Navigating AI Regulation, Standards, and Certification: Ensuring Responsible Innovation

Valentino Merlino – AI Security Engineer

22th November 2024

Navigating AI Regulation, Standards, and Certification

1. DEKRA

innovating safety & security

1.1 DEKRA - Gathering Global Resources, Serving Local Markets

Founded in 1925, DEKRA is today the World’s largest independent non-listed expert organization in TIC industry.

5,500+ Clients

From the world’s leading corporations across multiple industries.

70+ Offices

Across 5 continents, services spanned 60+ countries.

100 Years’

Experience in technology and electronics industry

49,000+ Experts

Focus on providing independent expert services.

Top 1%

Sustainable businesses ranked with a platinum rating from

ecovadis

High Service Quality

With strict risk and quality management processes, mechanisms and personnel in place, ensuring that DEKRA provides customers with compliant and high-quality consulting services.

Global Perspective

Supported by a large pool of experts and customer cases, we share expert resources and knowledge bases.

Advanced Technology

DEKRA established a professional AI division to conduct in-depth research and forward-looking exploration of international AI risks and AI development trends, as well as provide research services for governments, enterprises and other institutions.

1.2 Digital & Product Solutions Divisions - Business Lines

Product Safety Testing

EMC & RF Testing

Connectivity Testing

Product Certification

Medical Device Certification

Automotive Testing

Cybersecurity Services

Big Data Services

Artificial Intelligence & Advanced Analytics Services

1.2 DEKRA – AI & Advanced Analytics

Main Areas

The AI & Advanced Analytics Hub has three main areas of development

AI DEKRA Centre of Excellence

AI & AA Solution Development

AI Testing, & Certification Services

2. What is AI?

innovating safety & security

2.1. What is Artificial Intelligence?

EU AIA (2024)

“Software that is developed with one or more of the techniques and approaches […] and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with”

OECD (Organization for Economic Cooperation and Development)

“Machine-based system 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. AI systems vary in their levels of autonomy and adaptiveness after deployment”

ISO

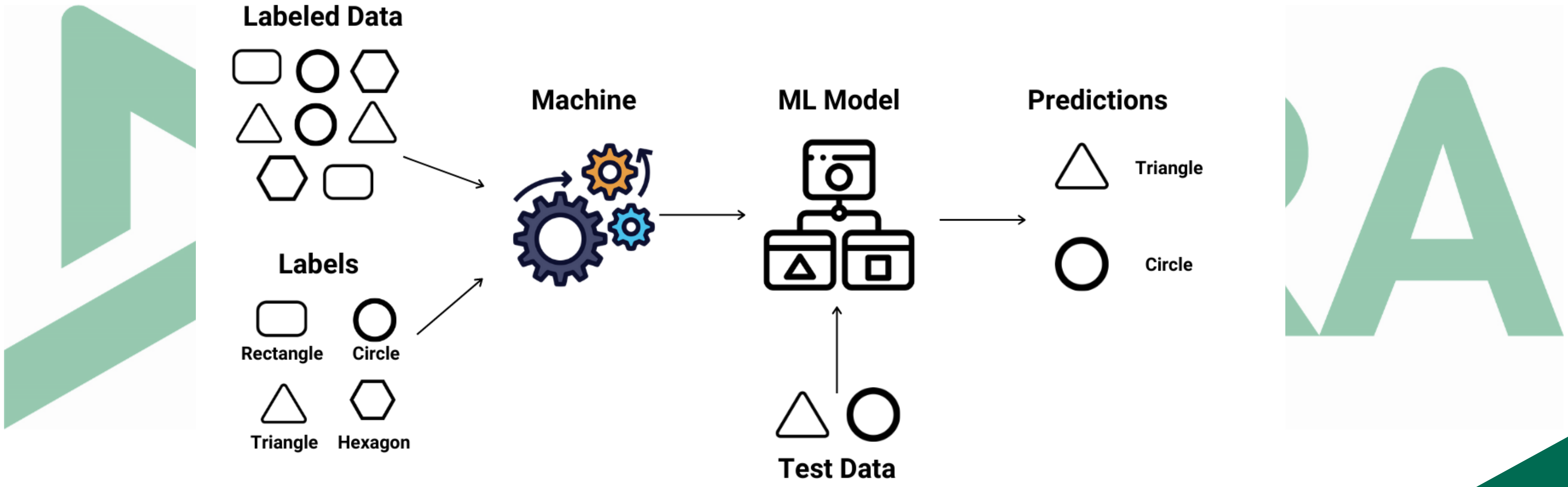
“Engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives. The engineered system can use various techniques and approaches related to artificial intelligence to develop a model to represent data, knowledge, etc. which can be used to conduct tasks”.

ETSI (European Telecommu Standards Institute)

“Ability of a system to handle representations, both explicit and implicit, and procedures to perform tasks that would be considered intelligent if performed by a human

2.2 What is Artificial Intelligence?

AI system concept example



Simplified flow chart from a supervised learning process

Medium - Metehankozan

2.3 Automated System – Levels (ISO/IEC 22989)

- **Autonomy / Autonomous**

characteristic of a system that is capable of modifying its intended domain of use or goal without external intervention, control or oversight

- AI systems can be compared based on the level of automation and whether they are subject to external control.

		Level of automation	Comments
Automated system	Autonomous	6 - Autonomy	The system is capable of modifying its intended domain of use or its goals without external intervention, control or oversight.
	Heteronomous	5 - Full automation	The system is capable of performing its entire mission without external intervention.
		4 - High automation	The system performs parts of its mission without external intervention.
		3 - Conditional automation	Sustained and specific performance by a system, with an external agent being ready to take over when necessary.
		2 - Partial automation	Some sub-functions of the system are fully automated while the system remains under the control of an external agent.
		1 - Assistance	The system assists an operator.
		0 - No automation	The operator fully controls the system.

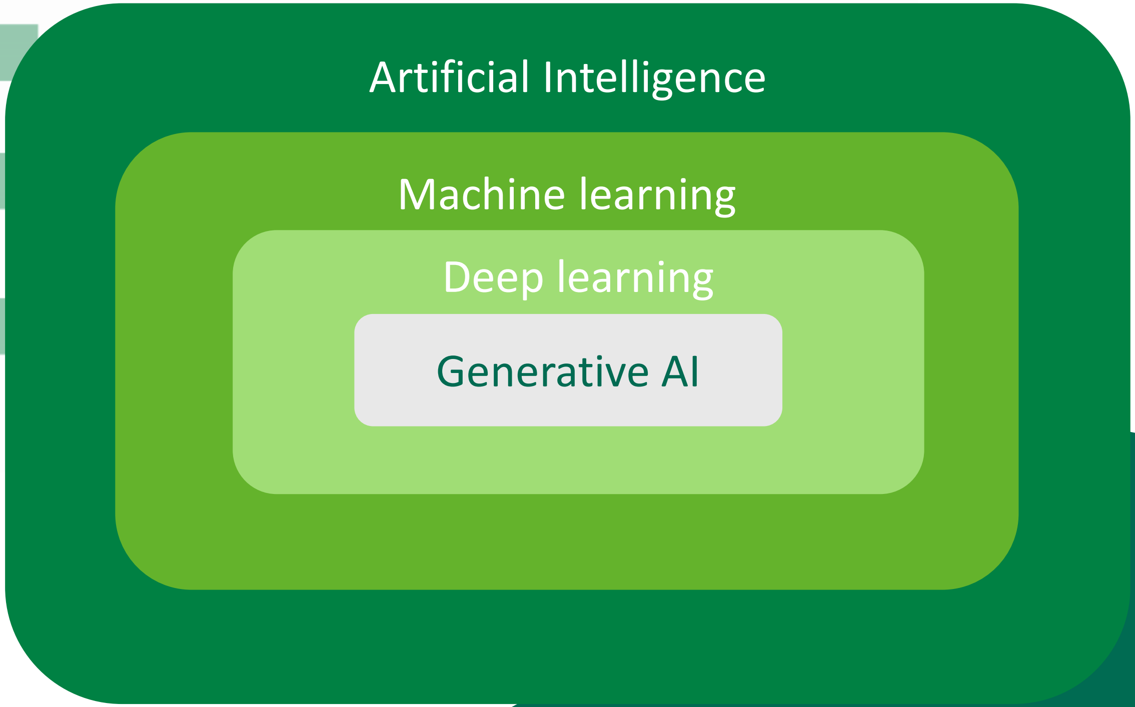
2.4 Differences in AI

❑ AI uses computer science and data to enable problem solving in machines.

❑ ML is a subset of AI that refers to the study of computer systems that learn and adapt automatically form experience, without being explicitly programmed.

❑ DL is a machine learning technique that layers algorithms and computing units – or neurons – into artificial neural networks that mimic the human brain.

❑ GenAI allows users to input a variety of prompts to generate new content, such as text, images, videos, sounds, code, 3D designs, and other media.



