).

1 Regulatory environment assessment

AI systems are typically subject to a variety of regulatory instruments that may force particular design choices, constrain functionalities, or, in extreme cases, make implementing a specific design or use case impossible. Understanding the regulatory environment is, consequently, important to prevent misplaced investments.

To develop a preliminary understanding of the AI system's regulatory environment, the AI System Owner should ensure that the organization conducts a regulatory environment assessment. The regulatory environment assessment provides the organization with basic information on the regulatory environment in which the AI system will be used.

The assessment should ensure that laws and regulations that may affect the AI system are identified and reviewed. The assessment should develop a knowledge base of the contents of the primary regulatory instruments and key constraints that could affect AI system design and operations. All parties active in developing or implementing an AI system within the organization should be aware of the findings of the assessment process.

2 Regulatory risks, constraints, and design parameter analysis

Regulation may impose critical constraints and requirements on AI system design.

Once a tentative understanding of the future intended use case and users of the AI system is reached, the AI System Owner should ensure that the organization conducts an appropriate analysis of the system and its regulatory environment to identify key regulatory risks, constraints, and design parameters.

The analysis should

1) assess regulatory risks associated with known design options,
2) identify key design constraints,
3) identify design areas with significant regulatory implications (key regulatory issues), and
4) outline possible design options and their implications.

These regulatory focal points should be clearly communicated to all parties active in developing or implementing an AI system within the organization.

3 Regulatory design consultation

During development, regulatory requirements may significantly impact design decisions. Development investments may be lost if the AI system has non-compliant features. To ensure efficient resource allocation, the legal function should be consulted prior to making design decisions that could affect the key regulatory focal points.

The AI System Owner should ensure that developers consult the legal function before making design decisions that could affect the key regulatory issues (G2).

4 Compliance assessment

Prior to deployment, a compliance assessment is required to verify that the AI system meets all applicable regulatory requirements. The AI System Owner should ensure that the legal function conducts and documents a compliance assessment before the AI system is approved to be taken into operational use or before deploying materially new versions. The AI System

Owner should review the compliance assessment results and ensure any identified issues are resolved before AI system approval (B7).

5 Compliance monitoring

The regulatory environment will likely change during the AI system's lifetime. In addition, changes made to the AI system may also affect its compliance.

The AI System Owner should ensure that the organization maintains an awareness of possible regulatory changes relevant to the AI system.

The AI System Owner should ensure that the organization designs and establishes appropriate workflows and technical interfaces to facilitate compliance monitoring.

The AI System Owner should also ensure that the organization implements the planned compliance monitoring processes.

If the compliance monitoring processes disclose a non-compliance event, the AI System Owner should initiate appropriate measures to address the non-compliance or restore alignment.

6 Compliance health checks

The AI System Owner should ensure that the organization conducts regular, comprehensive health checks of AI system compliance. The AI System Owner should ensure that the organization designs and establishes appropriate workflows and technical interfaces to facilitate periodic compliance health checks.

The health checks should assess whether the AI system's compliance processes align with the organization's objectives and risk tolerance. If a health check discloses a misalignment, the AI System Owner should initiate appropriate measures to restore alignment.