Nordic Ethical AI Guidebook

Insights from the Nordic Ethical AI Sandbox

Nordic Innovation

Executive Summary

This guidebook builds on insights generated through the Nordic Ethical AI Sandbox. This initiative is part of larger project, *Nordic Ethical AI & Data Ecosystem Building*, led by Accenture and Silo AI under the AI & Data program at Nordic Innovation.

The sandbox consisted of facilitated industry workshops and open webinars, aimed at generating and transferring knowledge about best practices for implementing ethical and responsible AI practices. This guidebook has been written by Accenture and Silo AI on behalf of Nordic Innovation.

The contents of this guidebook are aimed at Nordic businesses primarily and are meant to provide guidance and inspiration to individual organization's implementation of ethical and responsible AI. 25 organizations across the Nordics have contributed to this guidebook, through their participation in the sandbox events.

The guidebook consists of four sections, each summarizing an industry workshop and webinar held on the same topic:

- Organizational Approach to Ethical and Responsible AI
- Risks and Safeguards for Generative AI Systems
- Robust and Secure AI Systems
- Transparent and Explainable AI Systems

Toolkit

Based on the insights presented in the guidebook, a toolkit has been developed to support organizations with accelerating their ethical and responsible AI journey. The toolkit consists of methodologies and exercises used in the sandbox workshop series and are meant to be leveraged by individual organizations.

Table of Contents

01.
Introduction

Building a thriving ethical AI ecosystem

The Nordics is well positioned to build a thriving ethical AI ecosystem, enabling the region to become leading in ethical AI and responsible use of data

COMMON VALUES

Nordic countries share similar history, values, regulatory systems, and business approaches. The Nordics also has a common goal to become the most sustainable and integrated region in the world (Nordic Innovation). Enabling a cross-border ecosystem can enable businesses in the Nordics to expand into new markets quicker and draw on the experience of their peers.

TRANSPARENT & COLLABORATIVE

The Nordic countries value transparency and collaboration. Our democratic systems rely on transparency as a core principle, and there are many examples of high-degree of transparency of data and information regarding individuals, for example personal security numbers, population registration, and income tax. Nordic businesses can build on this tradition as they increase adoption of AI, and create a competetive advantage in AI by emphasizing on transparency.

INDUSTRY LEADERSHIP

The Nordic business landscape boosts of some of the world's leading companies in their industry, that have existed for several decades. Naturally, these organizations possess deep expertise about their respective industry. When AI is integrated in their core business processes, having domain expertise in the organization will be critical to validate AI outputs, and ensure effective human oversight which is central to ethical and responsible AI.

AI MATURITY & COMMON CHALLENGES

Digital and AI maturity in the Nordic countries is somewhat equal. Norway, Sweden, Finland and Demark are all ranked among the top 25 globally for investments, innovation, and implementation of AI, only Iceland ranks slightly lower at number 40 (Tortoise Media Global AI Index 2023). Being at the same maturity level means that businesses will face common challenges once they attempt to scale AI. Enabling collaboration and knowledge sharing between these businesses will be beneficial for all.

THRIVING START-UP NETWORK

The Nordics boasts a thriving start-up network with innovative AI products and responsible AI-specific solutions. Creating a joint ecosystem can help increased awareness of and access to Nordic providers of tools, services or platforms that enable operationalization of ethical and responsible AI. The Nordics are also the most impact-focused region globally with 38% of investments going to impact startups compared to 22% for Europe (Dealroom 2023).

Why are Nordic businesses focusing on ethical AI?

What are the main factors driving companies to adopt ethical and responsible AI?

- Compliance
- Trust
- Competitive Advantage
- Brand Reputation
- Not being able to do AI
- Efficiency
- Sustainability
- Corporate Responsibility
- Privacy & Security

What outcomes do Nordic businesses expect by adopting and responsible AI?

- Top Talent Attraction
- Competitive Advantage
- Increased Trust
- Increased control
- Positive Brand Reputation
- Better use cases
- Regulatory resilience
- Increased Value
- Robust processes

Source: Nordic Ethical AI Sandbox Workshop #1

02.

Organizational Approach to Ethical and Responsible AI

Chapter Introduction

Nordic perspectives on how to create and operationalize an **organizational approach for ethical and responsible AI**

Background

- This chapter is based on the **results of the first workshop** and webinar in the Nordic Ethical AI Sandbox, which took place during November-December 2023.

- **Ensuring that AI is developed and used in an ethical and responsible way**, necessitates organizational-wide implementation of a systematic approach.

- This chapter explores key building blocks of an organizational approach to ethical and responsible AI and includes suggested actions to mature each capability area. **Each organization needs to carefully consider their own unique context**, their role in the AI value chain, the regulatory landscape and AI risks when adopting an organizational approach to ethical and responsible AI. These recommendations should therefore be considered as a starting point, and need to be further detailed for each organization.

What this chapter will help organizations with

Understanding what the **key building blocks** are of an organizational approach to ethical and responsible AI

Understanding what **key actions** are needed to operationalize each building block

Understanding what **stakeholders** are normally involved in creating and implementing the organizational approach, and what their roles and responsibilities are

The building blocks of ethical and responsible AI

Ethical and responsible AI covers multiple layers of an organization and will therefore need involvement from stakeholders across the C-suite, risk, compliance, tech, data, procurement, and HR

Oversight & Control	*Defines the company governance model for AI (governance structure, principles, policies, risks), including roles and responsibilities*	Governance: AI Principles, Risks, Policies, Organizational Structures and Steering Mechanisms		
Risk Management	*Structured approach for how to evaluate and mitigate the risks of AI systems, connected into standard AI lifecycle processes*	Risk Assessment	Reporting & Escalation	
Good Practice	*Methodologies, tools and training that supports relevant teams with operationalizing the requirements defined per the governance model*	Best pratices & Methodologies	Tools	Training & Culture
Data Foundation	*Data capabilities that enables Responsible AI use-cases*	Data Quality	Data Lineage	Data Compliance

🟦 Organizational-level capability areas

Source: Nordic Ethical AI Sandbox Workshop #1. The workshop structure was based on Accenture’s Capability Framework for Responsible AI

Actions for oversight and control of AI

AI governance guides the design, development and deployment of AI across an organization

Governance: AI Principles, Risks, Policies , Organizational Structures and Steering Mechanisms

- **Identify regulatory frameworks** that will impact how the company and summarize the legal requirements imposed by these.
- **Define the scope for enterprise AI governance**, meaning what should be covered (for example legal, ethics and security).
- **Define what constitutes an 'AI'-system** (referencing industry and governmental standards) and create an inventory of all AI-systems developed and/or put on the market across the organization.
- **Create an accountability framework for AI** by defining what roles are involved in the governance of AI, and what their responsibilities are.
- **Define the company position on why commitment to ethical and responsible AI is important** and secure C-suite sponsorship.
- **Document responsible AI policies** and distribute to all relevant stakeholders within the organization to ensure that employees follow the guidelines and recommendations put in place.
- **Define the values that should drive ethical and responsible use of AI**. Responsible AI principles sets the company position internally and externally on AI use, defines the risk appetite and informs policies. Defining AI principles require analysis of existing company principles and policies, to avoid inconsistencies or duplications. These can for example include, the Code of Conduct, Code of Business Ethics, Corporate Governance, Privacy Policy, Procurement Guidelines, Diversity & Inclusion Policy.
- **Define rules for governance and compliance with the defined principles and any legal obligations**, these should guide employees' activities in relation to AI development and usage, e.g. human-in-the-loop requirements.
  - Consolidate both local and international policies that are part of the AI value chain.
- **Define risk categories that AI-systems can fall into and categorization logic**. These should be mutually exclusive, and collectively exhaustive. There can also be categories for prohibited systems. The categorization logic determines the criteria for classifying risk categories for individual AI systems.

Source: Nordic Ethical AI Sandbox Workshop #1. The workshop structure was based on Accenture's Capability Framework for Responsible AI

Actions for AI risk management

Adopting a risk-based approach to AI helps with building regulatory readiness and mitigating potential harms

Risk Assessment

- **Define a methodology for risk screening AI-systems**, meaning assessing and categorizing AI-systems by level of risk. This methodology should cover both already deployed AI-systems and to-be developed.
- **Create a methodology for assessing the impact of AI systems**, including identifying, assessing and measuring risks of AI-systems, and controlling alignment with regulations, principles and policies:
    - Define assessment questionnaire
    - Define scoring and aggregation methodology
- **Define documentation requirements and process** for identified impacts, risks and mitigation strategies.

