Intelligent Maintenance Conference (IMC) 2024

Lausanne, Switzerland

Responsible AI Development and Legal Compliance: *Navigating the New Landscape of AI Regulation*

Rialda Spahic, PhD

Task Manager of Responsible AI, Equinor

Image source: Equinor Media Bank

NEWS ARTICLE | 1 August 2024 | Directorate-General for Communication | 2 min read

AI Act enters into force

On 1 August 2024, the **European Artificial Intelligence Act (AI Act) enters into force**. The Act aims to **foster responsible artificial intelligence development and deployment in the EU**.

WORLD ECONOMIC FORUM

Join us | Sign in

Why we need to care about responsible AI in the age of the algorithm

Mar 22, 2023

Businesses need to take responsible AI seriously to remain competitive and avoid liability.
Image: Unsplash/Christopher Burns

Ayesha Gulley
Public Policy and Governance Associate,
Holistic AI

Uncertainty and risk mitigation

Environmental impact

AI Race

Security risks

Bias and fairness

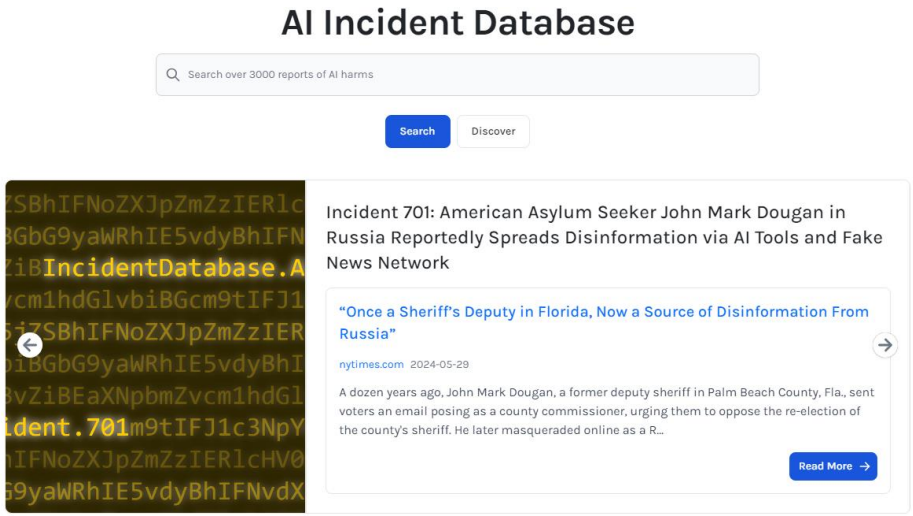
Change in workplace

Image source: Stock Images

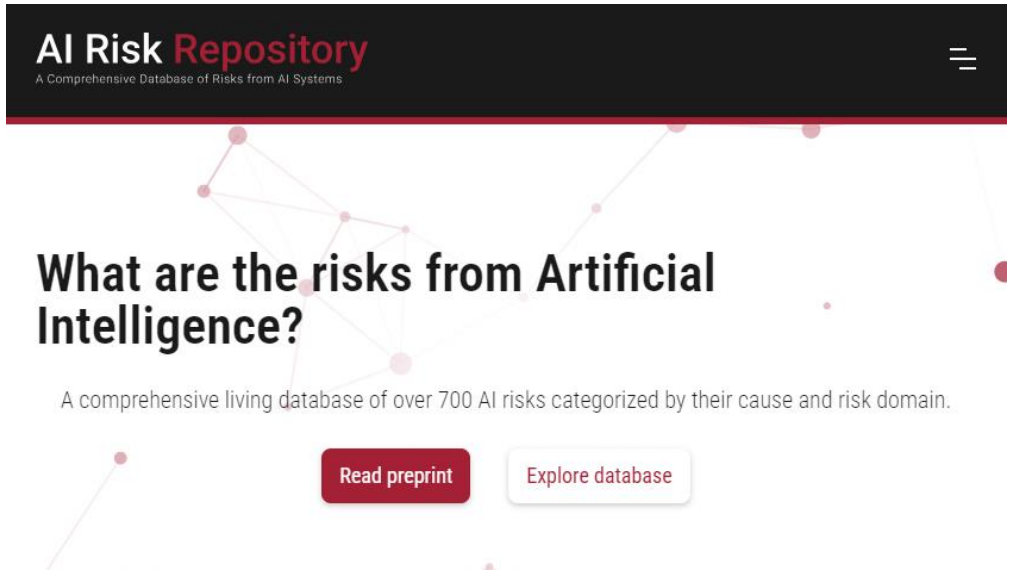
Beyond legal compliance

Why Responsible AI?

https://mlco2.github.io/impact/



https://incidentdatabase.ai/



https://airisk.mit.edu/

Responsible AI Principles

for designing, implementing and using AI

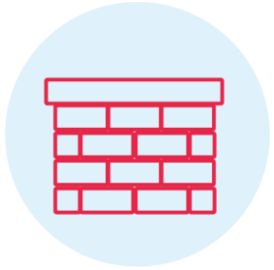
equinor

Governance and Accountability

Safety and Security

Fairness and Openness

Human Oversight



Robustness and Resilience

Sustainability

NEWS ARTICLE | 1 August 2024 | Directorate-General for Communication | 2 min read

AI Act enters into force

On 1 August 2024, the **European Artificial Intelligence Act (AI Act) enters into force**. The Act aims to **foster responsible artificial intelligence development and deployment in the EU**.

