



Ethical and Trustworthy AI

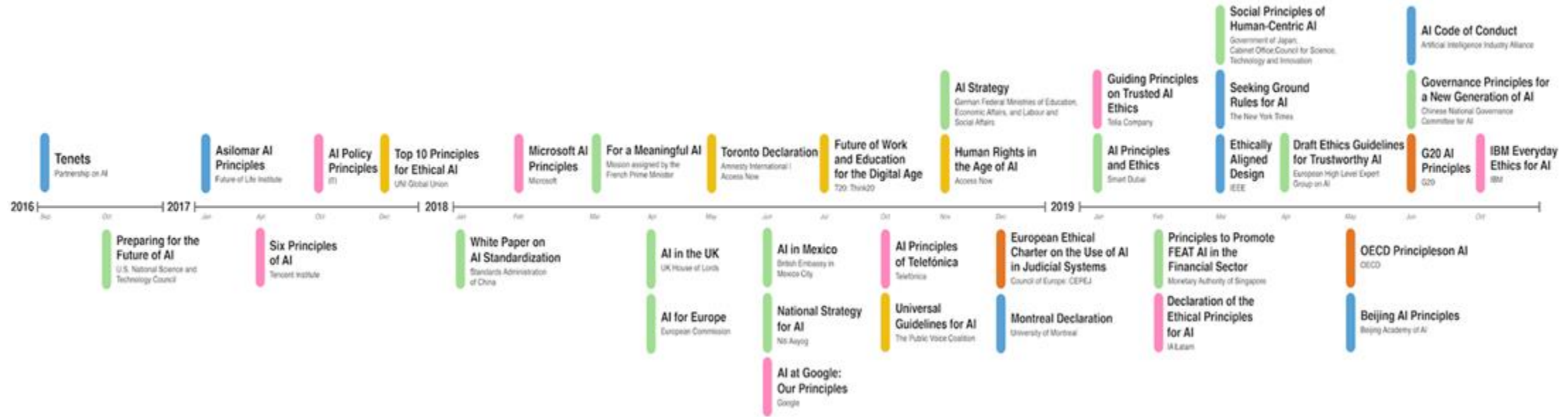
EBDVF 2023

Agenda

1. Intro on the **etami** task force
2. **Three use-cases** presented by etami members
3. Keynote on the **AI Act**
4. Panel discussion on all above

Situation

Back in 2019



Complication

On YouTube, a network of paedophiles is hiding in plain sight

Score
comm
distur

RETAIL | OCTOBER 11, 2018 / 1:04 AM / UPDATED 2 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

04-22-15 | STRONG FEMALE LEAD

The Hidden Gender Bias In Image Search

A new study finds that the stereotypes and underrepresentation in Google image searches for some careers affect unconscious bias.

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals – and highlights ways to correct it.

AI and me: friendship chatbots are on the rise, but is there a gendered design flaw?

Most chatbots are designed by men and tend to replicate gender stereotypes. But as the few women involved in the industry can testify, getting AI to emote requires input from all genders

Humans and technology | Jan 29

YouTube's algorithm seems to be funneling people to alt-right videos

Algorithms define kids by postcode: UK election results chaos reveal too much reliance on analytics

ed: AI, ML, and data analytics are valuable tools -- but they can be too easily lost.

Solution

etami (2020 – 2022)

The initial consortium

ABB

Atos

AVL 

Continental 

DB

T..

 LEONARDO

SIEMENS

VOLKSWAGEN
AKTIENGESELLSCHAFT

 zalando

 German
Research Center
for Artificial
Intelligence

UNTER
NEHMER  INITIATIVE FOR
TUM APPLIED ARTIFICIAL
INTELLIGENCE

 Eötvös Loránd
Universität

KU LEUVEN 
CENTRE FOR IT & IP LAW

 TU
berlin

MIRMI

 TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

etami

today

countries

24

120

organisations

64

people



etami today

What we do?

Quality

Translate the ethical guidelines into actionable practices and methods

European

Define and promote the EU branding for trustworthy AI

Legal

Support with the AI Act compliance

etami today

How we do it?