Reporting & Escalation

- **Design the escalation paths in the organization**, meaning the organizational model for making critical decisions regarding risks and negative impacts of AI-systems:
    - Define who is involved in what step of the escalation chain
    - Define what issues, scenarios or topics gets escalated when
- **Define documentation requirements and process** for escalations.
- **Define how to perform internal, and potentially external, reporting on AI** development and usage related to the defined organizational AI principles and policies.

Source: Nordic Ethical AI Sandbox Workshop #1. The workshop structure was based on Accenture's Capability Framework for Responsible AI

Actions for ensuring good AI practices

AI practitioners across the organization require support from standardized practices, tools and training

Best practice & Methodologies

- **Define standardized methods for how to ensure alignment with any principles** or policies for responsible AI. For example, statistical methods for various responsible AI dimensions, such as, fairness, explainability, robustness, soundness, human-centered design and human-in-the-loop approaches, and transparency mechanisms.
- **Develop standardized instructions and templates for documentation** along the development process, including risk assessment and mitigation results, system design, testing results, monitoring and functioning.

Tools

- **Identify effective tools that supports relevant stakeholders** across functions in operationalizing principles or policies related to responsible AI that support with risk assessment, monitoring, governance, and AI solution evaluation.
  For example (non-exhaustive):
  - Governance platform to perform conformity assessment with regulations or policies
  - Bias detection tool
  - Automated record-keeping tracker of AI system activities to maintain a log of developments and changes.
  - Post-deployment monitoring tool for registering all input, output and supporting with overseeing control that requirements are met

Training & Culture

- **Upskill the organization in ethical and responsible AI** by creating learning paths and training material tailored for different roles and responsibilities:
  - C-suite and board training
  - Technical training for AI teams
  - Upskilling business functions on the responsible AI approach
  - Company wide education series on current and upcoming AI regulations
  - Company wide education series on the organization's governance structure, risk categories and everyone's role in the organization
- **Foster a responsible AI culture** by engaging the whole organization in viewing responsible AI as a critical business imperative.

Source: Nordic Ethical AI Sandbox Workshop #1. The workshop structure was based on Accenture's Capability Framework for Responsible AI

Actions for making data AI-ready

Operationalizing responsible AI is dependent on high-quality and traceable data and robust supporting infrastructure

Data Quality

- **Incorporate user feedback** to validate and improve data quality.
- **Use data that is relevant to the problem being solved,** and address bias in training data to prevent biased outcomes in AI models.

Data Lineage

- **Ensure data lineage** by mapping data sources, data processes, data transformations, and data assets.
- **Create documentation for how data is transformed** for final input of AI model.

Data Compliance

- **Implement proper data archiving, versioning, and deletion** policies.
- **Safeguard user privacy and secure sensitive data** by implementing strong data encryption, data anonymization, and access controls.
- **Document the data compliance processes and regularly review,** and assess the compliance level of the data process and AI models.

Source: Nordic Ethical AI Sandbox Workshop #1. The workshop structure was based on Accenture's Capability Framework for Responsible AI

Stakeholders involved (I/II)

Ethical and responsible AI covers multiple layers of an organization and will therefore need involvement from a wide variety of stakeholders. Each organization needs to define what functions and roles are involved in operating responsible AI

Non-exhaustive

C-suite (Leadership)

- **Budget:** Allocates financial resources for responsible AI initiatives.
- **Communication:** Communicates the importance of responsible AI to internal and external stakeholders.

Product Owners

- **Responsible AI integration:** Ensures that responsible AI principles and policies are integrated into product development, from design to deployment.

R&D (Research & Development)

- **Innovation with responsibility:** Focuses on developing AI solutions and technologies that align with the organizations principles and policies on responsible AI.

Data Owners

- **Data governance:** Ensures the quality, privacy, and security of data used in AI systems.
- **Data compliance:** Ensures data usage aligns with legal and ethical standards.

Legal

- **Regulatory compliance:** Interprets laws and regulations and defines the legal obligations put on the organization's AI operations
- **Risk management:** Identifies and manages legal risks associated with AI use.

Risk & Compliance

- **Risk management:** Integrates AI specific risk management in the organization's existing risk management practices.

Note that the roles and role descriptions provided here are examples and might not be applicable to all organizations. Each organization needs to identify their critical roles and respective responsibilities needed to operationalize ethical and responsible AI.

Source: Nordic Ethical AI Sandbox Workshop #1. The workshop structure was based on Accenture's Capability Framework for Responsible AI

Stakeholders involved (II/II)

Ethical and responsible AI covers multiple layers of an organization and will therefore need involvement from a wide variety of stakeholders. Each organization needs to define what functions and roles are involved in operating responsible AI

Non-exhaustive

Communication

- **Stakeholder engagement:** Communicates responsible AI practices internally and externally.
- **Public relations**: Manages the public image and perception of the company regarding its AI practices.

Procurement

- **Vendor responsibility:** Helps evaluate that third-party AI vendors adhere to ethical and responsible AI principles and policies set forth by the organization
- **Due diligence:** Conducts assessments to ensure ethical practices in the AI supply chain.

HR & Learning

- **Skill development:** Ensures that employees have the necessary knowledge and skills to operationalize the organization's principles and policies for ethical and responsible AI
- **Culture development:** Promotes a culture of ethical AI use within the organization.

Security

- **Security:** Ensures the security of AI systems to prevent malicious use and data breaches.
- **Threat mitigation:** Addresses potential risks and vulnerabilities associated with AI applications.

Internal Audit

- **Audit trail:** Ensure that all data and actions are documented and can be traced back to explain outcomes as well as mitigate compliance risks.
- **Internal Control:** Defines and evaluates internal controls created to ensure compliance

End-users & Civil Society

- **Feedback:** Provides feedback from the perspective of end-users and civil society on the organization's use of AI, to help validate and improve internal responsible AI approach.

Note that the roles and role descriptions provided here are examples and might not be applicable to all organizations. Each organization needs to identify their critical roles and respective responsibilities needed to operationalize ethical and responsible AI.

Source: Nordic Ethical AI Sandbox Workshop #1. The workshop structure was based on Accenture's Capability Framework for Responsible AI

03.
Risks and Safeguards for Generative AI Systems

Chapter Introduction

Nordic perspectives on **potential risks and safeguards when developing and using Generative AI**

Background

- This chapter is based on the **results of the second workshop** and webinar in the Nordic Ethical AI Sandbox, which took place during February 2024.
- Generative AI has become a top strategic priority for Nordic leaders and businesses, and many organizations are piloting and/or deploying the technology. The attention to **the capabilities of Generative AI has raised concerns about potential risks, and how to effectively mitigate or manage them**.
- This chapter explores **Nordic perspectives on potential risks with developing and adopting Generative AI solutions, and how to safeguard against them**. Each organization needs to carefully consider their own unique AI risk landscape. These recommendations should therefore be considered as a starting point, and need to be further detailed for each organization.

What this chapter will help organizations with

Insight into **Generative AI adoption** among selected Nordic organizations

Understanding **important factors when assessing and prioritizing** between Generative AI use-cases

Understanding **Generative AI risks and mitigation strategies** explored by selected Nordic organizations

Generative AI use-cases

Generative AI is being piloted and deployed by Nordic businesses where a common theme across workshop participants is that the initial use-case(s) are internal-facing, focused on increasing productivity

Examples of Generative AI Use-cases Explored

Non-exhaustive responses

Content Creation & Editing

- Text generation & summarizing (text-to-text)
- Image generation & editing (text-to-image)
- Video generation & editing (text-to-video)
- Speech synthesis (speech-to-text)

Information Retrieval

- Intelligent Search engine (internal & external data)
- Domain expert chatbots
- Summarize & assess internal information

R&D & Innovation

- *Security:* Network alarms & cyberattack defence
- *Sustainability:* Accelerate green transitions

Customer Experience

- Customer service chatbots
- Hyper personalization & customization

Data Science & Analysis

- Coding (generation & assessment)
- Analysis (sentiment analysis, text classification)

Source: Nordic Ethical AI Sandbox Workshop #2

Note: These are aggregated results from the workshop, and do not necessarily apply for all participating organizations

Generative AI tools & techniques

Nordic businesses are exploring different Generative AI tools and techniques. Using pre-built applications lowers barriers to adoption, but custom solutions are also built with different techniques or solution patterns.