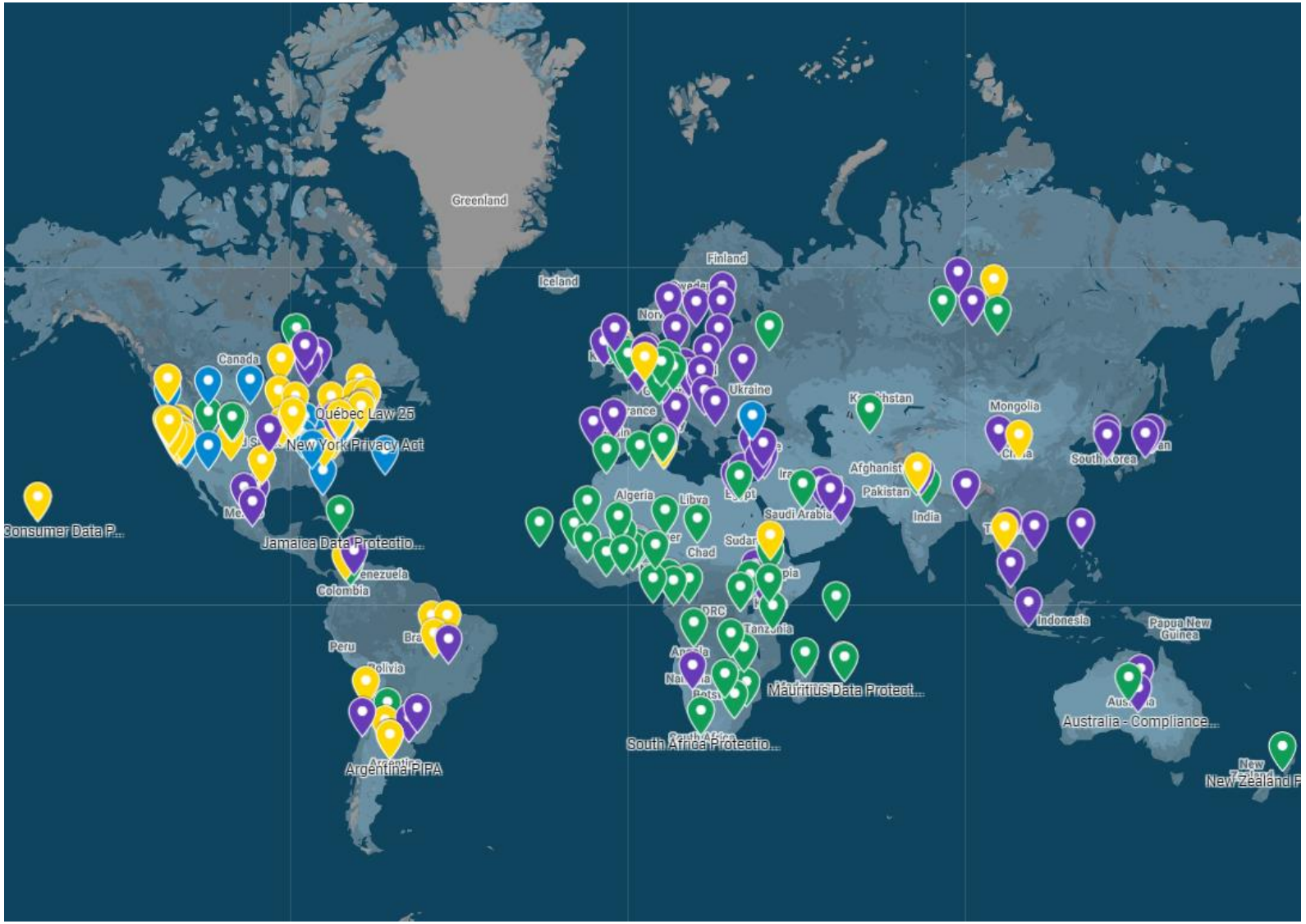
Why we need to care about responsible AI in the age of the algorithm

Our legal obligations

Our obligations as responsible members of society

Image source: Stock Images

A Snapshot of Regulatory Landscape



Map of Global AI Regulations (April 2024)

Source: Fairly AI Map of Global Regulations April 23, 2024 [https://www.fairly.ai/blog/map-of-global-ai-regulations]