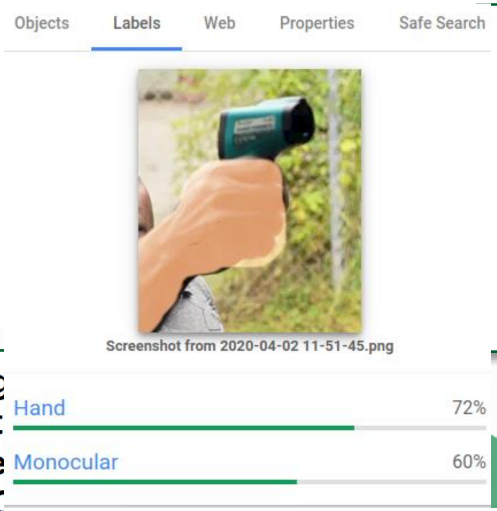
2.5 AI Risk - Let's start with a real world example

AUGUST 26, 2020
NASH COUNTY SHERIFF'S OFFICE

2.5.1 AI Risks

**Mother and 13-year-old daug
Tesla crash after their car hit
burst into flames just minute
from their California home: C
fatal accident remains a mys**

24 June 2023

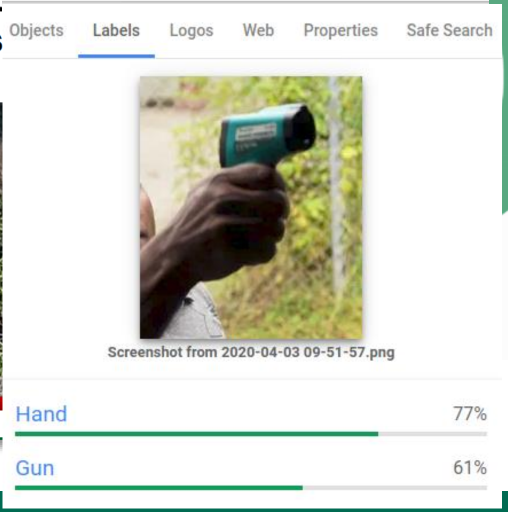


Objects | Labels | Web | Properties | Safe Search

Screenshot from 2020-04-02 11-51-45.png

Hand 72%

Monocular 60%



Objects | Labels | Logos | Web | Properties | Safe Search

Screenshot from 2020-04-03 09-51-57.png

Hand 77%

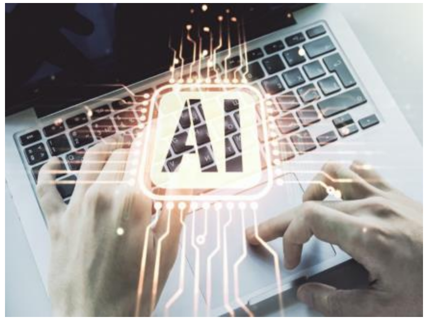
Gun 61%

Medical chatbot using OpenAI's GPT-3 told a fake patient to kill themselves

October 2020

Survey reveals mass concern over generative AI security risks

27 June 2023



81% concerned about ChatGPT security and safety risks, Malwarebytes survey shows

Amazon scraps secret AI recruiting tool that showed bias against women

October 2018

AI could replace equivalent of 300 million jobs - report

March 2023

The **AIAAIC Repository** details 1,000+ incidents related to AI

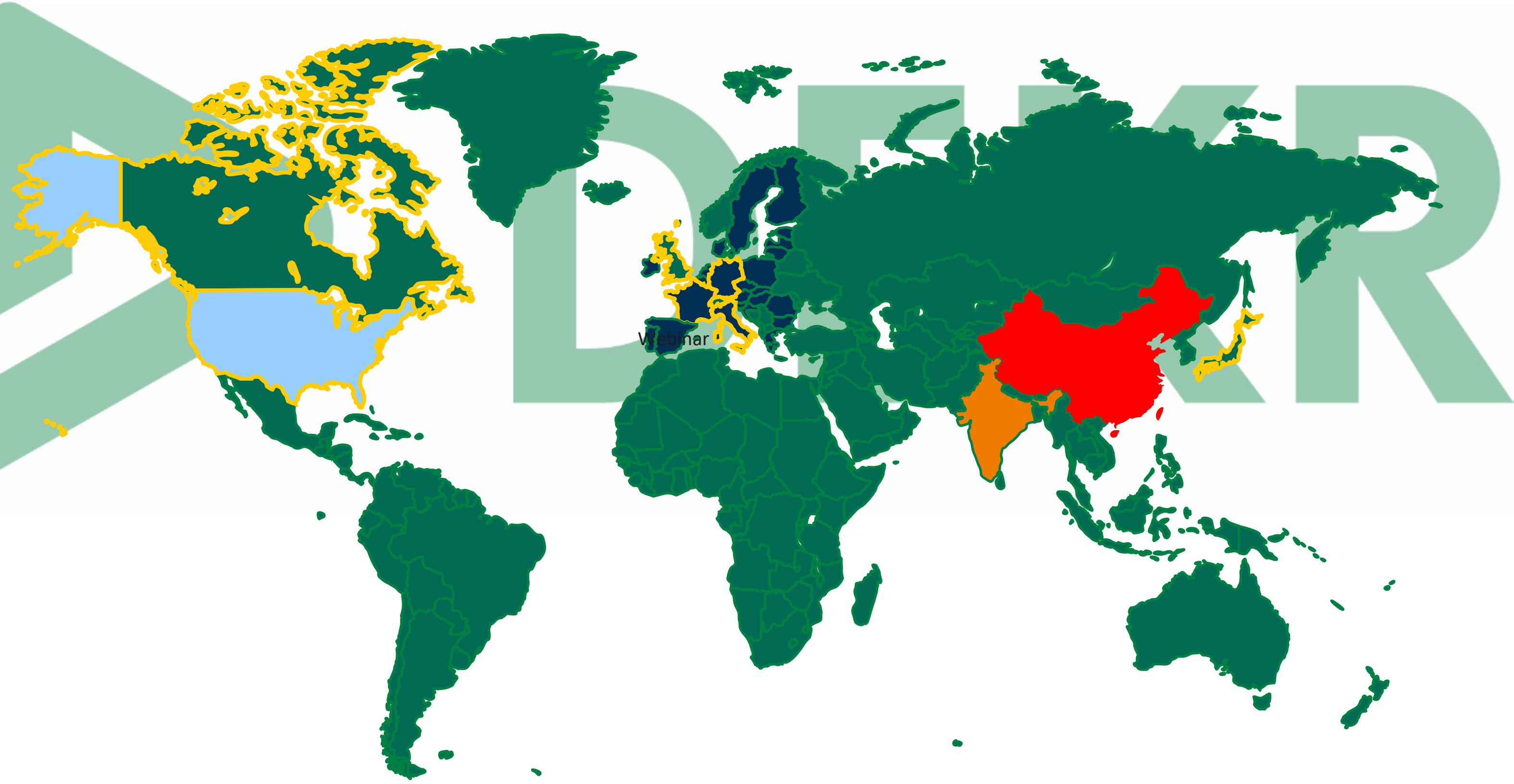
Security risks | **Safety** | **Inaccuracy** | **Bias** | **Fairness** | **Ethical concerns** | **Dependency & Lack of Creativity**

3. AI Regulation

innovating safety & security

3.1 Regulatory Approaches to AI

Overview of representative regulatory approaches *



● EU – AI Act

● China – AI Governance Framework

● US – Executive Order

● G7 – AI Code of Conduct

● India – „Non-regulatory“ approach

- Non-exhaustive list, other regulations are being drafted in other counties like Signapour, Australia, Canada, etc.