Examples of Generative AI Tools & Techniques Explored

Non-exhaustive responses

Tools & Applications

Third-party APIs

External software interfaces for accessing off-the-shelf LLMs, for example OpenAI's API

GitHub Copilot

AI code completion tool by GitHub and OpenAI

Chat-GPT

Interface for accessing OpenAI's text-to-text models, for example GPT3.5, GPT-4

Google Duet AI

Google's Gen AI co-pilot, integrated with Google's various creative applications and platforms

O365 Copilot

Microsoft's Gen AI Co-pilot, integrated with Microsoft Office 365

Techniques & Solution Patterns

RAG

Retrieval-augmented generation model for improved performance in Gen AI solutions & NLP tasks

Multimodal

Processing and understanding data from multiple modalities, for example text, image, audio, simultaneously

Opensource LLMs

Publicly available large language models with accessible source code, for example Llama 2, BLOOM

Distributed and federated AI

Spreading AI tasks across multiple devices in a network, enabling AI to be independently trained through federated learning

Fine-tuned models

Pre-trained AI models that are trained on contextual data to achieve a higher performance for specific tasks

Improve LLMs' explainability

Enhancing the interpretability of transformer models to understand their decision-making

Edge deployment

Run on devices at the network edge, closer to where data is generated, to reduce latency and bandwidth usage

Source: Nordic Ethical AI Sandbox Workshop #2

Note: These are aggregated results from the workshop, and do not necessarily apply for all participating organizations

Use-case assessment framework

Based on Accenture’s Use-case Assessment Framework, insights were captured on what factors Nordic businesses consider when prioritizing between Generative AI use-cases

- This framework can be used as a starting point when assassing and prioritizing between different Generative AI (and traditional AI) use-cases. In addition to factors included in this framework, there might be other factors specific to each organization that should be considered.
- In the second Nordic Ethical AI Sandbox workshop, participants mapped different factors considered into three overarching categories: **1) Value 2) Effort & Feasibility, and 3) Risks & Ethics.**
- Identified assessment factors were also priotized to better understand which ones are more important than others according to the workshop group.

1	**Value** *Why we do it?*	*The benefits and value potential of deploying the use-case*
2	**Effort & feasibility** *Are we able to do it, and how?*	*The required technical capabilities, resources and costs to realize the use-case*
3	**Risks & ethics** *Can and/or should we do it?*	*The potential risks and ethical implications that could occur because of the use-case which could hinder value creation*

Source: Nordic Ethical AI Sandbox Workshop #2, based on Accenture’s Responsible AI Use Case Assessment Framework

Use-case assessment framework

Nordic companies identified different factors relevant for them when assessing and prioritizing Generative AI use-cases

1 Value

Why we do it?

- Efficiency gains
- Customer experience & communication
- Positive sustainability impact (Environmental, Social, Economical)
- Innovation & unlocked insights
- Business growth

The benefits and value potential of deploying the use-case

2 Effort & feasibility

Are we able to do it, and what is the cost?

- Technical feasibility and complexity
- Compliance burden
- Time to market
- Cost
- Skills and competences
- Change management and culture

The required technical capabilities, resources and costs to realize the use-case

3 Risks & ethics

Can/Should we do it?

- Unreliable output
- Workforce and talent risks
- Bias & harm
- Privacy and security
- Fast evolving technology
- Negative sustainability impact (Environmental, Social, Economical)
- Lack of understanding and control
- Liability and compliance

The potential risks and ethical implications that could occur because of the use-case which could hinder value creation

Source: Nordic Ethical AI Sandbox Workshop #2, based on Accenture's Responsible AI Use Case Assessment Framework

Note: These are aggregated results from the workshop, and does not necessarily apply for all participating organizations

Value considerations

Factors that contributes to valuable outcomes because of the use-case

Value

Why we do it?

Non-exhaustive

1

Efficiency gains

Increased operational efficiency and productivity that could be gained from for example automating repetitive tasks, streamlining processes, and using data-driven insights to optimize resource allocation and decision-making

Customer experience & communication

Improved customer experience and communication, enabled by for example personalized interactions and delivering timely and relevant content

Positive sustainability impact

Increased positive impact on society or the environment, enabled by for example facilitating development of more sustainable products and services and optimizing resource usage

Innovation & unlocked insights

Supporting innovation through generation of ideas, designs, and solutions, fostering creativity and augmenting human ingenuity across various domains

Business growth

Supporting generation of new revenue streams through product and service development, and optimizing revenue through tailored sales and marketing activities

Source: Nordic Ethical AI Sandbox Workshop #2, based on Accenture's Responsible AI Use Case Assessment Framework

Effort & feasibility considerations

Factors that could contribute to complexity and cost when realizing valuable outcomes of the use-case

2

Effort & feasibility

Are we able to do it, and what is the cost?

Non-exhaustive

Technical feasibility & complexity

The practicality and viability of implementing the Generative AI solution within existing technological infrastructures, considering factors such as data availability, data accessibility, solution lock-in etc.

Compliance capabilities

Technical and organizational factors that affect the capacity to adhere to regulatory requirements and adapt to changing legal frameworks while maintaining agility in adjusting to new guidelines, policies or constraints.

Time to market

The urgency of bringing the use-case from conception to deployment, requiring efficient deployment strategies and rapid iteration cycles, for example through AI prototyping tools or automating and validating tasks

Costs

The costs of developing and implement the solution, for example licensing costs, cloud storage costs, hardware costs and maintenance costs.

Skills & competences

The required knowledge and skills of the team involved in bringing the use-case from conception to deployment, including potential need to hire new employees or consult external experts

Change management & culture

The required change in existing processes, skills profiles or culture that could affect the ability to adopt and realize value from the use-case , for example lack of AI literacy or stakeholder buy-in

Source: Nordic Ethical AI Sandbox Workshop #2, based on Accenture's Responsible AI Use Case Assessment Framework

Risks & ethics considerations

Potential risks or ethical considerations that could have business, reputational or regulatory effects unless mitigated

3

Risks & ethics

Can and/or should we do it?

Non-exhaustive

Unreliable outputs

The risk that the outputs generated are thought to be correct, but instead may be false, misrepresenting or misleading, for example from hallucinations

Workforce & talent risks

The resulting impact on workforce and required skills profiles from utilizing Generative AI in the organization, for example from displacing certain roles or reducing resource needs

Bias & harm

The risk that the outputs generated express prejudice, toxicity or include harmful content of any kind due to bias in data, model or human review

Privacy & security risks

Uncertainty around protection of proprietary data and sensitive information used to train and prompt the model

Fast evolving technology

The uncertainty that comes with a fast evolving, less tested technology, making it difficult to fully foresee all potential risks with deploying it

Negative sustainability impact

Negative environmental and social consequences resulting from Generative AI, for example climate impact or power imbalance of technology development

Lack of understanding & control

The risk of not understanding the outputs or behavior of Generative AI solutions, caused by for example low transparency and explainability of the AI model

Liability & compliance

Legal obligations and compliance with regulations related to the use of Generative AI, including ownership of content and potential implications of misuse or harm caused

Source: Nordic Ethical AI Sandbox Workshop #2, based on Accenture’s Responsible AI Use Case Assessment Framework

Most important prioritization factors

As a general takeaway, participants in the workshop consider the ability to generate efficiency gains and business growth, while maintaining privacy and security as the most important factors to consider when evaluating individual Generative AI use-cases

Efficiency gains

Nordic businesses see a lot of value potential from optimizing business operations both to cut costs, as well as speed up processes in order to quicker deliver their services. Using generative AI can potentially enable improved productivity, automation of repetitive tasks, reduction manual labor, reduction of required resources, identification of inefficiencies and faster decision making processes.

Business growth

Nordic businesses see a lot of value potential from using generative AI to contribute to growth of the business by strengthening brand reputation, providing more personalized services to their customers and finding new market opportunities.

Privacy & security

Nordic businesses prioritize the protection of sensitive data, with privacy and security being critical factors when evaluating whether to pursue a generative AI use-case, as they are wary of potential risks such as data leakage, security breaches or privacy violations.

Source: Nordic Ethical AI Sandbox Workshop #2, based on Accenture's Responsible AI Use Case Assessment Framework

Note: This is a conclusion made by the writers of this guidebook based on the discussions in the Nordic Ethical AI Sandbox Workshop #2. It does not mean that this applies for all participating organizations.

Identifying and mitigating AI and generative AI risks

- The following section includes AI and generative AI risks identified in the Nordic Ethical AI Sandbox, together with examples of risk mitigation methods and learnings from implementing and using generative AI in Nordic organizations.
- Mitigating actions can target both the root case of the risk and the effect. In this guidebook, both technical and non-technical mitigations are included.
- Risks and risk mitigation methods are highly contextual and specific to individual AI use-cases. Organizations therefor need to screen each use-case for the level and type of risk to identify the correct mitigations.