equinor

EU AI Act

GDPR, Data Protection Law

Regulation on Cybersecurity

Machinery Directive

Competition Law

Intellectual Property Law

Consumer Protection

Liability Law

Labor Law

Industry-specific regulations

Other laws, policies, regulations

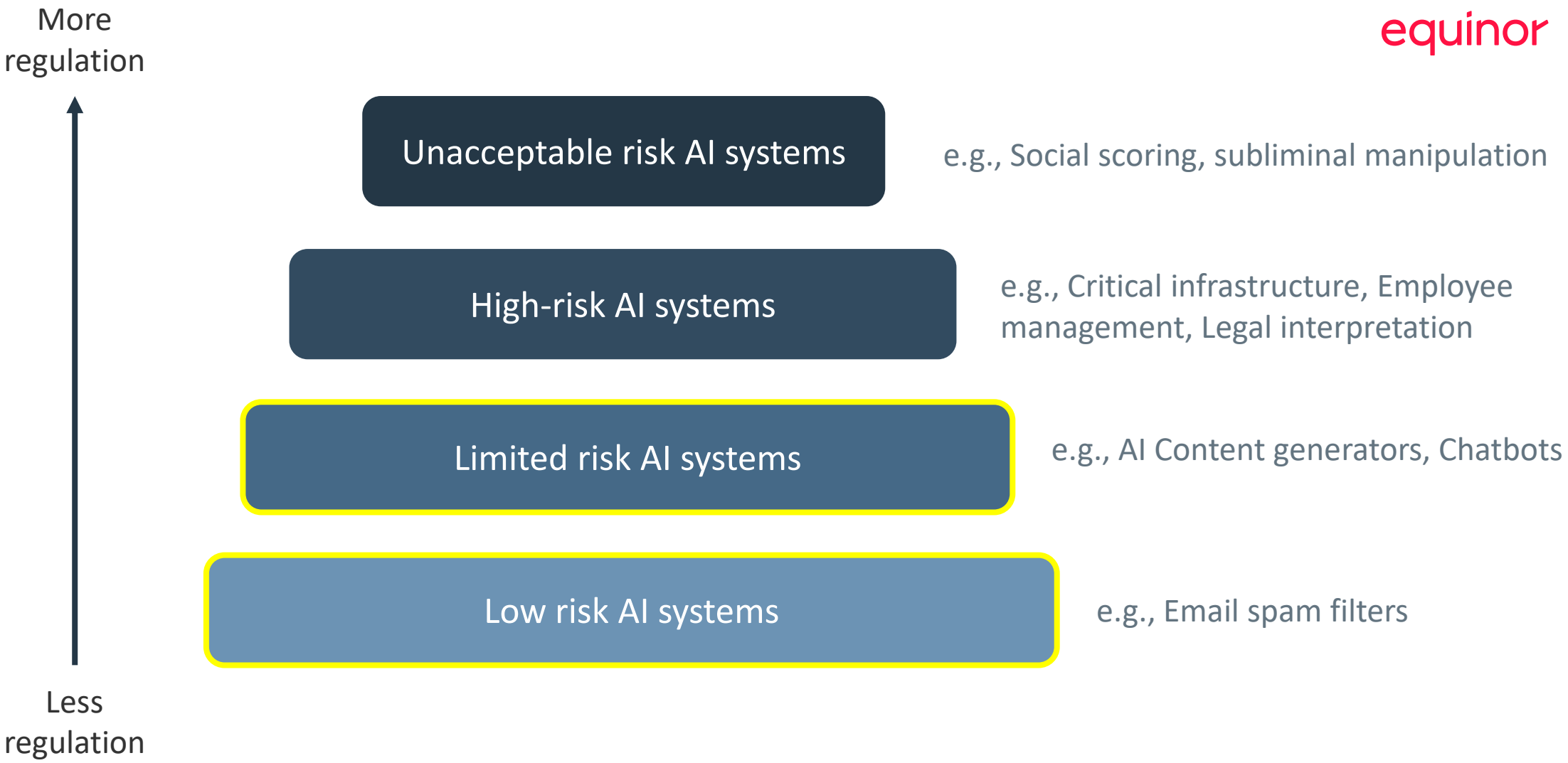
Image source: Wikipedia

AI Definition

EU AI Act, OECD

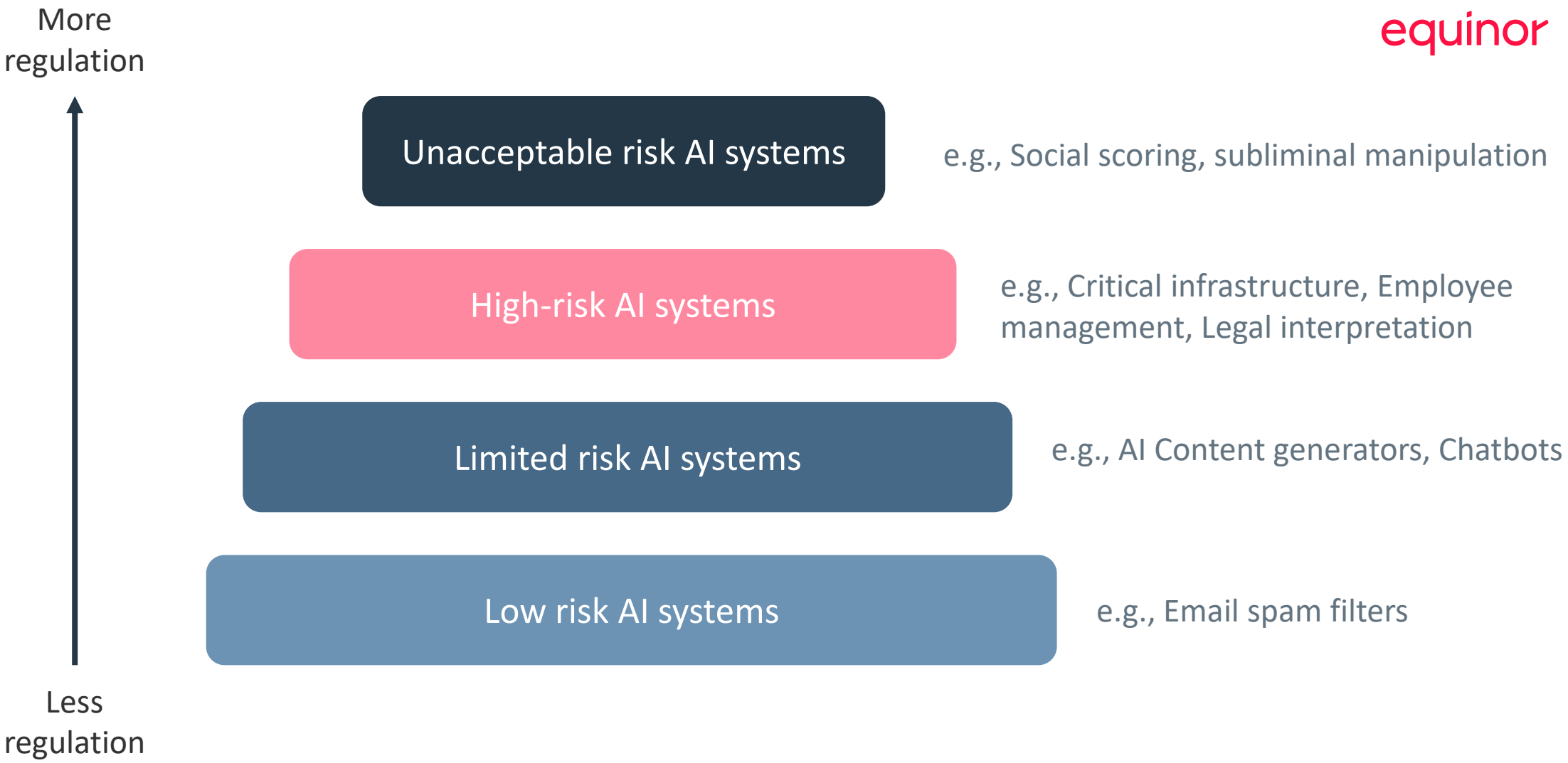
- ✓ ‘AI system’ means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, *for explicit or implicit objectives*, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments;

EU AI Act



Ref. EU AI Act

EU AI Act



Ref. EU AI Act

High-risk AI Systems

AI systems that can result in significant harm to people's health, safety, fundamental rights or the environment.