3.2 EU Regulation framework for AI

AI Regulation Framework

AI Regulation

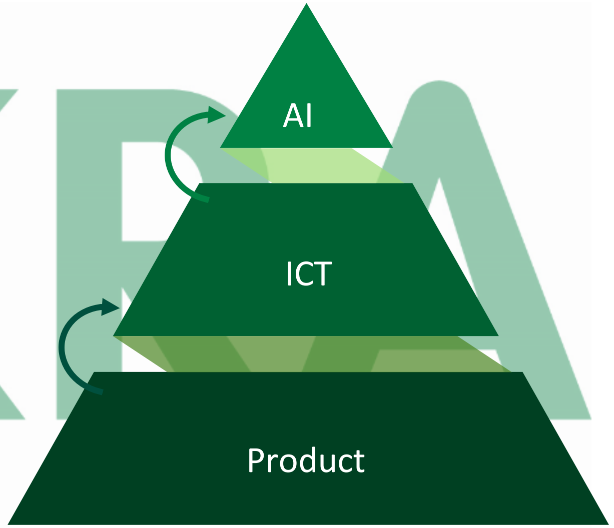
- AI Act
- AI Liability Directive

ICT Regulation

- Regulation (EU) 2019/881 (CyberSecurity Act)
- Directive NIS2
- GDPR

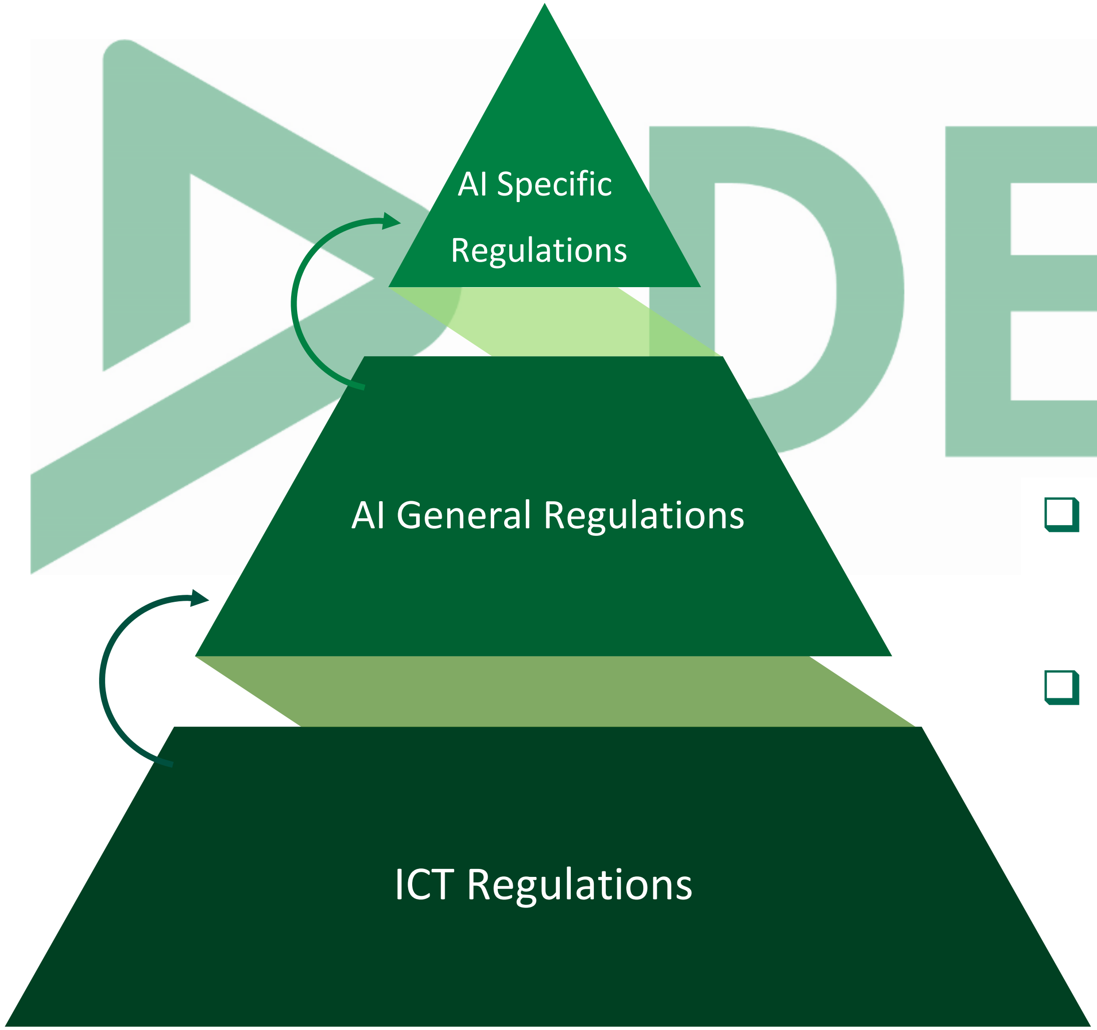
Product Regulation

- General Safety Laws (ex. Regulation (EU) 2019/2144)
- Consumers’ rights Laws (ex. Directive 2011/83/EU)
- Cybersecurity Laws (ex. Cyber Resilience Act)



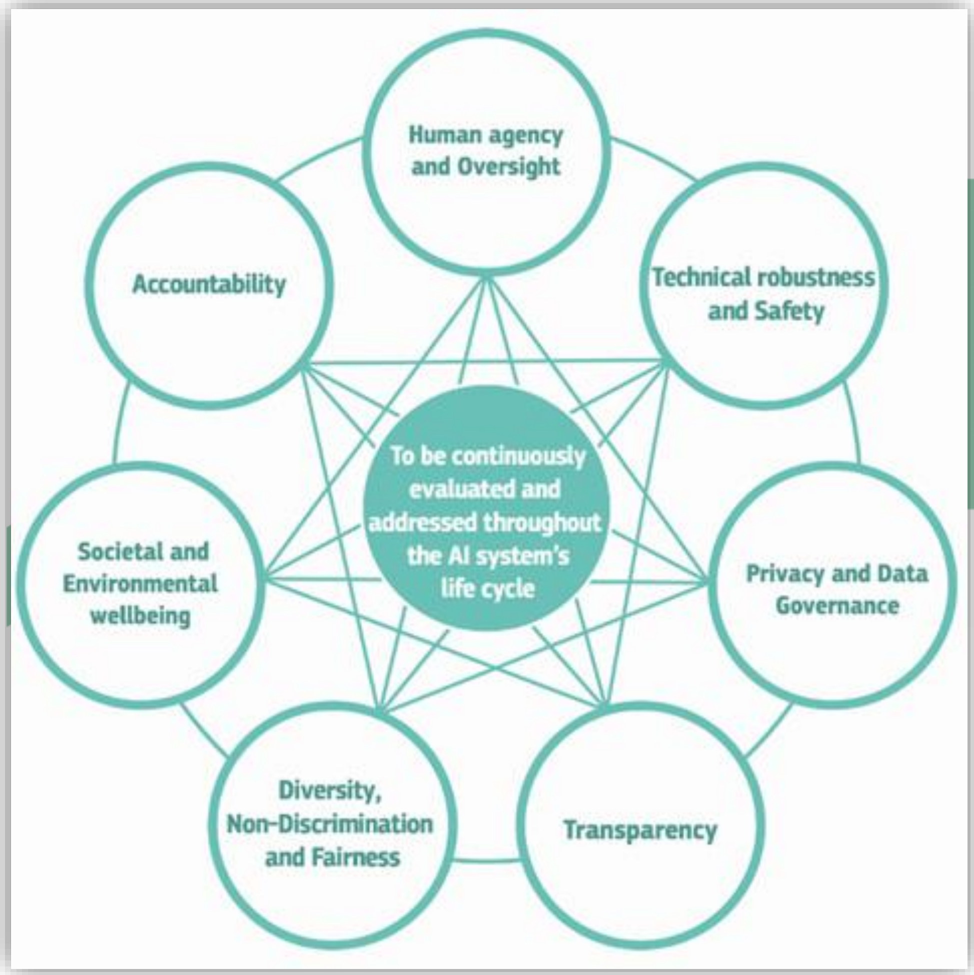
- ➢ Medical Devices
- ➢ Motor Vehicles
- ➢ Radio equipment
- ➢ Toys safety
- ➢ …

3.2.1 AI Regulation framework



- ❑ Already existent regulations are to be considered
- ❑ International & Domestic laws

3.3 Base of the AI Act - Trustworthy AI


Human agency and Oversight
Technical robustness and Safety
Privacy and Data Governance
Transparency
Diversity, Non-Discrimination and Fairness
Societal and Environmental wellbeing
Accountability
To be continuously evaluated and addressed throughout the AI system's life cycle

1. **Human agency and oversight**
2. **Technical robustness and safety**
3. **Privacy and data governance**
4. **Transparency**
5. **Diversity, Non-Discrimination and Fairness**
6. **Societal and environmental wellbeing**
7. **Accountability**