Identify and Assess

Risks ⚠

The potential risk of which an organization is threatened by, with consideration to both adverse impact and likelihood of occurrence

Effects

The consequences of AI risks materializing

Mitigate and Remedy

Organizational

Non-technical practices that help identify, assess, reduce or eliminate different risks across the AI lifecycle

Technical

Tools, techniques or technical methods that help identify, assess, reduce or eliminate different risk factors in the AI lifecycle

**Risks and risk mitigation methods are use-case specific, and organizations need to assess which are applicable for them and for specific use-cases*

Risks identified by Nordic companies

Risks

The potential risk of which an organization is threatened by, with consideration to both adverse impact and likelihood of occurrence

Unreliable outputs	Workforce & talent risks	Bias & harm
Privacy & security risks	Fast evolving technology	Negative sustainability impact
Lack of understanding & control	Liability & compliance	

Effects

Non-exhaustive

The consequences identified by Nordic businesses that come as an effect of the risks

Non-compliance (effect)

- *Legal and compliance (for example with industry standards, GDPR, EU AI Act)*
- *Copyright infringement*

Reputational damage (effect)

- *Loss of trust*
- *Negative brand reputation*

Loss of Control / Power Imbalance

- *Power and control imbalance (for example market oligopolies)*
- *Loss of Control*

Business & Financial Loss

- *Loss of income*
- *Loss of intellectual property*
- *Regulatory fines*
- *Loss of innovation & competitive advantage*

Adverse Sustainable impact (effect)

- *Long term societal implications (for example on human rights, democracy, propaganda, misinformation)*
- *Negative environmental impact*

Source: Nordic Ethical AI Sandbox Workshop #2

Note: This list is non-exhaustive. The risks and effects mentioned were identified during the workshop and may not apply or be exhaustive for to all organizations.

Organizational risk mitigation for (Gen) AI risks

Organizational mitigations are in this context non-technical practices and structures that help identify, assess, reduce or eliminate different risks across the AI lifecycle. These risk mitigations were identified by workshop participants and could in most instances also apply for other AI technologies as well.

Organizational Mitigation Strategies

Non-exhaustive

Principles & Governance

AI Policy
AI policy which defines what AI is and how it can be deployed and used across the organization

Gen AI Council
Gen AI Council or Board that reviews and accept GenAI models and use-cases

Accountability Framework
Documented roles, responsibilities, and accountability structures and processes in relation to AI

AI System Inventory
Central oversight of all deployed (Gen) AI systems (applies for all AI, not just Generative AI)

Standardized Processess

Risk Classification
Standardized process for defining the risk level of individual Gen AI systems, based on predefined risk categories

Responsible AI-by-Design
Documented guidance on how to integrate ethical and resopnsible AI requirements during each AI lifecycle stage

Human-in-the-Loop
Processes and guidance for reviewing and quiality-checking AI outputs

Pre-deployment Monitoring
Process to continiously review deployed AI systems and monitor for significant changes

Training & Culture

Leadership Sponsorship
Clear committment from leadership to foster a culture which promites ethical and responsible deployment and use of AI

Upskilling and AI Literacy
AI educational programs, AI certifications and AI awareness programs tailored for different roles

Gen AI Community of Practice
Educational and knowledge sharing forums for practitioners for learn and share practical examples

Culture Building on RAI
Establish regular AI focus groups, Document organizational values on responsible AI

Ecosystem Collaboration

Industry Sandboxes and Toolkits
Participation in industry sandbox to learn from peers, and leverage existing and guides and toolkits provided by the ecosystem

Vendor Collaboration
Collaboration with AI-model providers, including upstream transparency requirements and data sharing agreements

Stakeholder Engagement
Involvement of external stakeholders to understand the broader impact of AI on society and environment

Source: Nordic Ethical AI Sandbox Workshop #2

Note: This list is non-exhaustive. The risk mitigation strategies mentioned here were identified during the Nordic Ethical AI Sandbox workshop. These are non-exhaustive and and may not be relevant for all organizations.

Technical risk mitigation for (Gen) AI risks

Technical mitigation strategies are in this context tools, techniques or technical methods that help identify, assess, reduce or eliminate different risk factors in the AI lifecycle. These risk mitigations were identified by workshop particpants, and could in most instances also apply for other AI technologies as well.

Technical Mitigation Strategies

Non-exhaustive

Data	Development	Deployment	Cross-Lifecycle
Data Loss Prevention Data encryption, Data masking redaction, Antivirus software, Data loss prevention software	**Prompt Engineering** Best practice for creating prompts to balance specificity with openness to optimize prompt effectiveness	**Monitoring System** Network Failure Monitoring, Output Monitoring, Error Monitoring	**Technical Sandbox & Toolkits** A protected technical environement that allows developers and engineers to test software or system
Sensitivity Labels & Access Controls Classification of data assets based on sensitivity and with clear access rights	**Fine-tuning** Adapting pre-trained models to for specific tasks or use-cases to achieve higher accuracy	**Record-Keeping** Techniques and tools to automatically record events and enable traceability	**AI Frameworks** Open-source and proprietary frameworks to architect, train, validate and deploy AI systems
AI Model & Data Cards Documentation format to provide standardized information to downstream users – "nutrition label" for AI models	**Grounding the Model** Set up parameters, settings and boundaries for what is accurate behaviour for intended use	**Harmful Content Classification** Techniques and methodologies for reviewing generated content to flag harmful or toxic content	**Governance Platform & Toolkit** A platform or toolkit that allow organization to direct, manage, and monitor AI activities according with internal policy
	AI Model Explainability Techniques and methodologies for improving explainability of AI-outputs, for example decision trees, LIME, SHAP	**Guardrails** Technical guardrails that limits the types of user prompts that can be made	**System Deactivation** Tool that can deactive or disable entire system or certain features or services

Source: Nordic Ethical AI Sandbox Workshop #2

Note: This list is non-exhaustive. The risk mitigation strategies mentioned here were identified during the Nordic Ethical AI Sandbox workshop. These are non-exhaustive and and may not be relevant for all organizations.

04.
Robust & Secure AI

Chapter Introduction

Nordic perspectives on **challenges and mitigation strategies related to ensuring robust and secure AI**

Background

- This chapter is based on the **results of the third workshop** and webinar in the Nordic Ethical AI Sandbox, which took place during April 2024.
- AI systems comes with different **vulnerabilities, hacking threats, and complexities of operations**, compared to other software systems. This necessitates attention at several levels of organizations and businesses.
- This chapter explores **Nordic perspectives on challenges related to robustness and security when developing and adopting AI solutions, and what strategies can be applied to mitigate these.** Each use-case comes with its unique set of challenges, and these recommendations should therefore be considered as a starting point for further analyses.

What this chapter will help organizations with

Understanding the **phases of the AI lifecycle** and their influence on AI robustness and security

Identifying key **challenges** related to ensuring robustness and security of AI systems

Understanding what **methods and strategies** exist to mitigate challenges related to AI robustness and security

Robust and secure AI

New vulnerabilities emerge with the adoption of complex machine learning models. The robustness and security of AI models pose challenges distinctly different from those associated with traditional technologies.

Robustness

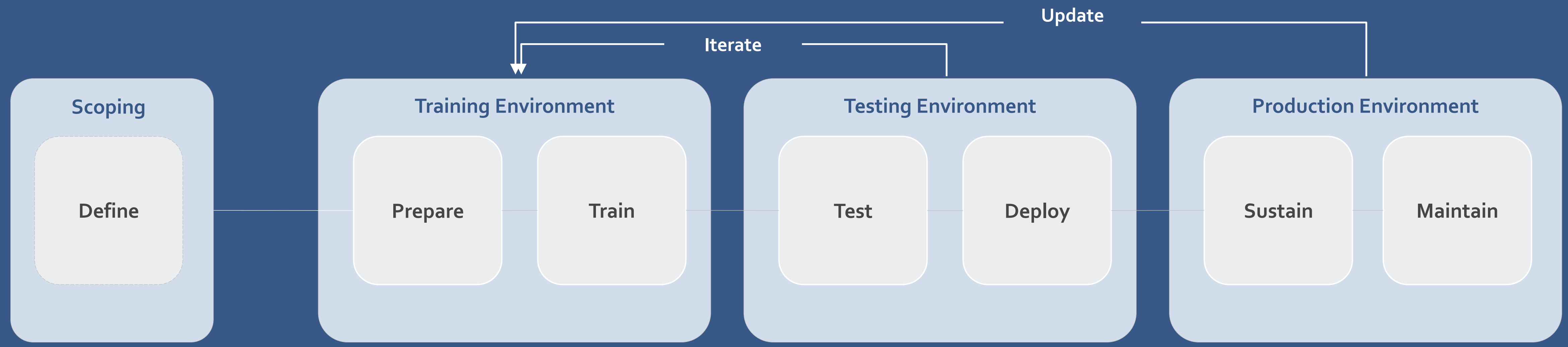
- AI robustness is the ability of an AI system to maintain its performance level under varying conditions, including unexpected conditions.
- This is related to the system's sensitivity to minor changes, in some cases changes imperceptible to humans
- Such changes typically are caused by one of these phenomena:
  - Minor differences in data collection, for example in medical imaging where different installations and/or equipment from different vendors will create outputs with minor differences
  - Minor (or major) changes in the statistical properties of the input data change, for example caused by seasonality factors