Safety components for machinery

Safety or protective equipment used in potentially explosive environment

Lifts or safety components of lifts

Medical devices

Recreational crafts

Toys

AI systems used in the following areas are not automatically high-risk, *unless* they pose a **significant risk**:

Management and operation of critical infrastructure

Biometric identification and categorization of natural persons

Employment, worker management

Essential private and public services and benefits

Assistance in legal interpretation and application of the law

Education and vocational training

Law enforcement

Migration, asylum and border control management

Ref. EU AI Act
Image source: Stock Images

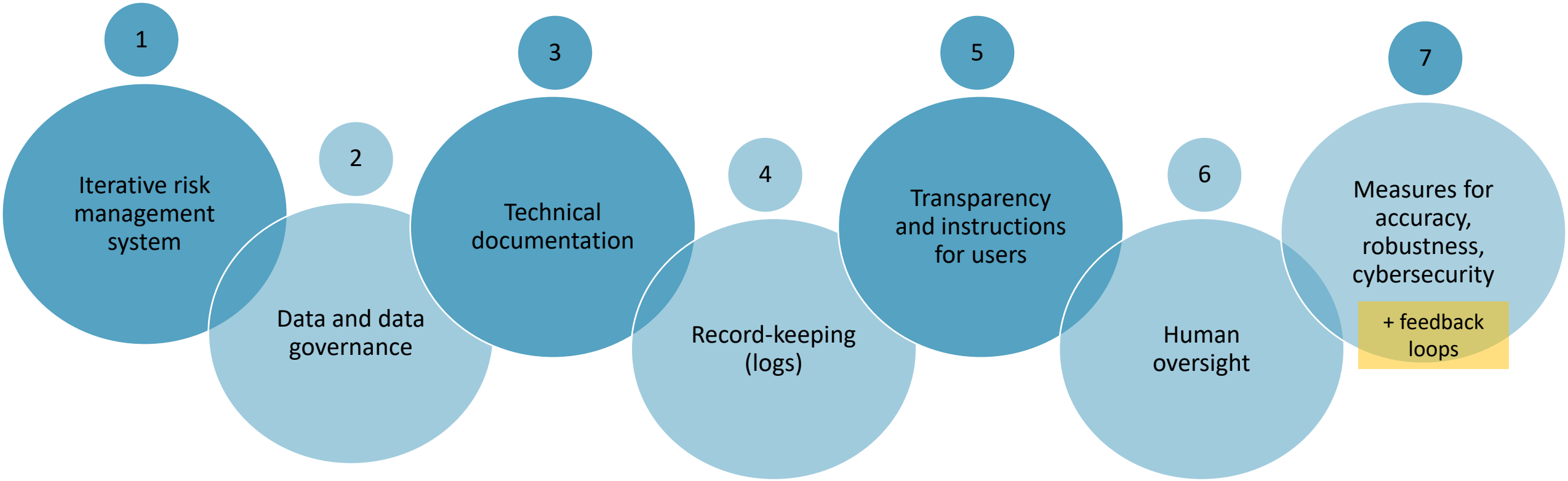
High-Risk AI System Criteria

Criteria used to assess whether an AI system poses a risk of adverse impact on fundamental rights:

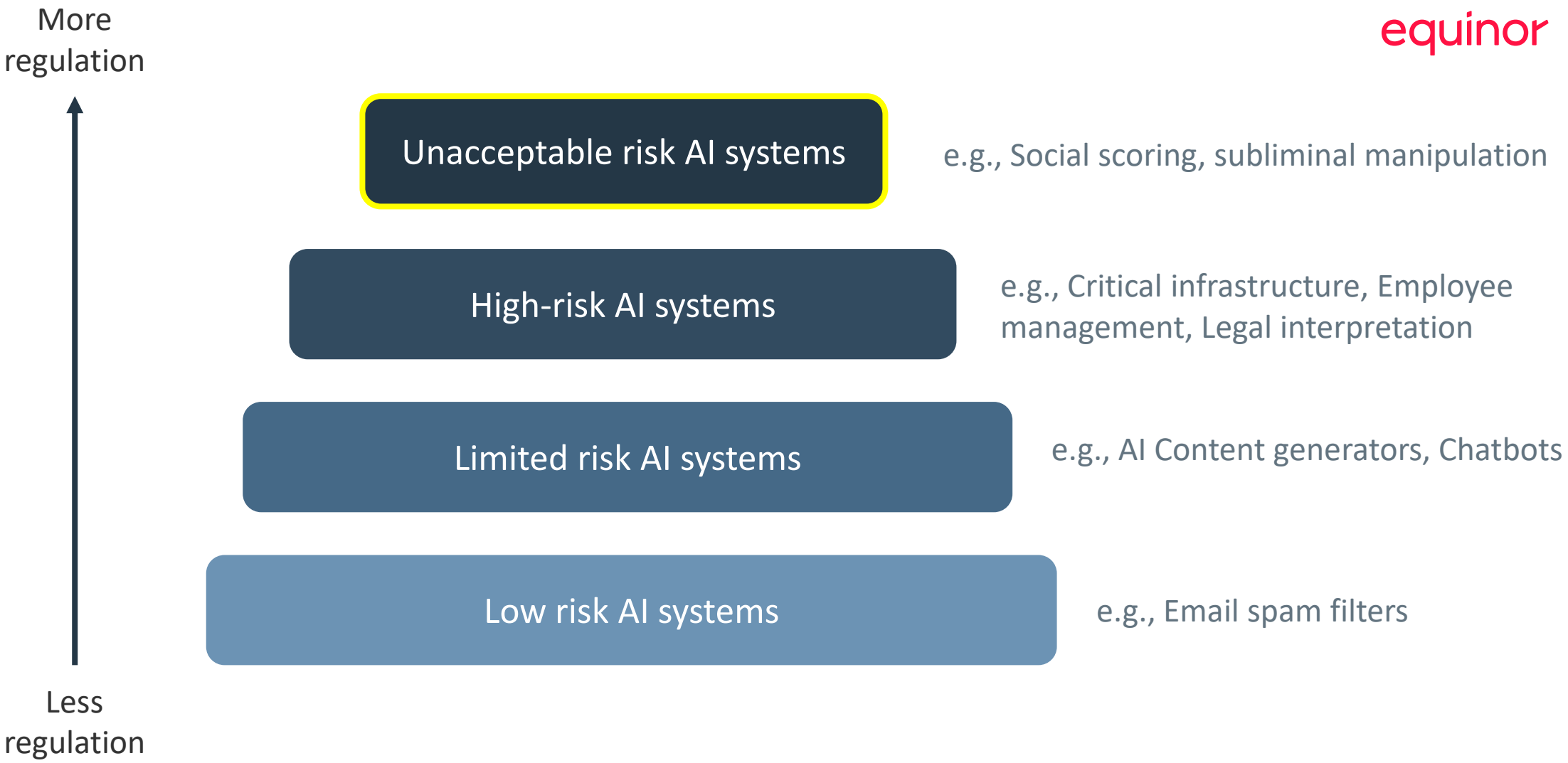
- Intended **purpose** of AI system
- Potential extent of the **harm**
- The extent to which harmed persons are in a **vulnerable** position
- The extent in which the outcome of the system is **reversible**
- The extent in which **existing legislation** provides for effective **measures** to address and minimize the risks