Collaborations

Industry x Academia; horizontal, and vertical

Research

Interdisciplinary research teams

Publications

White papers, academic papers, an open guidebook

Communicate

Awareness campaigns

etami

outputs

RESEARCH-ARTICLE

RESEARCH-ARTICLE

AI Regulation Is (not) All You Need

Authors: Laura Lucaj, Patrick van der Smagt, Djalel Benbouzid [Authors Info & Claims](#)

FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency • June 2023 • Pages 1267–1279
• <https://doi.org/10.1145/3593013.3594079>

Published: 12 June 2023 [Publication History](#)

RESEARCH-ARTICLE OPEN ACCESS

RESEARCH-ARTICLE OPEN ACCESS

A sector-based approach to AI ethics: Understanding ethical issues of AI-related incidents within their sectoral context

Authors: Dafna Burema, Nicole Debowski-Weimann, Alexander von Janowski, Jill Grabowski, Mihai Maftel, Mattis Jacobs, Patrick van der Smagt, Djalel Benbouzid [Authors Info & Claims](#)

AIES '23: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society • August 2023 • Pages 705–714
• <https://doi.org/10.1145/3600211.3604680>

Published: 29 August 2023 [Publication History](#)

etami open guidebook

etami open guidebook

About
Principles
Contribute
Licence

Concepts
Glossary
Accountability
Legal Sandboxing & Policy Prototyping
Law & AI
Guide: AI & GDPR
Guide: ALTAI & law
AI Act & documentation
Policy Prototyping the AI Act

Trustworthiness by Design
Quality-Centric Lifecycle Models
Documentation
Risk Taxonomy
Assurance Cases
AI Security

Auditing
Process

etami The open guidebook on legal, trustworthy, and ethical Artificial Intelligence

about

This guidebook offers a window onto the activity of the etami consortium. It covers actionable guidelines to develop legal, trustworthy, and ethical Artificial Intelligence. The focus is put on quality-centric lifecycle models for AI systems, legal compliance, and AI auditing practices.

etami - Ethical and Trustworthy Artificial and Machine Intelligence - is a non-profit organisation that works on making **ethical AI principles** actionable. etami understands that the best path towards trustworthy and ethical AI starts with quality-centric lifecycle models, processes involving all the stakeholders, and auditing practices. By developing systems with care while following a number of simple principles, it is straightforward to make such systems safe and trustworthy, and follow ethical and legal principles - even where these change. This handbook, of which you are only able to see a snapshot in time, lies down these principles. If you are researching, developing, or applying AI methodologies, this handbook is for you.

[More about etami](#)

overview the open guidebook

This guidebook, which is a continuing effort, is structured around four sections. The first section, the most theoretical one, includes a **glossary of terms** and defines **concepts** that are useful to reason about accountability. The rest of the guidebook can be divided into trustworthiness by design practices, such as a redefined **lifecycle models** and **documentation methods**, **AI auditing procedures**, and legal toolboxes.

etami

Public seminar series

session #1:

Legal requirements for explanation of decisions: what makes an explanation?

session #2:

GDPR in the context of machine learning

session #3:

Data Regulation

session #4 to come soon in 2024



Lessons learnt



The diagram is structured as a house. At the top is a grey triangle with the text 'Trustworthy AI pillars'. Below it is a grey horizontal bar with 'AI Governance'. The middle section consists of four trapezoidal pillars: 'Risk Management' (orange), 'Compliance processes and structure' (teal), 'Platforms integration' (blue), and 'Communication & training' (purple). At the bottom is a purple horizontal bar with 'AI legal compliance'. A yellow shape is at the very bottom of the page.