AI Security

- The deployment of AI models introduces new threat surfaces and requires new defence methods compared to traditional IT technologies
- Attacks typically are designed to manipulate machine learning models into:
  - Extracting sensitive information from the data that the model has been trained on. This could lead to the disclosure of business or person sensitive data.
  - Inducing erroneous outputs, manipulating the models to cause harmful outputs harmful to the deployer and/or beneficial to the attacker

AI lifecycle process flow

The AI lifecycle describes the different phases from the inception of AI systems to operation and is an iterative process



The **Definition phase** is important for scoping the intended use and setting internal requirements (for, e.g., performance and availability), but also for making sure potential fairness issues (if applicable) will be addressed properly.

The **Training environment** contains data management, data preparations, and model building activities (including algorithm selection, training, and optimization).

The role of the **Testing environment** is to test data quality, model performance, prediction uncertainty, to make decisions whether to deploy or not, and to deploy models for production if tests pass.

The **Production environment** contains activities required for flawless operation; monitoring, logging, model update logic, drift detection & countermeasures, provision of robust inference service, and mechanisms for failure recovery

Managing robustness and security challenges

Nordic companies explored **challenges and pitfalls** along the different phases of the AI lifecycle regarding robustness and security

CHALLENGES IDENTIFIED FROM WORKSHOP INSIGHTS

Scoping

- Lack of clarity regarding the goal and accessible data, including sensitivity levels
- Benefits of the AI use-case for different user groups not well understood, together with the value levers
- Unclear value of leveraging AI compared to other existing approaches
- Lack of clear business case analysis:
    - *Identifying the challenge to be solved*
    - *Determining if AI is the appropriate tool for the challenge*

Training

- Lack of explainability of the model and its decisions
- Fairness based on factors like location and occupation not sufficiently evaluated
- Lack of trustworthy training data, due to unreliable data sources, bias, etc.
- Unreasonable expectations for model accuracy
- Poor handling of sensitive data sharing and consent issues
- Highly automated model training processes excludes human judgement from the training phase
- Lack of target group involvement in the model setup process

Testing

- Retention of unnecessary data
- Testing efforts not sufficiently prioritized
- Lack of budgets for red teaming
- Models lacking robustness to environmental effects
- Model developers not detecting combinations causing underperformance, leading to blind spots
- Unclear criteria for "go/no-go" decisions for production deployment.
- Lack of control and specifications of the data on which results are based

Production

- Data changing meaning or context over time
- Overly specialized model will not be applicable in other business contexts
- Operations teams not following a/the governance process
- Adversarial inputs
- Understanding of when to reassess privacy and risk not shared by all stakeholders, especially with changes to models and data
- Over-reliance on the model's output as being superior to other analyses
- Lack of human operators who can competently infer issues and intervene when necessary

Source: Nordic Ethical AI Sandbox Workshop #3

Note:These are aggregated results from the workshop and do not necessarily apply for all participating organizations. The list should not be considered exhaustive.

Robustness and security mitigation methods

Nordic companies explored **solutions and mitigation methods** relevant in different phases of the AI lifecycle

RECOMMENDATIONS BASED ON WORKSHOP INSIGHTS	Scoping	Training	Testing	Production
ORGANIZATIONAL	• Define clear business case • Define AI risk categories/thresholds • Establish procurement • Conduct fundamental rights impact assessment	• Establish procurement guidelines • Conduct fundamental rights impact assessments • Ensure effective monitoring	• Consider manual moderation	• Implement a change control board Communicate clear expectations of use and reliability • Consider relevant organizational reporting lines • Discuss potential responsibilities to withdraw the AI system from use
TECHNICAL	• Ensure effective human oversight to implement risk screening	• Consider techniques as poison filters and differential privacy • Remove unneeded attributes in the data engineering process • Investigate the relevance of using synthetic data • Document model transparency	• Find model blind spots (underperforming combinations) • Discuss minimum requirements for quality and fairness	• Facilitate the regular monitoring of adversarial actions
LEGAL & POLICY			• Agree minimum requirements for quality and fairness	• Create AI relevant procurement practices, including how to mitigate risks in contractual terms • Create an AI incident policy (what to do if something goes wrong)

Source: Nordic Ethical AI Sandbox Workshop #3

Note:These are aggregated results from the workshop and do not necessarily apply for all participating organizations. The list should not be considered exhaustive.

Key takeaways for ensuring robust and secure AI

Workshop participants considered the following factors to be critical for ensuring responsible, robust and secure AI

Overarching factors

- **Task Definition**: Defining tasks as simply as possible aids in clarity and effective execution
- **Human Factors**: Considering the long-term impact of AI and ML on humans, including discussing human factors in deployment, is crucial for ethical AI implementation
- **Complexity and Mitigation**: Addressing the complexity of problem solving and emphasizing the importance of getting it right, including considering both mitigation measures and the involvement of domain experts
- **Framework and Sensitivity**: Understanding the sensitivity of sharing how models work in the real world and iterating models based on real-world performance and failures helps in refining and improving AI systems
- **Sector Challenges**: Discussing common challenges across different sectors, such as health and maintenance models, highlights the universality of certain issues and facilitates shared learning

Data and training

- **Data Supply**: Ensuring access to the right data for the project is vital. Validating data availability is a crucial initial step
- **Training Data Issues**: Focusing on cleaning and resolving issues with training data is fundamental to developing reliable models

Testing

- **Standardized metrics**: Utilizing standardized metrics (based on models from ISO or other standardization organizations) to measure success in AI and ML applications is important for benchmarking
- **Robustness**: Exploring the robustness of models, including testing for vulnerabilities and potential hacking, is necessary for maintaining security
- **Validation and Stress Testing**: Conducting thorough validation and stress testing of models ensures they can perform reliably under various conditions

Production

- **Deployment**: Safety and mitigation of failure modes during deployment are essential. Establishing governance processes for both deployment and solution phases helps maintain oversight and accountability
- **Usage Drift**: Monitoring and addressing how AI and ML models may deviate from their intended use over time ensures relevance and accuracy

Source: Nordic Ethical AI Sandbox Workshop #3

Note:These are aggregated results from the workshop and do not necessarily apply for all participating organizations. The list should not be considered exhaustive.

05.
Transparent and Explainable AI

Chapter Introduction

Nordic perspectives on **challenges and mitigation strategies related to transparency and explainable AI**

Background

- This chapter is based on the **results of the fourth workshop** and webinar in the Nordic Ethical AI Sandbox, which took place during April/May 2024.
- The **lack of transparency and explainability is a common criticism of AI system deployments**. The urgency to integrate AI into strategic operations across Nordic businesses underscores the necessity to mitigate such issues.
- This chapter explores **Nordic perspectives on challenges related to transparency and explainability when developing and adopting AI solutions, and what strategies can be applied** to ensure that solutions are understandable and accountable. Each use case comes with its unique set of challenges, and these recommendations should therefore be considered as a starting point for further development.

What this chapter will help organizations with

Emphasizing the **importance of building trust** through transparent and explainable AI systems

Identifying key **challenges** related to the lack of transparency and explainability of AI systems

Highlighting some of the **methods and strategies** that exist to aid with transparency and explainability of AI systems

Transparent & Explainable AI

Use of AI technologies impacts trust on individual as well as societal levels. Transparency and explainability of models are crucial elements in responsible and ethical deployment of AI systems.

Transparency

- Transparency seeks to ensure that all stakeholders can understand how an AI system arrives at a result, such as a decision or recommendation.
- Several factors are important for AI transparency, most notably:
  - Interpretability: the capability to provide information about the relationships between model inputs and outputs.
  - Explainability: the capability to explain the model’s decision-making process in terms understandable to the end user.
  - Accountability: the capability of AI systems to learn from mistakes and improve over time, while organizations should take suitable corrective actions to prevent similar errors in the future.
- Social transparency focuses on the implications of AI deployment on society as a whole.