Ref. EU AI Act

Design requirements for High-risk AI Systems



EU AI Act



Ref. EU AI Act

EU AI Act

Non-compliance

Non-compliance fines up to 35 million EUR
or **7%** of a company’s annual turnover

Ref. EU AI Act

Preparing for compliance

- ✓ **AI Register/Database**
- ❑ AI Risk Management
- ❑ AI Governance Framework and Guidelines

AI Register → High-Risk AI Systems Database

Internal database listing AI systems (In-house, Third Party, Vendor) and their documentation.

EU AI Act:
Requirement to register all high-risk AI systems in EU AI Database of High-Risk AI Systems, along with documentation

Preparing for compliance

- ✓ AI Register/Database
- ✓ **AI Risk Management**
- ❑ AI Governance Framework and Guidelines

AI Risk Library, Risk Taxonomy

- First step towards risk assessments
- Internal AI risk landscape and a focus area heatmap
- Assist audit, procurement and product stakeholders

AI Incidents

Any outcome of the system that could cause harm

equinor

Attacks

Backdoors
Data Poisoning
Model extraction
…

Intentional abuse

Ethnic Profiling
AI-enhanced Cybercrime
Misuse of autonomous bots
…

Failures

Data Drift
Discrimination
Opaqueness
…

Adopted from Machine Learning for High-Risk Applications, Approaches to Responsible AI (Patrick Hall, James Curtis, Parul Pandey) O'Reilly 2023

EU AI Act Requirement: AI Serious Incident Reporting

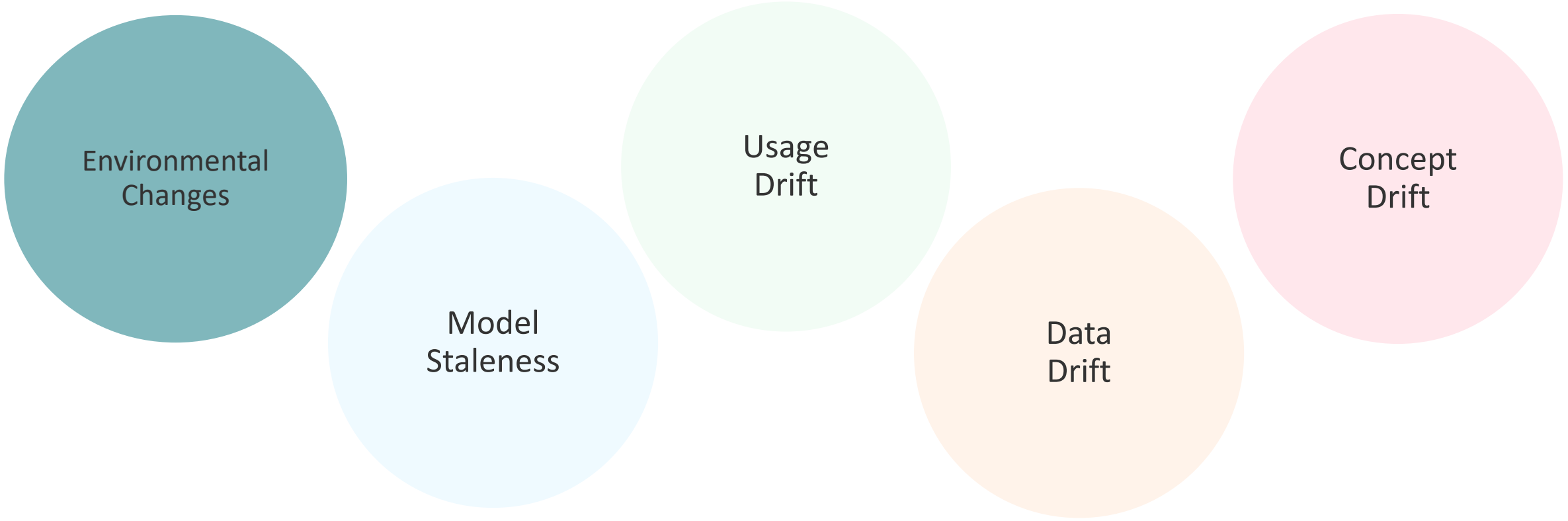
Example

AI Model Decay and Ripple Effects

An AI system in development may ***not*** be the same once *operational*.

