3ª EDIZIONE

CYBERSEC 2024

ROMA, ITALIA

LA CYBERSECURITY NELL'ERA DELL'AI

6-7 MARZO

www.cybersecitalia.events

CYBERSECURITY ITALIA EVENT

CYBERSECURITYITALIA

RELATORE

Analisi delle criticità nell'integrare l'AI di terze parti nelle aziende;

Marco Tulliani

Partner, *Intellera Consulting*
Partner
M + 39 331-6703155

marco.tulliani@intelleraconsulting.com
www.intelleraconsulting.com

*AI is rooted in the **Good AI** concept and includes the following three pillars:*

Good AI

AI for the common good

- Development of **use cases** to be used as vertical solutions in the Good AI area, e.g., **AI against organized crime**

Compliance with EU policies

- **Policy analysis** of EU digital policies
- **Legal analysis** of the EC's **Digital Acts**
- **Training services** offered in collaboration with **legal partners**

Human-centric AI governance

- **AI governance model** to ensure its quality, efficiency and effectiveness
- AI-specific **risk management**
- ***Augmented human intelligence***, AI to support people and implementation of change management systems

The definition of AI as per the AI Act

An **Artificial Intelligence** (AI) system is: *a **machine-based** system, designed to operate with **varying levels of autonomy**, that is capable of influencing the environment by **producing an output (predictions, recommendations or decisions)** for a given set of objectives.*

AI systems:

They often have **machine learning capabilities** to adapt **and perform new tasks autonomously**.

They can be used as **standalone software systems**, **embedded in a product**, **used to provide product functionality**, or **utilized as an AI component** in a larger system.

Generative AI

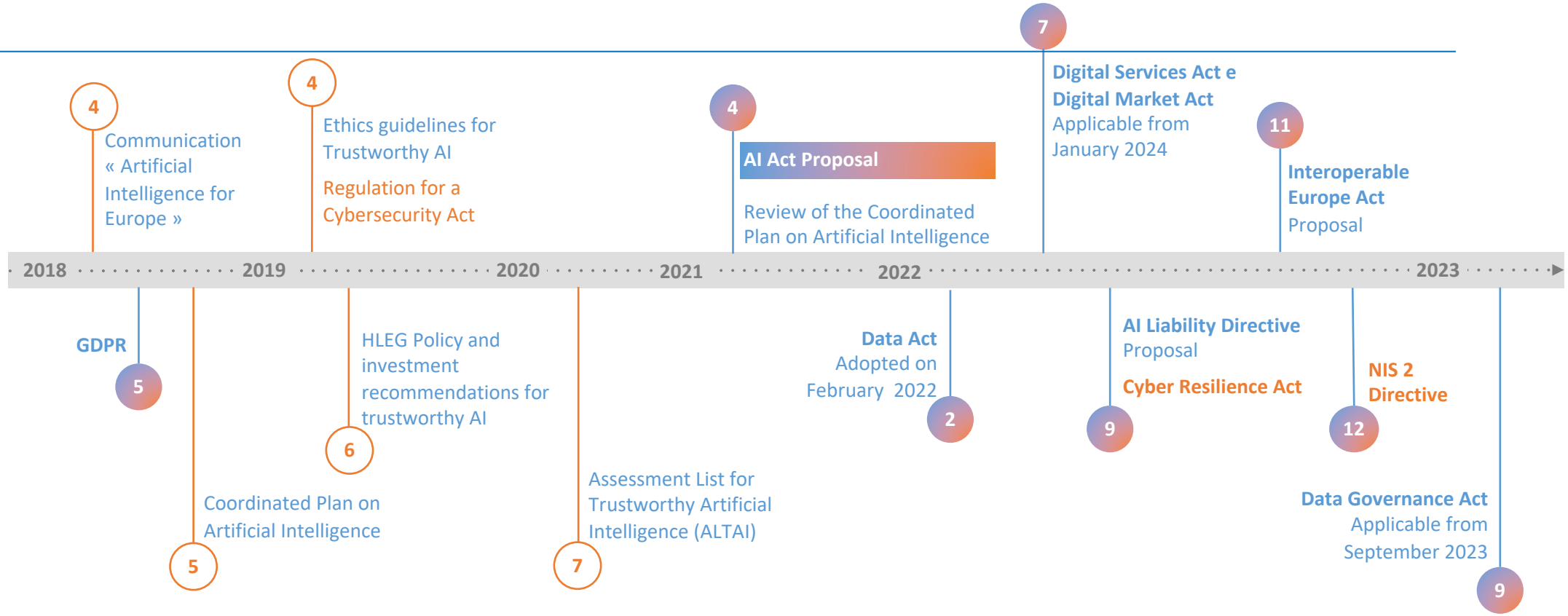
They include **foundation models**: systems that utilize a vast amount of data and are designed to provide general-purpose results that can then be customized