Explainability

- Explainable AI (or XAI) refers to the ability of an AI system to provide easy-to-understand explanations for its decisions and actions.
- Many modern AI models are hugely complex, which tend to obfuscate the model’s decision-making process to a level where not even experts can explain it.
- XAI helps build trust with the users by enabling them to understand, trust, and effectively manage these AI systems.
- This is especially important in high-stakes domains and safety critical applications of AI.

Transparent and explainable AI challenges

Nordic companies explored challenges in ensuring transparency and explainability in the different phases of the AI life cycle

CHALLENGES IDENTIFIED FROM WORKSHOP INSIGHTS

Scoping	Training	Testing	Production
Balancing transparency with competitive advantage: There is a challenge in determining how much of the AI's operational logic can be disclosed without compromising business secrets or competitive advantages.	**Determining the right level of transparency:** It is challenging to establish the appropriate level of transparency for different stakeholders without negatively impacting the functionality or security of the AI.	**Explaining errors:** Explainability aids in debugging and refining AI systems, especially in handling corner cases or errors arising from inadequate data or model understanding.	**Monitoring and adjustment:** Organizations must have the capability to continuously monitor, evaluate, and adjust AI systems, ensuring they remain effective and aligned with ethical standards over time.
Inclusion of sensitive features: It is essential to consider when and how sensitive features like gender or ethnicity should be incorporated into the AI model, ensuring they are used objectively and justifiably.	**Bias and fairness considerations**: The need to recognize and mitigate biases, which may vary culturally and regionally, is crucial during the deployment phase to ensure fairness and ethical use.	**Data representation and blind spots:** Understanding whether the data used is representatively and free of biases is vital for the effective functioning of the AI, necessitating mechanisms to identify and address any blind spots.	**Educating users:** There is a need to educate end users on the AI's functionalities and limitations to ensure responsible usage.
Interpretability of outcomes: The AI system must be designed to not only perform tasks but also provide understandable outputs. This includes creating models that can explain decisions in a way that end users can comprehend.	**Ensuring model understandability**: For AI systems that operate at higher levels of autonomy, there is a heightened demand for transparency to ensure that stakeholders can conceptually understand AI outputs and their implications.		**Relevance over time:** The ability to incorporate new knowledge or correct misunderstandings within the AI system is crucial for its ongoing relevance and accuracy.

Source: Nordic Ethical AI Sandbox Workshop #4

Note:These are aggregated results from the workshop and does not necessarily apply for all participating organizations. The list should not be considered exhaustive.

Transparent and explainable AI mitigation mechanisms

Nordic companies explored solutions and mitigation mechanisms relevant in different phases of the AI life cycle

RECOMMENDATIONS BASED ON WORKSHOP INSIGHTS

Scoping

- Understanding the intended audience of an AI system is key in determining the necessary level of explainability.
- Opening up data used for training AI models can enhance transparency and accountability.
- Understanding which data is used and why it is chosen for training of AI systems is crucial for ensuring the quality and trustworthiness of the system.

Training

- Understanding how a well-designed user interface can enhance the explainability of an AI system.
- Understanding the relevance and functionality of the algorithms and models in use is crucial for enhancing explainability.
- Ensuring transparency at every stage of model training aids the development team in detecting shortcomings in the model's performance.

Testing

- Ensuring that the AI system is explainable to the end user prior to deployment.
- Transparently describing the architecture of the AI system for deployment to ensure there are no leaks or potentially harmful stages.
- Assessing that the active learning and process of injecting more data to the model is reliable and robust.

Production

- Assessing whether the necessary resources to actively monitor, modify, and adjust the AI over time are available.
- Assessing whether the organization possesses the capacity and skills required to comprehend potential issues.
- Maintaining the system in a way that ensures it remains transparent and explainable is crucial.

Source: Nordic Ethical AI Sandbox Workshop #4

Note:These are aggregated results from the workshop and does not necessarily apply for all participating organizations. The list should not be considered exhaustive.

Key takeaways for ensuring Transparent and Explainable AI

Workshop participants considered the following factors to be critical for ensuring responsible, robust and secure AI

Transparency and Explainability

Transparency alone is insufficient; hence explanations are crucial

Transparency and Demand

Higher transparency is in some cases demanded from AI than from human operators

Transparency and Competitive Edge

In some cases, transparency and the protection of someone's competitive edge needs to be balanced

Purpose and Representation

AI models need to be representative of tasks and inputs. Step-by-step explanations can aid understanding

User Responsibility

End users responsible for not misusing AI systems

Adaptability and Modification

AI systems should be designed to avoid architectural lock-in. This could help in cases where the regular monitoring identifies needs for adjustments

Bias and fairness

Address bias and fairness considering cultural contexts and use this to ensure representation in training data

Explainability and Debugging

Explainability aids in debugging and improving AI systems

Resource Availability

Assess to organizational resources and skills for ongoing AI maintenance is critical. In some cases this includes the education of end users

Trust and Understanding

Provide tools and meaningful user interfaces to help users understand AI decisions

Learning From Mistakes

Use transparency to learn and improve data preparation

Source: Nordic Ethical AI Sandbox Workshop #4

Note:These are aggregated results from the workshop and does not necessarily apply for all participating organizations. The list should not be considered exhaustive.

06.
Accelerating Ethical and Responsible AI in the Nordics

Summary: accelerating ethical and responsible AI

The ethical and responsible AI journey will look different for organizations, depending on factors such as company size, capabilities, organizational complexity and exposure to global markets and regulation

KEY TAKEAWAYS FOR DIFFERENT TYPES OF ORGANIZATIONS

Start-ups

- For start-ups with few employees, ensuring responsible development, procurement and use of AI is likely going to be less about having strict governance with dedicated forums, processes and controls. Instead, fostering a culture from the ground up that puts ethical and responsible AI practices at the core of all AI operations will be critical. Making ethical and responsible AI a central part of the company’s core values and culture will enable all employees to feel accountability and responsibility for living up to those values.
- Unless AI is part of the core business, start-ups are less likely to develop their own AI models, instead they will rely on pre-built AI solutions. Therefore, they should focus on building knowledge about the vendor landscape, and make informed decisions about which vendor to partner with, considering their approach to responsible AI and use of data.
- Small organizations with no legacy also has great potential to ”do right” from the start, by adopting a responsible-by-design approach to AI.

SME

- As organizations grow, it will be more important to develop formal AI governance to keep an overview of AI operations and potential risks. Building on existing principles, core values and responsible AI culture can help enforce new policies, processes and controls. Ensure there is clarity on roles and responsibilities of ensuring ethical and responsible AI.
- Empower employees to adopt a responsible approach to their use of AI, by creating clear instructions and education material about potential risks and limitations with AI tools. For AI that is being deployed by the organization, standardize best practice for system development and testing based on key responsible AI dimensions such as explainability, fairness, robustness and fairness. Additionally, adopt robust documentation practices for AI development.
- As organizations grow, they might enter new markets and jurisdictions. Staying informed about new data and AI regulations will be essential to successful market entry.
- Treat ethical and responsible AI as an ongoing, iterative process, continually assessing and improving practices within resource constraints.

ENTERPRISE

- Most Nordic enterprises are not *AI-companies*, even though many might rely on a strong digital and technical foundation. These organizations are likely to already have significant structures in place already in terms of policies, governance and processes that impact how AI is developed, procured and used across the organization. Because of this, operationalizing ethical and responsible AI is likely to be centered around uplifting existing structures rather than creating new ones. Some organization could need to create new teams, roles and capabilities, but implementing this requires working with what is already in place. Some responsible AI capabilities might be centralized, and some federated. Creating a scalable operating model for ethical and responsible AI will be important.
- A major challenges for large organizations is ensuring oversight of AI. Identifying all AI systems, screening these for risks, and keeping an AI inventory will be immediate actions, especially with the EU AI Act.
- Some large enterprises could operate across multiple markets, with varying legal landscapes. Continuously monitoring legal developments, and harmonizing frameworks will be essential for multi-national organizations.

Source: Aggregated results from all workshops in the Nordic Ethical AI Sandbox

Toolkit: Discovery Workshop to Define the Ethical and Responsible AI Roadmap

Workshop methodology

How to create an approach for ethical and responsible AI, and define the roadmap to operationalize

Introduction & Objective

This workshop is designed for businessess to execute internally, to build an initial understanding about what ethical considerations their use of AI is associated with, and what actions are needed to fully operationalize ethical and responsible AI within their organization.