equinor

Degradation of model's performance over time
Gradual or sudden



Model Staleness

Example

AI Model Decay and Ripple Effects

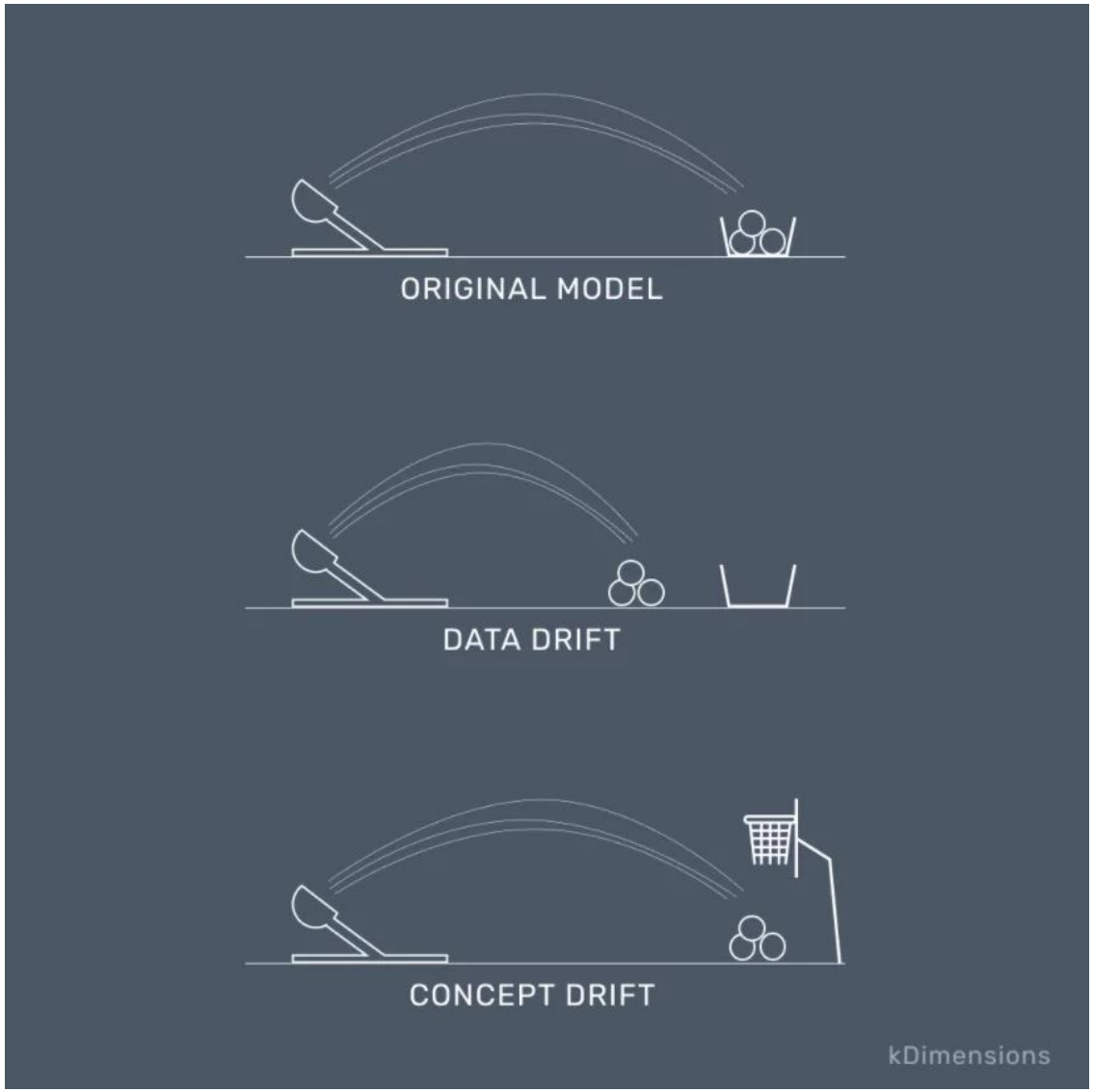
Aging algorithms and outdated assumptions;
Data used to train models does not represent current reality;

Historical bias in AI systems

Figure source" Australian Human Rights Commission
https://humanrights.gov.au/about/news/media-releases/historical-bias-ai-systems, Accessed on 08.07.2024

Example

AI Model Decay and Ripple Effects



Data Drift

Change in statistical characteristics of the **input data**;
The input data has changed rendering the trained model irrelevant on new data;