Trustworthy AI pillars

AI Governance

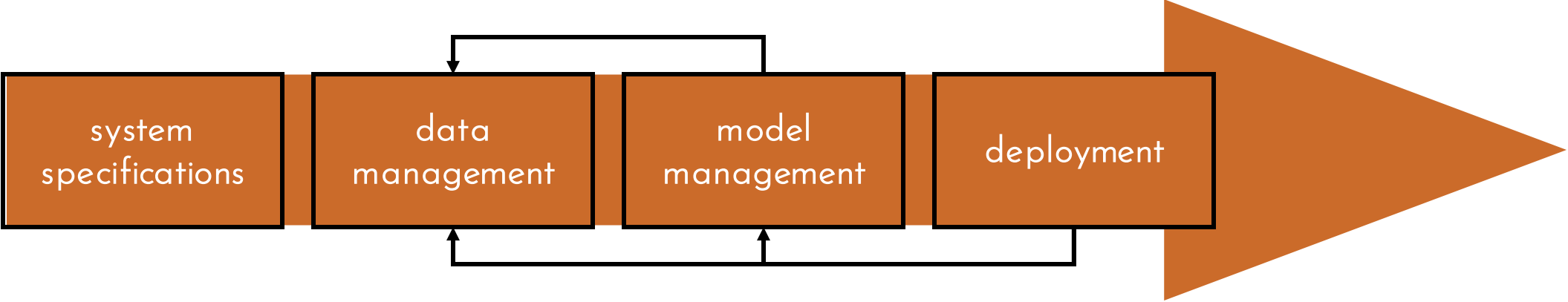
**Risk
Management**

**Compliance
processes and
structure**

**Platforms
integration**

**Communication
& training**

AI legal compliance



**System
specification
& evaluation**

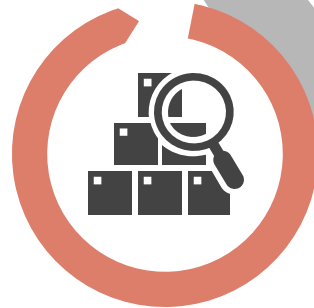


**Data
management**

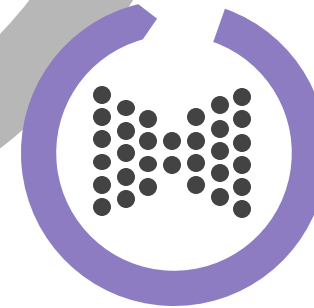


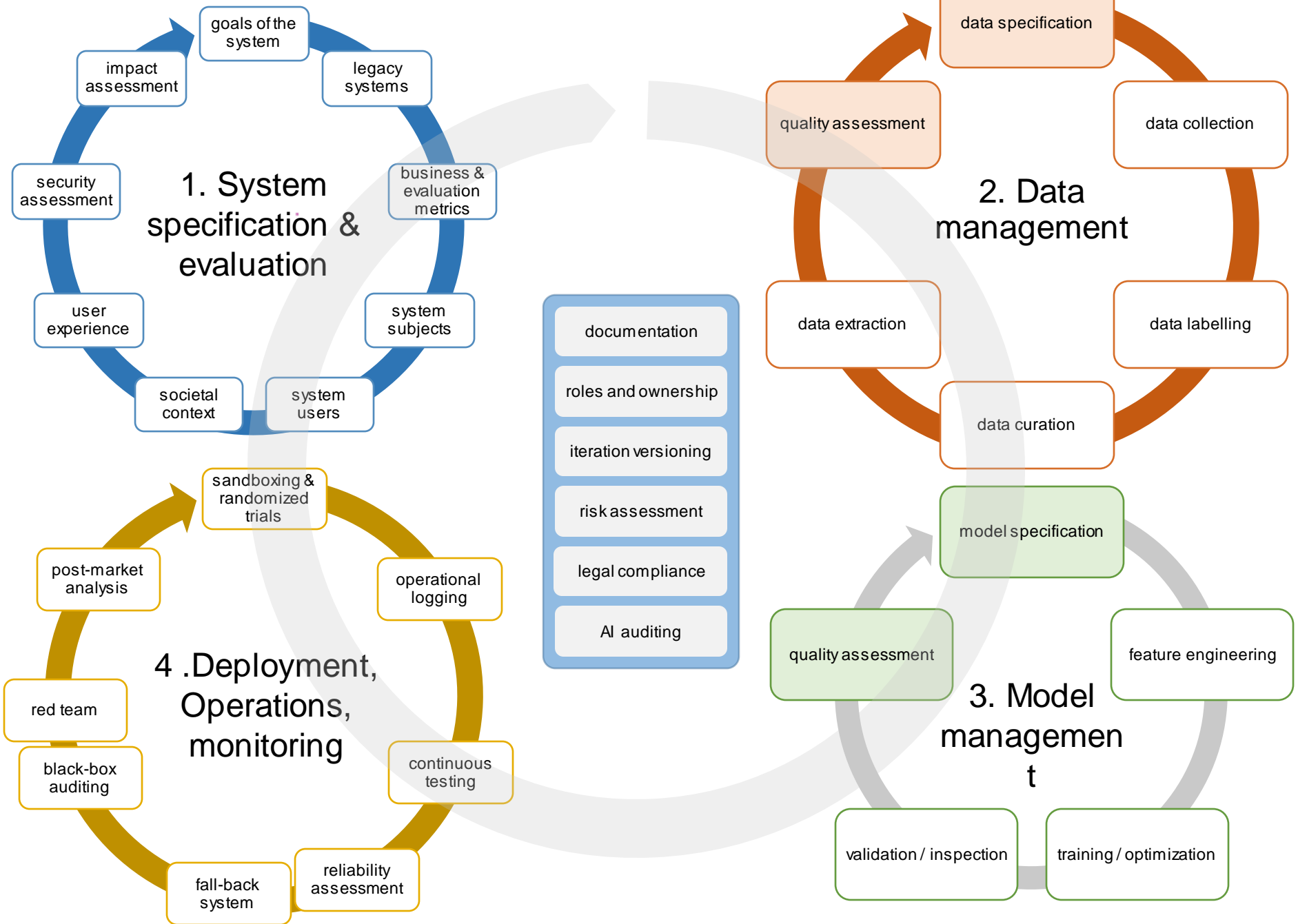
**Quality-centric
Lifecycle model for
ML systems**