Process & Timeline

The workshop consists of four exercises that guide the participants to identify their organization's AI opportunities and risks. The workshop begins with a high-level inventory of current practices to identify the organizations starting point, and ends with defining a roadmap and action plan to improve current processes or create new practices that help ensure alignment with AI principles. The workshop exercises are conducted in sequential order with a estimated completion time of 2,5 hours.

Exercises & Participants

The workshop is designed to be cross-functional, and the exercices require knowledge about current and future AI operatations from multiple perspectives. The recommendation is to form a core workshop team of not more than 10 people, who can later validate the findings with other relevant stakeholders. Recommended perspectives to have included in the workshop are AI/Data Science, Data Engineering, Legal, Product, HR/Learning & Development.

Expected Outcome

This workshop is meant to be a starting point for organizations to create a formal roadmap to mature critical capabilities needed to develop and/or use AI responsibly. After completing the workshop, the organization will have defined and prioritized short- and long-term initiatives to address critical gaps identified in current ways of developing and/or using AI. This is the first step towards establishing a unified approach for ethical and responsible AI.

Overview of the exercises & output

Exercise: 0.5 – Health check (time ~ 20min)

- Identifying the current ethical and responsible AI state by completing the check-list and identify areas where the organization is lacking to fully operationalize ethical and responsible AI .

Exercise: 1 – Define your AI context (time ~ 40min)

- Understanding your AI vision and associated risks, to better define ethical and responsible AI principles.

Exercise: 2 – Identify critical gaps (time ~ 40min)

- Mapping out the current known practices and processes that are connected to the current principles for ethical and responsible AI , as well as identifying any gaps.

Exercise: 3 – Define principles and roadmap (time ~50min)

- Identifying prioritized initiatives, and ethical and responsible AI principles,including both short- and long-term priorities.

Output:

- Outlined a short- and long-term plan of initiatives

Exercise 0.5: Health Check

Task

1 Together in the large group, fill out the Health check questionnaire to map the high-level state of your organization's current responsible AI practices (20 min)

Tools

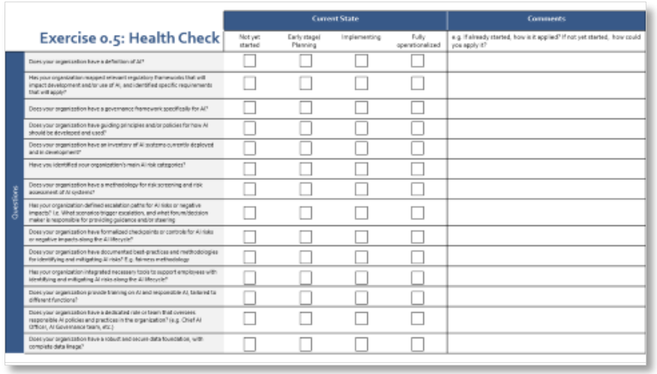
Template

- Exercise 0.5: Health Check

Time

Total time 20 min

1



Exercise 0.5: Health Check

		Current State				Comments
		Not yet started	Early stage/ Planning	Implementing	Fully operationalized	e.g. If already started, how is it applied? If not yet started, how could you apply it?
Questions	Does your organization have a definition of AI?	□	□	□	□	
	Has your organization mapped relevant regulatory frameworks that will impact development and/or use of AI, and identified specific requirements that will apply?	□	□	□	□	
	Does your organization have a governance framework specifically for AI?	□	□	□	□	
	Does your organization have guiding principles and/or policies for how AI should be developed and used?	□	□	□	□	
	Does your organization have an inventory of AI systems currently deployed and in development?	□	□	□	□	
	Have you identified your organization's main AI risk categories?	□	□	□	□	
	Does your organization have a methodology for risk screening and risk assessment of AI systems?	□	□	□	□	
	Has your organization defined escalation paths for AI risks or negative impacts? I.e. What scenarios trigger escalation, and what forum/decision maker is responsible for providing guidance and/or steering	□	□	□	□	
	Does your organization have formalized checkpoints or controls for AI risks or negative impacts along the AI lifecycle?	□	□	□	□	
	Does your organization have documented best-practices and methodologies for identifying and mitigating AI risks? E.g. fairness methodology	□	□	□	□	
	Has your organization integrated necessary tools to support employees with identifying and mitigating AI risks along the AI lifecycle?	□	□	□	□	
	Does your organization provide training on AI and responsible AI, tailored to different functions?	□	□	□	□	
	Does your organization have a dedicated role or team that oversees responsible AI policies and practices in the organization? (e.g. Chief AI Officer, AI Governance team, etc.)	□	□	□	□	
	Does your organization have a robust and secure data foundation, with complete data lineage?	□	□	□	□	

Exercise 0.5: Health Check

Questions	Current State				Comments
	Not yet started	Early stage/ Planning	Implementing	Fully operationalized	e.g. If already started, how is it applied? If not yet started, how could you apply it?
Does your organization have a definition of AI?	□	□	□	□	
Has your organization mapped relevant regulatory frameworks that will impact development and/or use of AI, and identified specific requirements that will apply?	□	□	□	□	
Does your organization have a governance framework specifically for AI?	□	□	□	□	
Does your organization have guiding principles and/or policies for how AI should be developed and used?	□	□	□	□	
Does your organization have an inventory of AI systems currently deployed and in development?	□	□	□	□	
Have you identified your organization's main AI risk categories?	□	□	□	□	
Does your organization have a methodology for risk screening and risk assessment of AI systems?	□	□	□	□	
Has your organization defined escalation paths for AI risks or negative impacts? I.e. What scenarios trigger escalation, and what forum/decision maker is responsible for providing guidance and/or steering	□	□	□	□	
Does your organization have formalized checkpoints or controls for AI risks or negative impacts along the AI lifecycle?	□	□	□	□	
Does your organization have documented best-practices and methodologies for identifying and mitigating AI risks? E.g. fairness methodology	□	□	□	□	
Has your organization integrated necessary tools to support employees with identifying and mitigating AI risks along the AI lifecycle?	□	□	□	□	
Does your organization provide training on AI and responsible AI, tailored to different functions?	□	□	□	□	
Does your organization have a dedicated role or team that oversees responsible AI policies and practices in the organization? (e.g. Chief AI Officer, AI Governance team, etc.)	□	□	□	□	
Does your organization have a robust and secure data foundation, with complete data linage?	□	□	□	□	

Exercise 1: Define your organization's AI context

Task

1. Together in the large group, evaluate your organization's **current and future use of AI** (10 min)
   a) What is your organization's definition of AI
   b) How is your organization currently using AI
   c) How will your organization use AI in the future

2. Together in the large group, define and document your organization's **main AI risk categories** (15 min)

3. Together in the large group, outline and document relevant **guiding principles** to help operationalize AI responsibly within your organization (15 min)

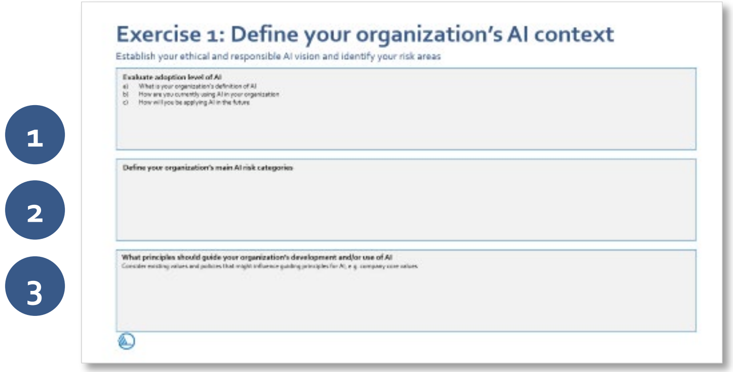
Tools

Template

- Exercise 1: Define your organization's AI context

Time

Total time 40 minutes



Exercise 1: Define your organization's AI context

Establish your ethical and responsible AI vision and identify your risk areas

Evaluate adoption level of AI
a) What is your organization's definition of AI
b) How are you currently using AI in your organization
c) How will you be applying AI in the future

Define your organization's main AI risk categories

What principles should guide your organization's development and/or use of AI
Consider existing values and policies that might influence guiding principles for AI, e.g. company core values

Exercise 1: Define your organization's AI context

Establish your ethical and responsible AI vision and identify your risk areas

Evaluate adoption level of AI

a) What is your organization's definition of AI

b) How are you currently using AI in your organization

c) How will you be applying AI in the future

Define your organization's main AI risk categories

What principles should guide your organization's development and/or use of AI

Consider existing values and policies that might influence guiding principles for AI, e.g. company core values