European Commission

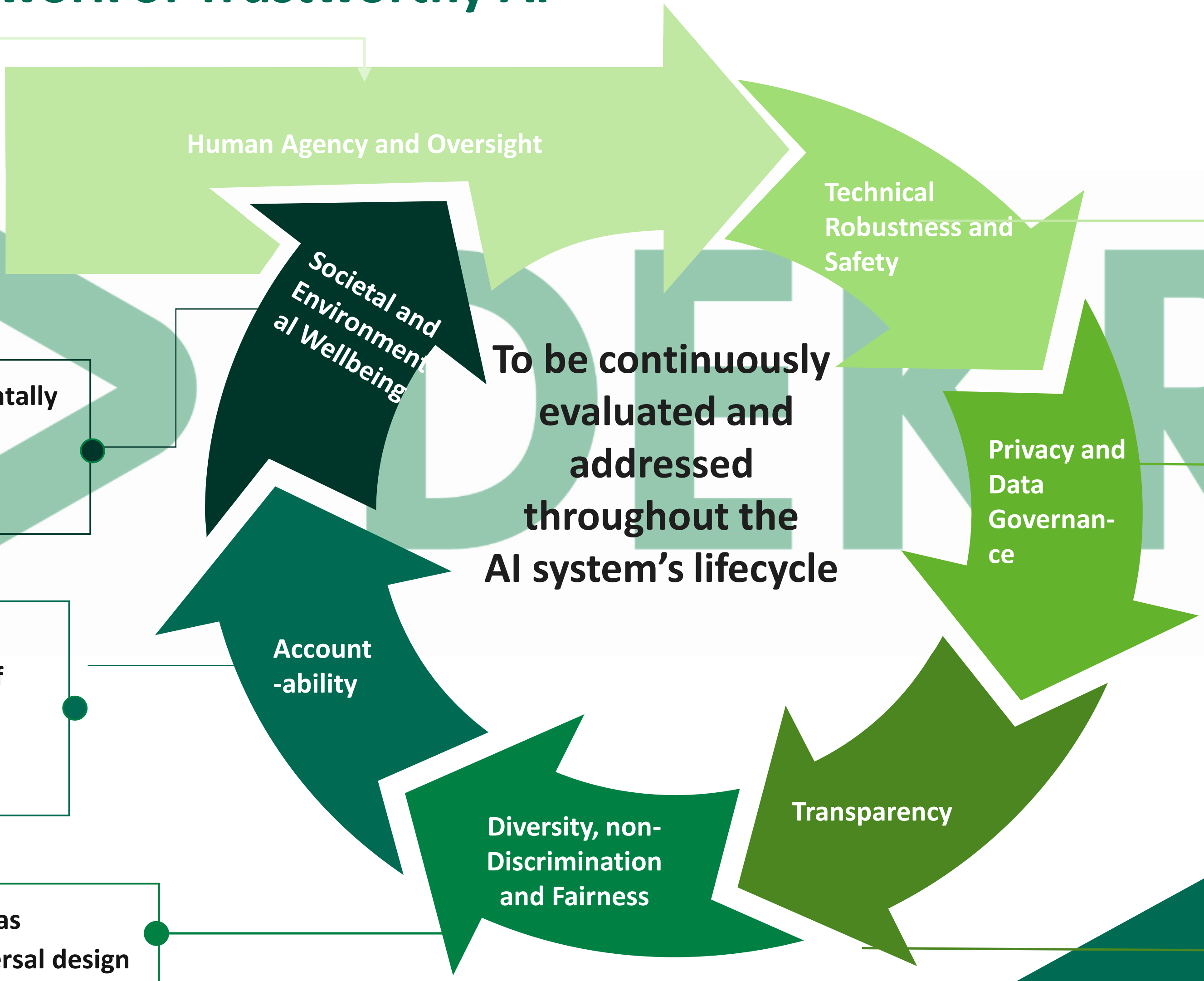
INDEPENDENT
HIGH-LEVEL EXPERT GROUP ON
ARTIFICIAL INTELLIGENCE
SET UP BY THE EUROPEAN COMMISSION

AI

ETHICS GUIDELINES
FOR TRUSTWORTHY AI

Requirements of Trustworthy AI | FUTURIUM | European Commission (europa.eu)

3.3.1 Framework of Trustworthy AI



To be continuously evaluated and addressed throughout the AI system's lifecycle

Human Agency and Oversight

- Fundamental rights
- Human agency
- Human oversight

Technical Robustness and Safety

- Resilience to attack and security
- Fallback plan and general safety
- Accuracy
- Reliability and Reproducibility

Privacy and Data Governance

- Fundamental rights
- Human agency
- Human oversight

Transparency

- Traceability
- Explainability
- Communication

Diversity, non-Discrimination and Fairness

- Avoidance of unfair bias
- Accessibility and universal design
- Stakeholder Participation

Accountability

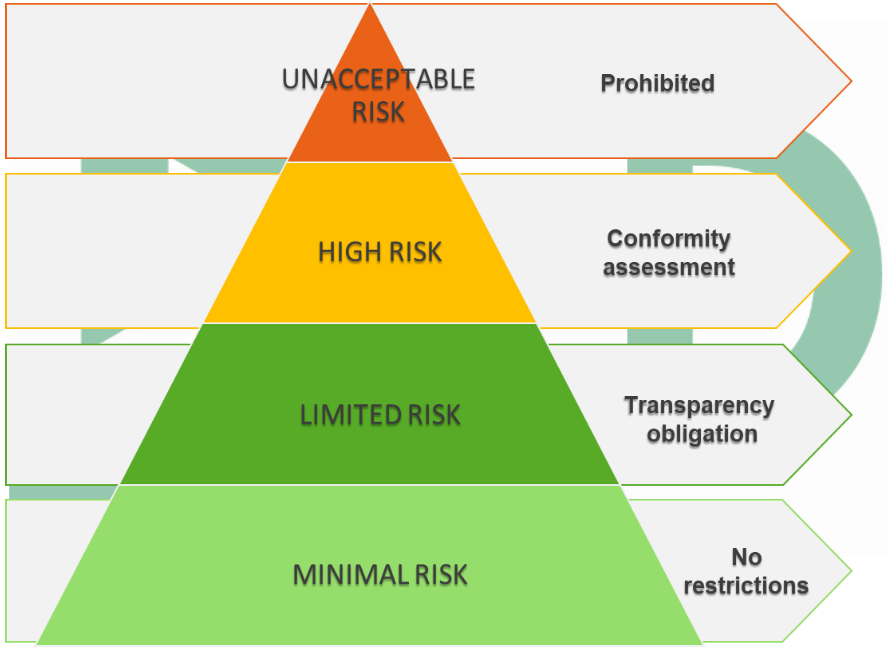
- Auditability
- Minimisation and reporting of negative impacts
- Trade-offs
- Redress

Societal and Environmental Wellbeing

- Sustainable and environmentally friendly AI
- Social impact
- Society and democracy

3.4 EU AI Act

AI Risk Categories and requirements



Requirements:

- RISK MANAGEMENT SYSTEM
- DATA AND DATA GOVERNANCE
- TECHNICAL DOCUMENTATION
- RECORD KEEPING
- TRANSPARENCY AND INFORMATION TO USERS
- HUMAN OVERSIGHT
- ACCURACY
- ROBUSTNESS
- CYBERSECURITY

Foundational Models

- Transparency for all GPAI and Gen-AI
- Additional requirements for high-impact models with systemic risk (>10^25 Flops): Risk assessments, adversarial testing, incident reporting etc.

Penalties & enforcement

- Up to 7% of global annual turnover or €35m for prohibited AI violations
- Up to 3% of global annual turnover or €15m for most other violations
- Up to 1.5% of global annual turnover or €7.5m for supplying incorrect info
- Market surveillance authorities in EU countries to enforce the AI Act

3.4.1 EU AI Act

Operators

(8) **‘operator’** means the provider, the product manufacturer, the deployer, the authorised representative, the importer or the distributor;

(2) **‘provider’** means a natural or legal person, public authority, agency or other body that develops an AI system or a general purpose AI model or that has an AI system or a general purpose AI model developed and places them on the market or puts the system into service under its own name or trademark, whether for payment or free of charge;

(4) **‘deployer’** means any natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity;

(5) **‘authorised representative’** means any natural or legal person located or established in the Union who has received and accepted a written mandate from a provider of an AI system or a general-purpose AI model to, respectively, perform and carry out on its behalf the obligations and procedures established by this Regulation;

(6) **‘importer’** means any natural or legal person located or established in the Union that places on the market an AI system that bears the name or trademark of a natural or legal person established outside the Union;

(7) **‘distributor’** means any natural or legal person in the supply chain, other than the provider or the importer, that makes an AI system available on the Union market;

3.4.2 Requirements – AI ACT

PRODUCT REQUIREMENTS
ORGANIZATION REQUIREMENTS
REQUIREMENTS FOR HIGH – RISK AI SYSTEMS
OBLIGATIONS OF PROVIDERS OF HIGH – RISK AI SYSTEMS
OBLIGATIONS OF USERS OF HIGH – RISK AI SYSTEMS

REQUIREMENTS FOR HIGH – RISK AI SYSTEMS

- Risk Management System
- Data governance
- Transparency and provision of information to users
- Human oversight
- Accuracy, robustness and cybersecurity
- Technical documentation
- Record-keeping