**Deployment,
Operations,
monitoring**



**Model
management**





AI Act compliance

Pro-active implementation thereof

AI definition

Validated by the legal department

Understood by the AI practitioners

AI Stocktaking

List and track all AI systems

Complex for large organizations

Risk Classification

Non-centrality of the topic in the current text

Needs a validated "flowchart"



Other lessons learnt

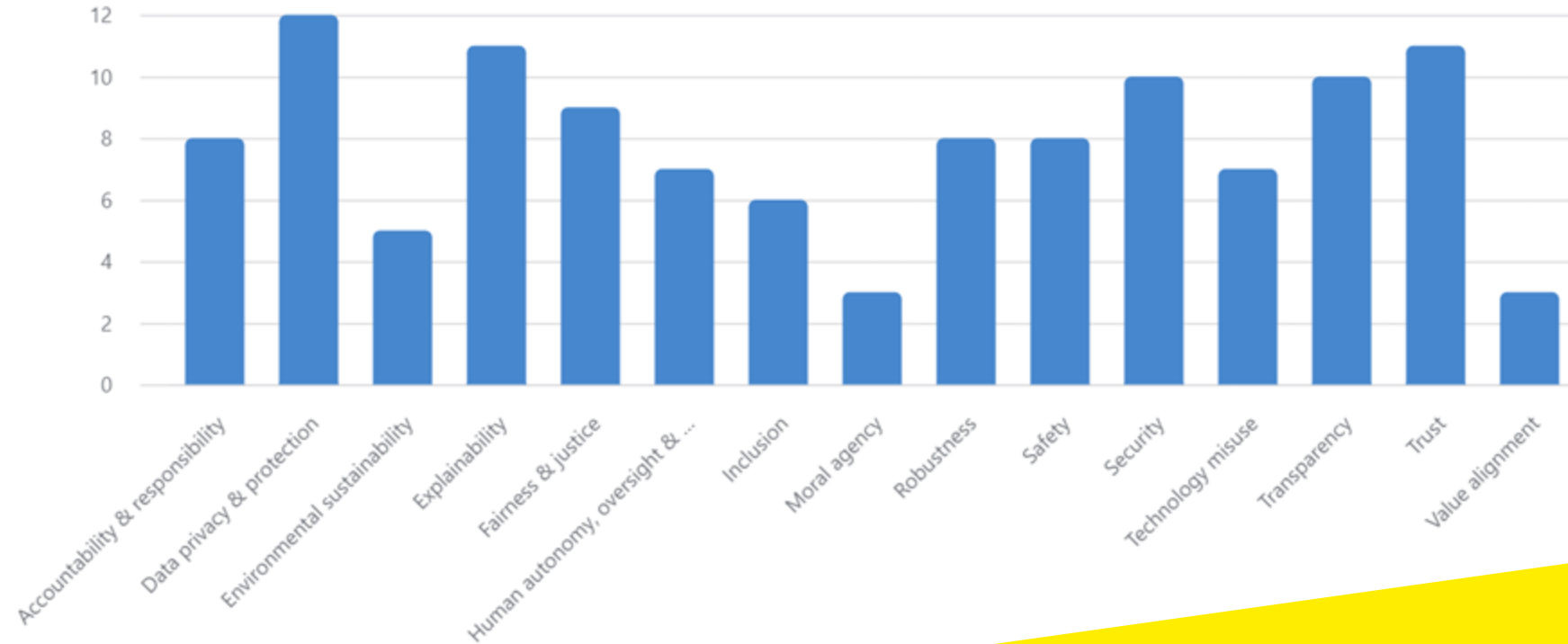
Ethical, Trustworthy, and Legal AI Techniques & Methodologies

What measures have you implemented/are you planning to implement to ensure compliance with the European AI act?

- **Systematic analysis** of regulatory documents (incl. categorization of AI-based products; defining a roadmap towards achieving compliance)
- Organisation-wide **frameworks** (w.r.t. ethical and regulatory compliance)
- Organisation-wide **policies** for managing AI-related systems (incl. control mechanisms)
- Appointment of a **Data Protection Officer / AI Compliance Officer**
- **Data governance** (incl. Cleaning, bias detection, anonymisation of personal data aspects that might cause discrimination, data storing terms and conditions)
- Implementation of **procurement policies** and **processes** for AI systems and solutions checking against compliance with **national legislation** and **EU GDPR**
- AI system **misuse, bias, and factual correctness** assessment
- Software **development, testing and validation** aligned with **EU regulatory requirements** (documentation approved by the EU Ethics Committee)
- **Trainings** (incl. on AI incident communication)

Ethical, Trustworthy, and Legal AI Techniques & Methodologies

If yes, what are the ethical principles associated with the need for the adaptation? (n=14)



Ethical, Trustworthy, and Legal AI Techniques & Methodologies

If you have already started addressing ethical and legal issues w.r.t. your AI systems, how?
(n=14)