Exercise 2: Identify critical gaps

Task

1. Individually identify **current practices and processes** that support upholding of the guiding principles defined in exercise 1. Write them down on post-it notes (5 min)

2. Individually identify **lacking capabilities** that are needed to uphold the guiding principles defined in exercise 1. If needed, leverage the framework provided in this guidebook to categorize the capabilities. Write them down on post-it notes (10 min)

3. Together in the large group, discuss the missing capabilities and **prioritize the most critical ones** (25 min)

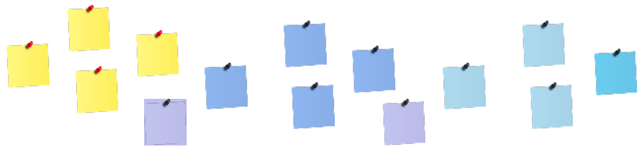
Tools

Post-it notes

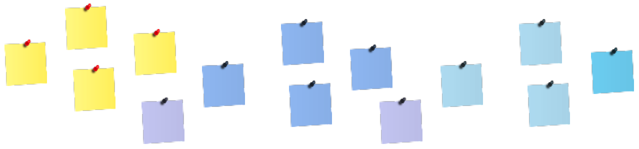
Time

Total time 40 minutes

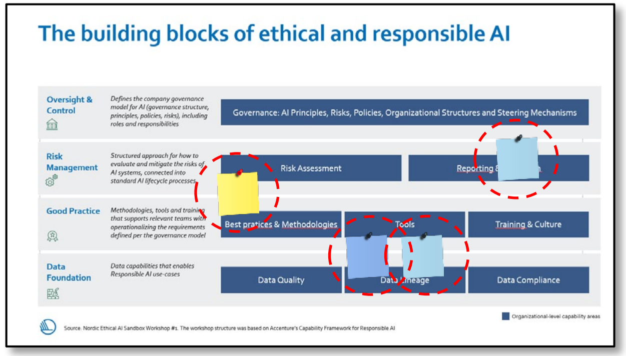
1 Current work practices & processes



2 Missing Capabilities



3



The building blocks of ethical and responsible AI

Oversight & Control	Defines the company governance model for AI (governance structure, principles, policies, risks), including roles and responsibilities	Governance: AI Principles, Risks, Policies, Organizational Structures and Steering Mechanisms		
Risk Management	Structured approach for how to evaluate and mitigate the risks of AI systems, connected into standard AI lifecycle processes	Risk Assessment	Reporting &	
Good Practice	Methodologies, tools and training that supports relevant teams with operationalizing the requirements defined per the governance model	Best pratices & Methodologies	Tools	Training & Culture
Data Foundation	Data capabilities that enables Responsible AI use-cases	Data Quality	Data Lineage	Data Compliance

■ Organizational-level capability areas

Source: Nordic Ethical AI Sandbox Workshop #1. The workshop structure was based on Accenture's Capability Framework for Responsible AI

Exercise 3: Prioritize and document initiatives

Task

1. Individually list up to 7 potential initiatives for your organization that address the identified gaps from exercise 2 and support with upholding the guiding principles identified in exercise 1 (5 min)

2. Together in the large group, discuss the proposed initiatives and prioritize 5 in total (15 min)

3. Together in the large group, detail the prioritized initiatives according to the provided template (30 min)

Tools

Template

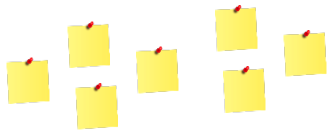
- Exercise 3: Responsible AI Initiatives

Post-it notes

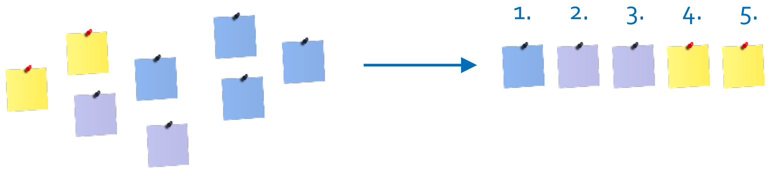
Time

Total time 50 minutes

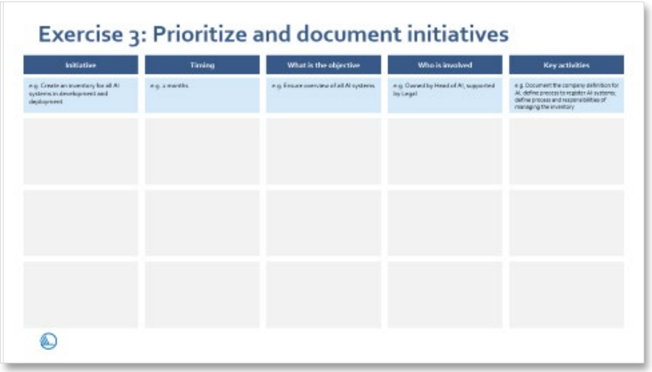
1



2



3



Exercise 3: Prioritize and document initiatives

Initiative	Timing	What is the objective	Who is involved	Key activities
e.g. Create an inventory for all AI systems in development and deployment	e.g. 2 months	e.g. Ensure overview of all AI systems	e.g. Owned by Head of AI, supported by Legal	e.g. Document the company definition for AI, define process to register AI systems; define process and responsibilities of managing the inventory

Exercise 3: Prioritize and document initiatives

Initiative	Timing	What is the objective	Who is involved	Key activities
e.g. Create an inventory for all AI systems in development and deployment	e.g. 2 months	e.g. Ensure overview of all AI systems	e.g. Owned by Head of AI, supported by Legal	e.g. Document the company definition for AI; define process to register AI-systems; define process and responsibilities of managing the inventory

Appendix

Workshop participants

1

Nov 2023

Building and Operationalizing an Approach to Ethical & Responsible AI

- Gjensidige
- Equinor
- Katapult Group & VC
- Zenseact
- Volvo Cars
- 2021.AI
- Loihde
- Saab
- Aidn

2

Feb 2024

Safeguards for Generative AI

- Equinor
- Volvo Cars
- Saab
- Aidn
- Apheris
- Innsikt.ai
- Ericsson
- Saidot
- Factiverse
- Islandsbanki

3

Mar 2024

AI Lifecycle Management for Security & Robustness

- Equinor
- Turre Legal
- Roosa AI
- Save the Children
- Volvo Cars
- Ardoq
- Visiba Group AB
- Saab
- Latticeflow AI
- Moventrac

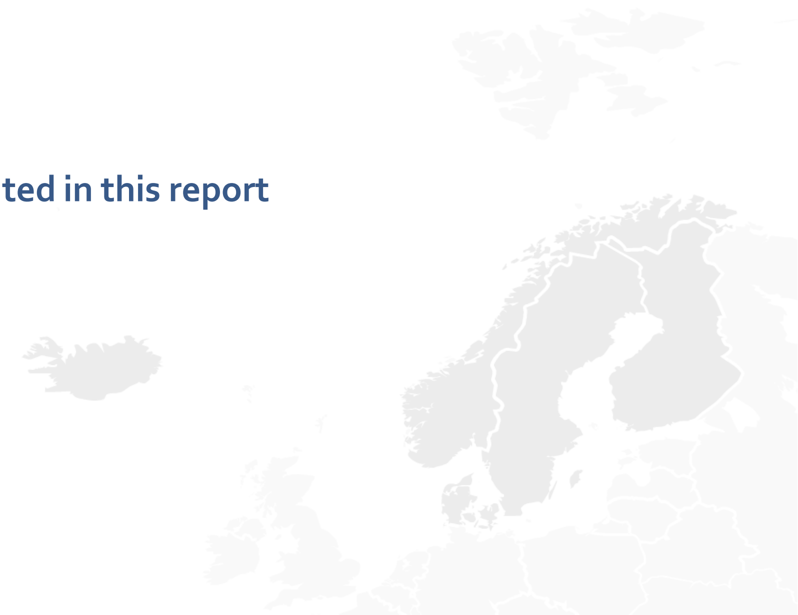
4

Apr 2024

AI Lifecycle Management for Transparency & Explainability

- Equinor
- Turre Legal
- Saidot
- Chalmers Industriteknik
- Resoniks
- Eficode
- Visiba Group

Thank you to the participants in the Nordic Ethical AI Sandbox that contributed to the insights generated in this report



About this publication

The Nordic Ethical AI Guidebook

Insights from the Nordic Ethical AI Sandbox

Us2024-422

Published: 2. September 2024