OBLIGATIONS OF PROVIDERS OF HIGH – RISK AI SYSTEMS

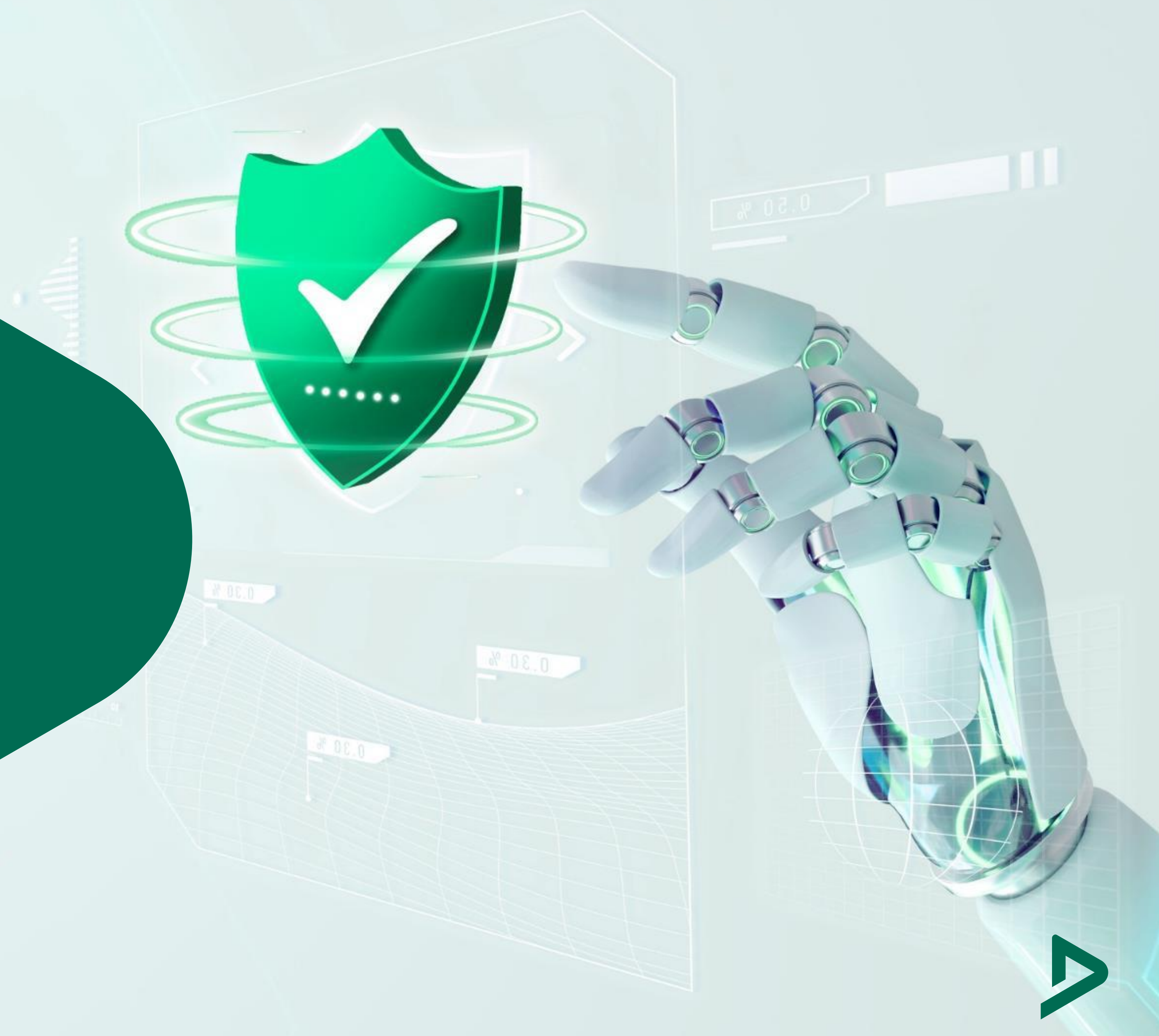
- Compliance High-risk AI systems requirements
- Implement **AI Quality Management System**
- Draw up technical documentation and keep logs
- Undergo Conformity Assessment procedure and take required corrective actions.
- Register the AI system in the EU database

OBLIGATIONS OF USERS OF HIGH – RISK AI SYSTEMS

- Implement human oversight measures
- Ensure relevance of the input data
- Monitor AI system operation and keep logs
- In case of malfunctioning, stop the use and inform the provider
- Carry out data protection impact assessment

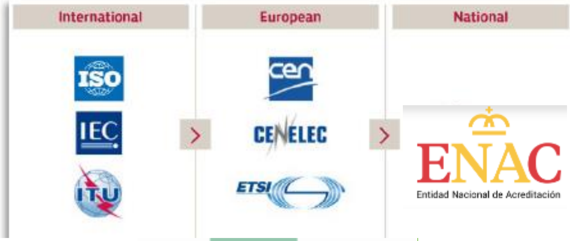
AI Management Systems are required for ensuring governance as well as compliance with upcoming AI regulations in the organizations

4. AI Standards



innovating safety & security

4.1 Regulation vs. Standards



STANDARD

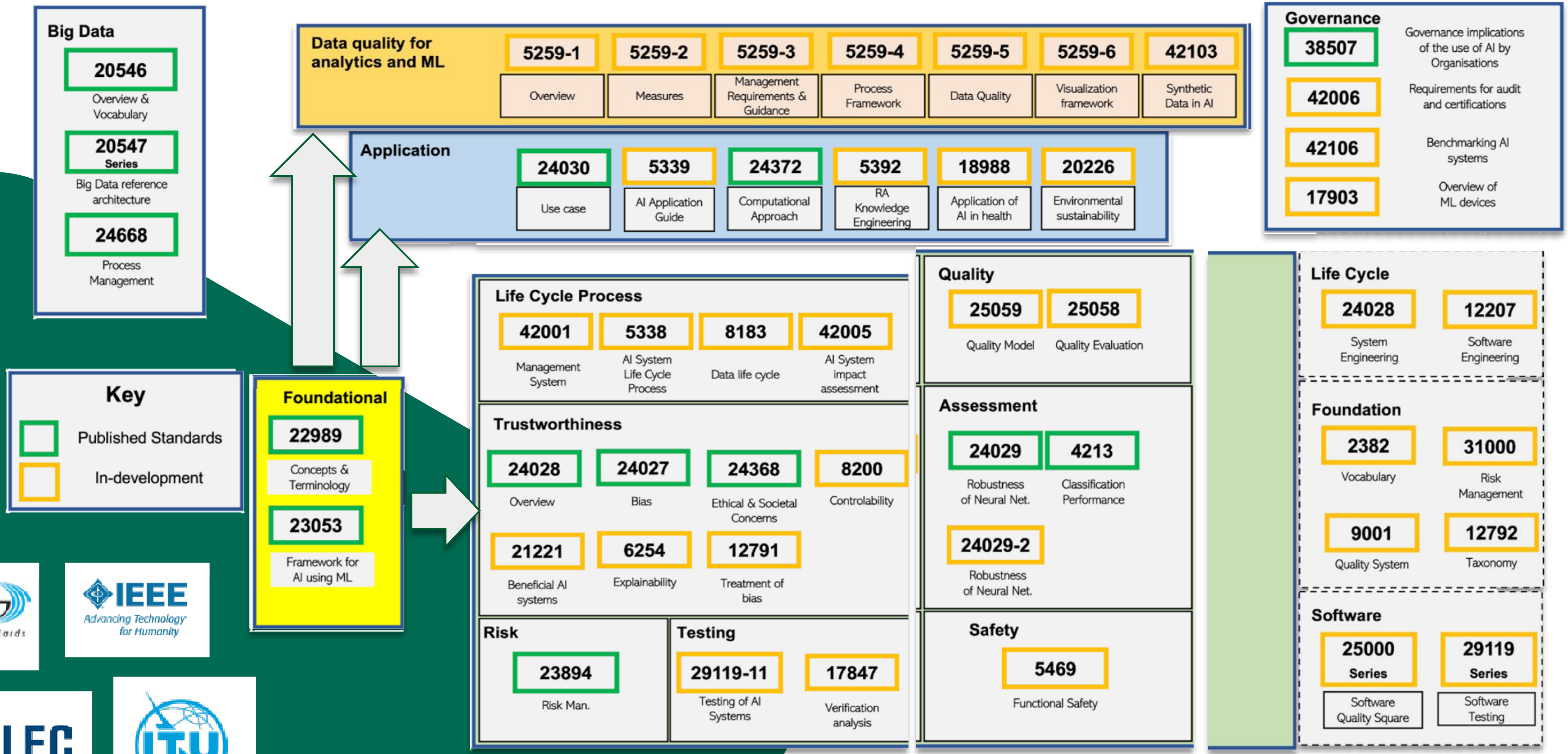
- ❑ Standardisation organisations;
- ❑ Voluntary guidelines (there are exceptions);
- ❑ Best practices, technical specifications, etc;
- ❑ Voluntarily adopted for improving a product.

REGULATION

- ❑ Regulatory bodies and national/international authorities;
- ❑ Compliance obligations;
- ❑ Legal liability and responsibility;
- ❑ Fines and sanctions;
- ❑ Regulatory requirements.