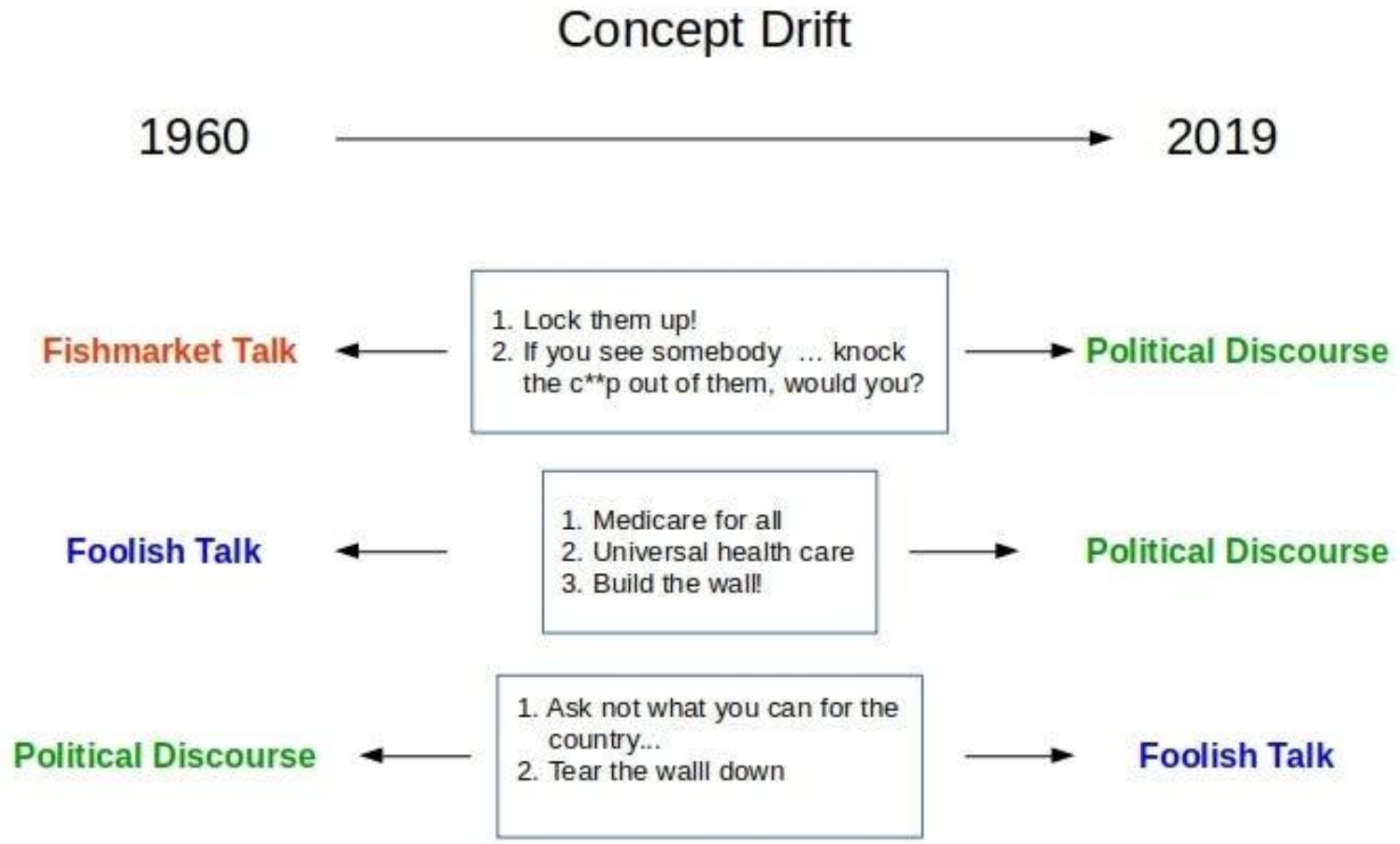
Concept Drift

Changes in the underlying ***relationships*** in data;
Objective change = what we are trying to predict has changed;

Figure source: https://medium.com/@vedantasingh28/sales-forecasting-on-given-historical-data-85c18f70d39f, Accessed on 08.07.2024

Example

AI Model Decay and Ripple Effects



Data has not changed, but our interpretation and assigned class have

Source: Concept Drift and Model Decay in Machine Learning, Ashok Chilakapati, Medium, Apr 25 2019

Mitigating risks

- EU AI Act: Risk Management System
- National Institute of Standards and Technology (NIST): AI Risk Management System
- ISO/IEC 23894:2023 Artificial intelligence - Guidance on risk management
- ISO/IEC 42001 Artificial intelligence – Management system
- US Federal Reserve SR 11-7 Model Risk Management
- …

EU AI Act Requirement for High-Risk AI Systems

Implementing Risk Management System

Risk analysis

Gather data:
1. Initial project plan
2. User requirements
3. Data from previous projects / versions
4. Data on similar systems
5. Standards

Draft statement:
1. Intended users
2. Expected levels of user experience
3. Use environment
4. System functionalities
5. Anticipated effect on the environment
6. Operating principles

Examples of use scenarios

Document identified:
1. Foreseeable use errors
2. Foreseeable intentional misuse and abnormal use

Risk evaluation

Estimate risk levels for every hazardous scenario and identify scenarios with high risk levels

Do a thorough analysis of high-risk scenarios, identify correlations and causations to help design adequate mitigation measures

Do a cross-impact analysis of all identified risks

Re-evaluate all high-risk scenarios

Risk mitigation

Identify potential mitigation measures including changes to user interface as well as other changes

Identify necessary additional safety information in the 'Instructions for use'

Assess effectiveness of individual measures

Document residual risks and corresponding risk levels

Formative evaluation of residual risks

Summative evaluation of residual risks

Residual risk acceptability evaluation

Assess and document residual risk evaluation and justify why these are acceptable