4.2 AI Standardization

General Overview



Big Data

- 20546 – Overview & Vocabulary
- 20547 Series – Big Data reference architecture
- 24668 – Process Management

Data quality for analytics and ML

- 5259-1 – Overview
- 5259-2 – Measures
- 5259-3 – Management Requirements & Guidance
- 5259-4 – Process Framework
- 5259-5 – Data Quality
- 5259-6 – Visualization framework
- 42103 – Synthetic Data in AI

Application

- 24030 – Use case
- 5339 – AI Application Guide
- 24372 – Computational Approach
- 5392 – RA Knowledge Engineering
- 18988 – Application of AI in health
- 20226 – Environmental sustainability

Governance

- 38507 – Governance implications of the use of AI by Organisations
- 42006 – Requirements for audit and certifications
- 42106 – Benchmarking AI systems
- 17903 – Overview of ML devices

Key

- 🟩 Published Standards
- 🟨 In-development

Foundational

- 22989 – Concepts & Terminology
- 23053 – Framework for AI using ML

Life Cycle Process

- 42001 – Management System
- 5338 – AI System Life Cycle Process
- 8183 – Data life cycle
- 42005 – AI System impact assessment

Trustworthiness

- 24028 – Overview
- 24027 – Bias
- 24368 – Ethical & Societal Concerns
- 8200 – Controlability
- 21221 – Beneficial AI systems
- 6254 – Explainability
- 12791 – Treatment of bias

Risk

- 23894 – Risk Man.

Testing

- 29119-11 – Testing of AI Systems
- 17847 – Verification analysis

Quality

- 25059 – Quality Model
- 25058 – Quality Evaluation

Assessment

- 24029 – Robustness of Neural Net.
- 4213 – Classification Performance
- 24029-2 – Robustness of Neural Net.

Safety

- 5469 – Functional Safety

Life Cycle

- 24028 – System Engineering
- 12207 – Software Engineering

Foundation

- 2382 – Vocabulary
- 31000 – Risk Management
- 9001 – Quality System
- 12792 – Taxonomy

Software

- 25000 Series – Software Quality Square
- 29119 Series – Software Testing

Source: Introduction to Standards for Artificial Intelligence, Standards Australia, May 2023

4.3 What is a harmonized standard in Europe?

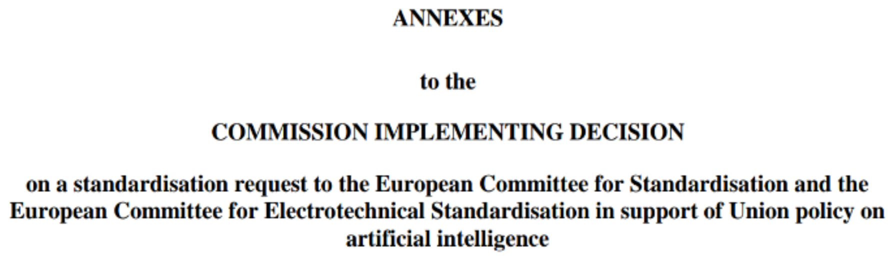
- A **harmonised standard** is a technical specification, adopted by a **European Standardisation Organisation**, for repeated or continuous application, with which compliance is **not compulsory** and that have been adopted on the basis of a request made by the **Commission for the application of Union Harmonisation Legislation**

  - European standardization is overseen by European Standardization Organization like **CEN** (European Committee for Standardization), **CENELEC** (European Committee for Electrotechnical Standardization), and **ETSI** (European Telecommunications Standards Institute).

  - It was founded on the principles of the **World Trade Organization** (WTO), emphasizing coherence, transparency, consensus, voluntary application, and independence.

European Parlament

4.4 EU Harmonized standards



ANNEXES

to the

COMMISSION IMPLEMENTING DECISION

on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence

ANNEX I

List of new European Standards and European standardisation deliverables to be drafted

European **standard(s)** and/or **standardisation deliverable(s)** on:

1. **risk management systems** for AI systems;
2. **governance** and **quality of datasets** used to build AI systems
3. **record keeping** through **logging capabilities** by AI systems
4. **transparency** and **information provisions** for users of AI systems
5. **human oversight** of AI systems
6. **accuracy specifications** for AI systems
7. **robustness specifications** for AI systems
8. **cybersecurity specifications** for AI systems
9. **quality management systems** for providers of AI systems, including **post-market monitoring** processes
10. **conformity assessment** for AI systems

5. AI Management System

innovating safety & security

5.1 What is a management system?

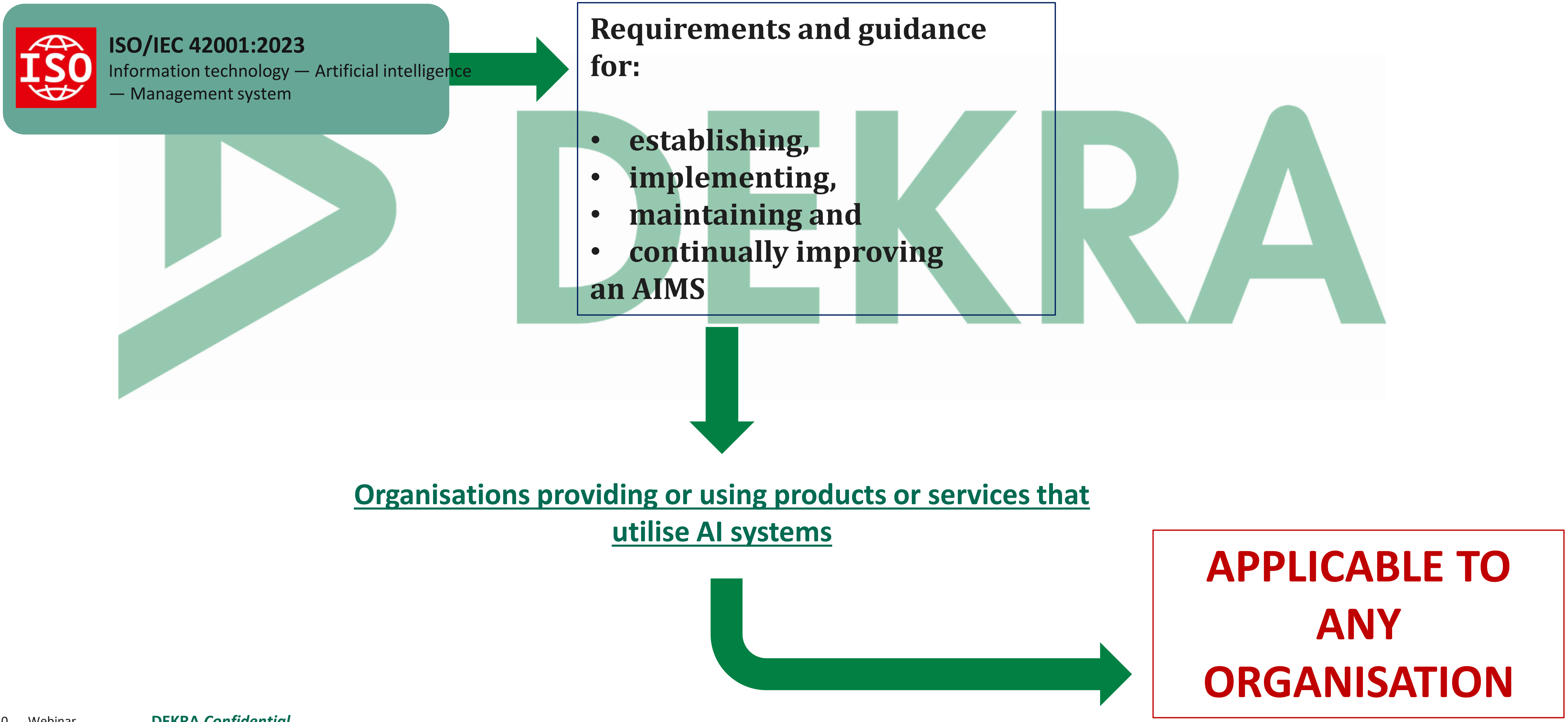
“A management system is the way in which an organization manages the interrelated parts of its business in order to achieve its objectives”

- ISO 9001 – Quality management systems
- ISO 14001 – Environmental management systems
- ISO/IEC 27001 – Information Security management systems

Some MS are for specific sectors:

- ISO 13485 – Medical devices quality management system
- ISO 22163 – Railway quality management system

5.2 Artificial Intelligence Management System (AIMS)



ISO/IEC 42001:2023
Information technology — Artificial intelligence — Management system

Requirements and guidance for:

- **establishing,**
- **implementing,**
- **maintaining and**
- **continually improving**

an AIMS

Organisations providing or using products or services that utilise AI systems

APPLICABLE TO ANY ORGANISATION

5.3 Artificial Intelligence Management System (AIMS)

Differences in governance from the common ICT governance

AI systems VS other ICT

Decision automation

❑ Output based on historical and current data

❑ Resultant prediction represented in probability

❑ Fast decision-making

Data driven problem solving

❑ Large amounts of data examination

❑ Data drive the analytical process

❑ Speeds up problem solving

❑ New governance problems

Adaptive systems

❑ Retraining / Ongoing training

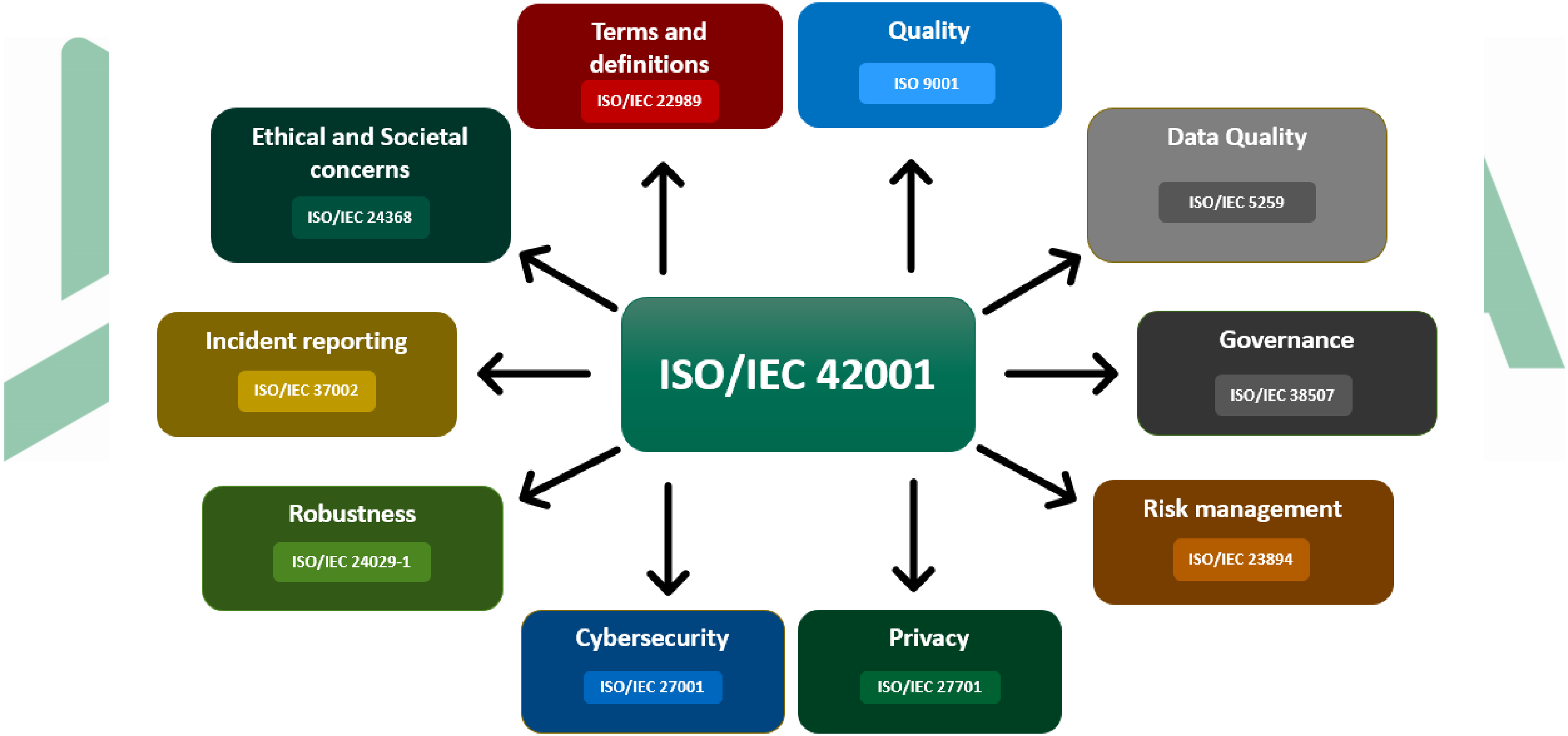
❑ Different outputs to same input

❑ Additional assessments for checking AI boundaries over time

❑ Ensure compliant AI

5.4 ISO/IEC 42001 – AI Management System

- References to AI Standards and other Management Systems



5.5 AIMS Implementation in the organizations

PRACTICAL ASPECTS:

- Substantial first-time **implementation effort** : set up policies, processes, monitoring procedures, etc.
- **Continuous** action: updating documentation and monitoring quality and well-functioning of the AI system.
- **Dedicated resources** are required.
- **Connection** to other **Management Systems**:
  - ISO 9001: Quality Management System
  - ISO 27001: Information Security Management System
- Compliance with **Data Protection requirements** (GDPR).

IMPLEMENTATION

5.6 Certification of AI Management System

Benefits of obtaining an AIMS certification go beyond regulatory compliance:

1 Trust and Credibility

Enhance **trust** with clients, partners, regulators, and the public.

2 Competitive Advantage

Differentiates organizations in the market through certified responsible AI practices.

3 Quality and Risk Management

Following best practices to ensure high quality standards and mitigate risks.

4 Accountability

Accountability in the development and usage of AI technology from the organization perspective

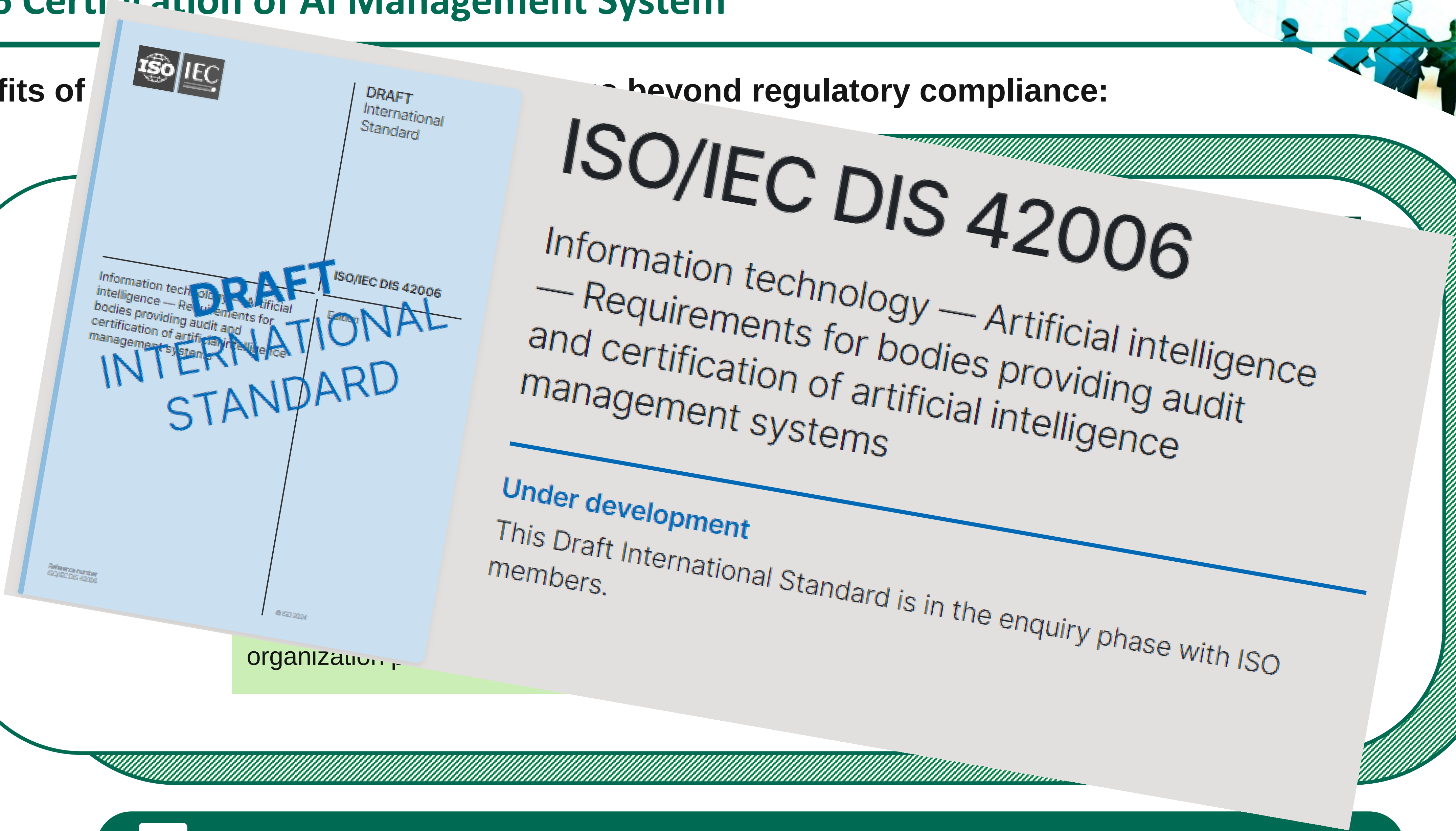
5 Broad scope

All these benefits applying to any AI-based system, process or service. Not restricted to High-risk AI.

ISO 42001 is currently the only certifiable AI Management System.

5.6 Certification of AI Management System

Benefits of … beyond regulatory compliance:



DRAFT International Standard

Information technology — Artificial intelligence — Requirements for bodies providing audit and certification of artificial intelligence management systems

ISO/IEC DIS 42006

DRAFT INTERNATIONAL STANDARD

Reference number ISO/IEC DIS 42006

© ISO 2024

ISO/IEC DIS 42006

Information technology — Artificial intelligence — Requirements for bodies providing audit and certification of artificial intelligence management systems

Under development

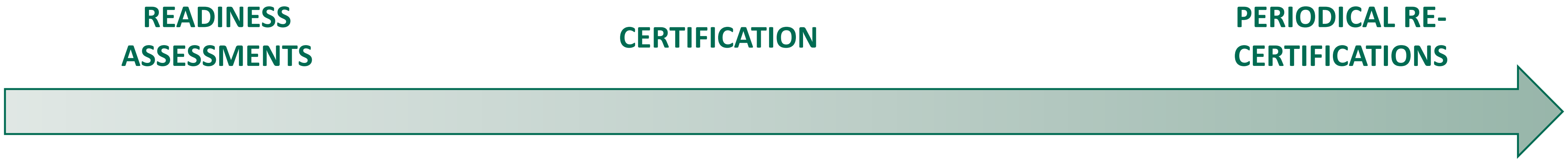
This Draft International Standard is in the enquiry phase with ISO members.

organization p

ℹ **ISO 42001 is currently the only certifiable AI Management System.**

5.7 Path to Certification of AIMS

- ISO 42001 **good basis** for AI governance
  - **Covers** most **EU AI Act** requirements
  - Type A Management System => **auditable**.
- Development of a **certification scheme**:
  - Definition of audit processes.
  - Definition of auditor qualifications.
  - Conversion of ISO requirements into measurable criteria.
- **Qualification** of **auditors**.
- Creation of **Certification Bodies** accredited with ISO42006 / ISO17021 / ISO17067.



READINESS ASSESSMENTS

CERTIFICATION

PERIODICAL RE-CERTIFICATIONS

6. Closing Remarks

innovating safety & security

6.1 We are active in AI Standardization and Policy discussions

AI standardization working groups participation:

- ISO/IEC JTC 1/SC 42 – AI: worldwide scope
- CEN/CENELEC JTC 21 – AI: European scope
- DIN/DKE Committee NIA: German mirror from CEN/CENELEC and ISO
- UNE - AI and Big Data: Spanish mirror from CEN/CENELEC and ISO
- ETSI ISG SAI: European scope – Electrotechnical scope
- ENISA – AI Cybersecurity: European Agency on Cybersecurity

Advisory work:

- TIC Council: participation as advisory role on AI Task Force
- German Standardization Forum: AI and Data Working groups
- Estrategia IA: andalusian regional AI strategy

6.2 DEKRA know-how at the service of the AI TIC Community

DEKRA contributes actively to the AI TIC Community:

- Long expertise on physical product testing
- Domain knowledge in multiple industries: Automotive, medical, Cybersecurity, Functional Safety, etc.

Collaboration with universities and R&D centers:

DFKI (German AI Research Center)

✓ AI Model testing tools for automatic Vehicle failure diagnosis

Shanghai SJTU University

✓ Digital human test standard

University of Malaga, University of Barcelona (Salle URL)

✓ Several R&D projects

Several startups on AI Validation software & tooling

6.3 - 1st Generation AI Testing & Certification Services

TRAINING & PRE-ASSESSMENTS

AI Training & Pre-assessment

Expert training and pre-assessment services on multiple aspects related to AI technology and regulations

- AI Risk Awareness
- AI Regulations and Standards
- Trustworthiness & Ethics
- Readiness assessment (DEKRA AI-Ready)

ASSESSMENTS

AI Audit & Certification

Assessment and conformity respect to standards and good practices for development and operationalization of AI solutions.

- Management Systems (ISO 42001)
- AI Risk management (ISO 23894)
- Road Vehicles Safety&AI (ISO 8800)
- Data Labelling (ISO 5259-4)
- A-Spice Machine Learning

AI Testing

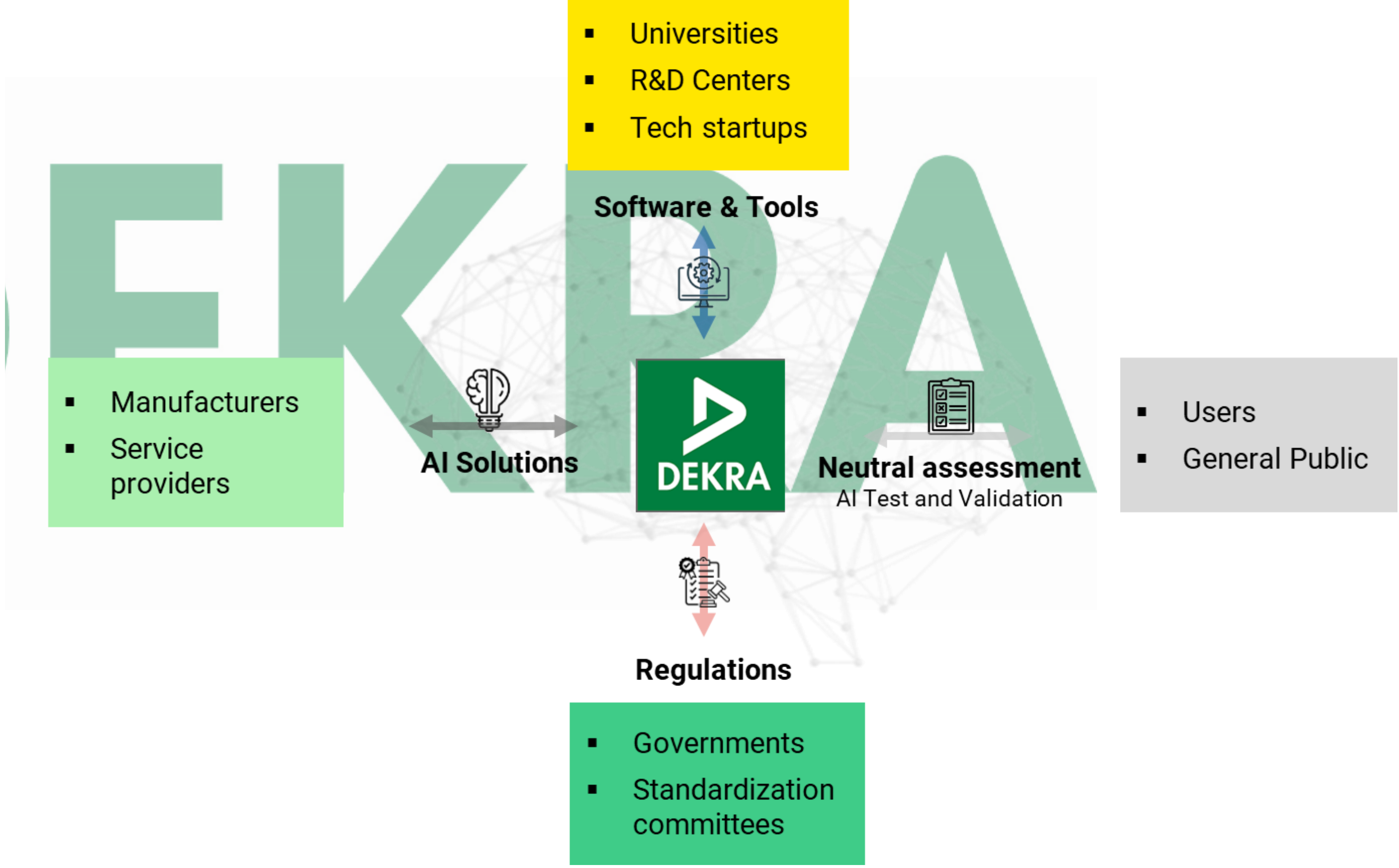
Our expert AI testers conduct thorough assessment leveraged by cutting-edge Software tools.

- Data Quality (ISO 5259)
- Model Robustness (ISO 24029)
- AI Bias & Fairness (ISO 24027)
- AI Security

6.4 Closing remarks

- **ISO 42001 (AIMS)** covers most of organizational requirements to ensure trustworthy AI and **compliance to the EU AI Act**.
- **Certification Scheme** could be implemented taking ISO 42001 requirements as basis.
- The certification of AIMS will allow the providers and developers to **build trust** of users **in AI technology** and take **advantage to competitors**.

- **<u>Let’s act now!</u>**



- Universities
- R&D Centers
- Tech startups

- Manufacturers
- Service providers

- Users
- General Public

- Governments
- Standardization committees

Your DEKRA contact for AI

PhD. Xavier Valero
Director AI & Advanced Analytics
Digital & Product Solutions
Innovating Safety & Security
DEKRA Testing and Certification, S.A.U.
+34 689 495 876
xavier.valero@dekra.com

Valentino Merlino
AI Security Consultant
AI & Advanced Analytics
Digital & Product Solutions
Innovating Safety & Security
DEKRA Testing and Certification S.A.U.
valentino.merlino@dekra.com

Gaia Fabbri
Country Sales Manager
Industrial & Consumer Goods
DEKRA Italia SrL
T +39 02 89929-226
M +39 338 6706268
gaia.fabbri@dekra.com