Consider conducting AI sandbox testing

Consider conducting real-world testing

Risk management report

Provide a risk management report

Design monitoring processes and define thresholds for re-assessment

Define fixed timeframes for mandatory re-assessments

Update report as necessary

Usage instructions, documentation, EU Register

Risk management report as a part of Technical documentation

Draft appropriate 'Instructions of use' based on risk assessment

Design training courses / materials if necessary to mitigate risks

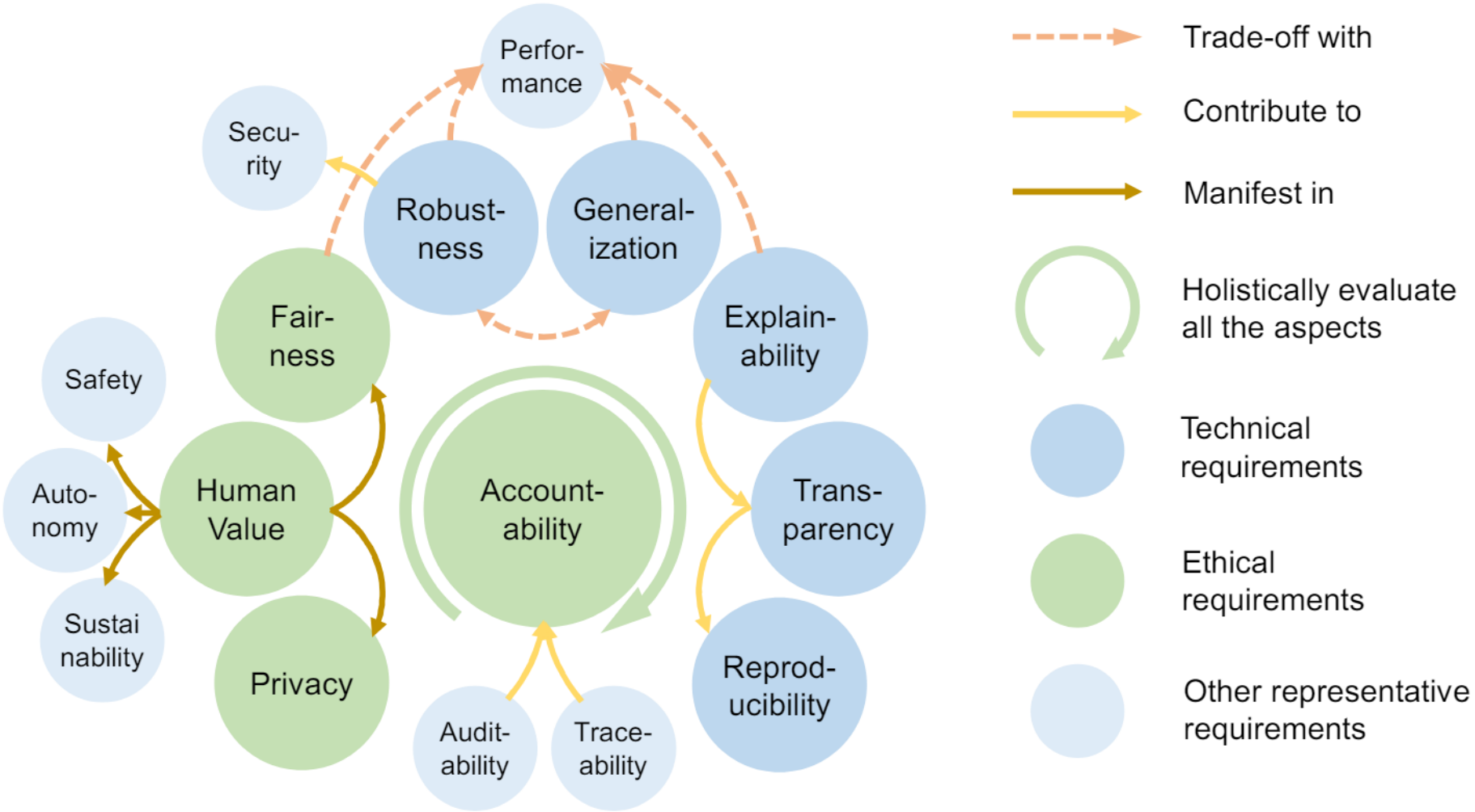
Submit risk management report when registering the system in EU Register

Adopted from Tea Mustac, European Law Institute, 2024

Moving beyond obvious risk

AI Trade-Offs

Prioritization of Responsible AI Practices



Source: Trustworthy AI: From Principles to Practices. Li et al. 2021

Cultural competencies for AI Risk Management

Organizational accountability

written policies and procedures
effective challenge
accountable leadership

Organizational processes

forecasting failure modes and knowing past failures
deliberating on **who** (customers, stakeholders),
what (well-being, opportunities),
when (frequently, over long period of time),
how (immediate response, altering processes)

Culture of effective change

diverse and experienced teams
challenging, evaluating, assessing at each step of the AI lifecycle

Partially Adapted from Machine Learning for High-Risk Applications, Approaches to Responsible AI (Patrick Hall, James Curtis, Parul Pandey) O'Reilly 2023

Preparing for compliance

- ✓ AI Register/Database
- ✓ AI Risk Management
- ✓ **AI Governance Framework and Guidelines**

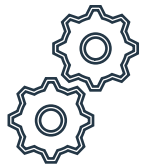
Responsible AI guidelines and best practices

Guidelines for each major process in AI lifecycle aligned with regulatory obligations, internal governance and best practices

- Ideation
- Development
- Procurement
- Deployment
- Operations and monitoring
- Use

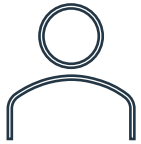
Lessons learned

AI is used to do things faster, more consistent and at scale
→
AI incidents can occur **fast**, **consistently**, and **at scale**



- ❑ Performance metrics
- ❑ Monitoring and continuous monitoring
- ❑ Regular model retraining
- ❑ Adaptive algorithms

Governance → Drastic increase in materiality of AI systems must be
followed by drastic increase of AI governance and safety measures



- ❑ Anticipate changes
- ❑ Feedback loops
- ❑ Human in the loop
- ❑ Continuous upskilling

Way ahead and what to expect

EU AI Act Timeline

- 6 months for prohibited AI systems.
- 12 months for GPAI.
- 24 months for high-risk AI systems under Annex III.
- 36 months for high-risk AI systems under Annex II.
- EU AI Office Codes of practice shall be ready 9 months after entry into force

EU AI Office work in progress

- facilitate compliance process

  (for high-risk AI systems- no later than 18 months after entry into force)
- facilitate the drawing up of codes of practice
- facilitate detection and labelling of artificially generated or manipulated content
- The EU Commission will provide guidelines and examples of high-risk and non-high-risk AI systems and add or remove conditions for high-risk classification based on evidence.

Image source: Wikipedia

Intelligent Maintenance Conference (IMC) 2024
Lausanne, Switzerland

Thank you

Responsible AI Development and Legal Compliance:
Navigating the New Landscape of AI Regulation

Rialda Spahic, PhD
Task Manager of Responsible AI, Equinor

Image source: Equinor Media Bank