They include **General-Purpose AI Systems (GPAIs)**: systems that have a wide range of potential uses, both intended and unintended by developers.

COMPANIES WITH EU POLICIES

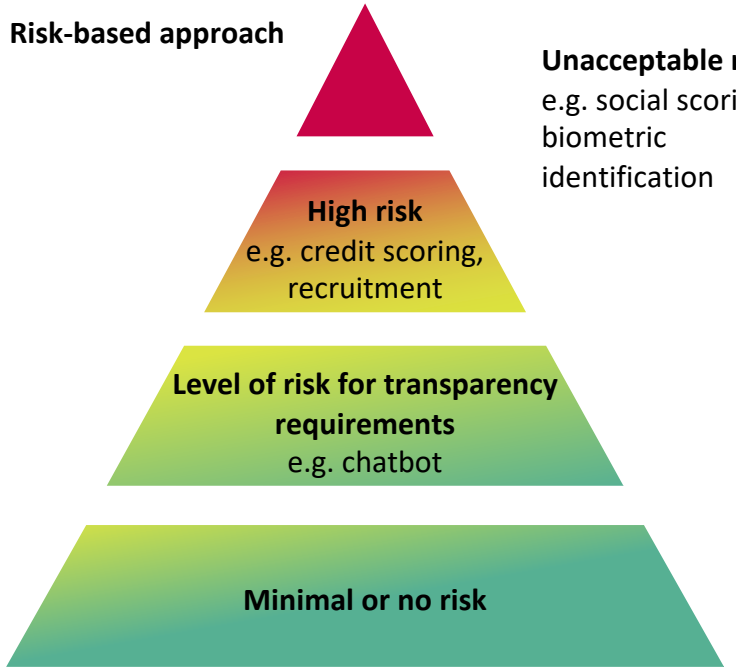
COMPLIANCE WITH EU POLICIES

The European Commission's journey to "Trustworthy" AI



AI ACT

The European Commission has proposed the first regulatory framework for AI regulation, following a risk-based approach



Prohibited

Permitted subject to compliance with AI requirements and ex-ante conformity assessment

Permitted subject to information / transparency obligations

Permitted with no restrictions, but suggested Code of Conducts

Who is subject to the AI Act?

It will apply to any AI provider, user, importer or distributor whose services or products reach the EU market:

- **Provider**: any natural or legal person, public authority or other body that develops an AI system, or has an AI system developed
- **User**: any natural or legal person using an AI system under its authority and places that system on the market or puts it into service

AI ACT

High Risk

What AI applications are high risk (1/2)?

AI systems used as **safety components** of products (or that are themselves products) that are normally subject to third-party ex-ante conformity assessment, covered by the legislation in Annex 2

- Annex 2 includes all **NLF legislation**, such as Machinery Regulation, Radio Equipment Directive, Medical Devices Regulation, In-Vitro Diagnostics Regulation
- It also includes sector-specific harmonisation legislation, including cars and aircrafts

AI ACT

High Risk

What AI applications are high risk (2/2)?

AI systems posing **risk of harm to the health, safety or fundamental rights of natural persons**, listed in Annex 3:

- **Biometric identification** (only after judicial authorization)
- **Access to essential private and public services and benefits** (social benefits or credit*)
- Management and operation of **critical and digital infrastructure**
- **Education**; **employment**, workers management and access to self-employment
- **Emergency first response; crime prediction** and court use; **migration**, asylum & border control management

** With the exception of applications used for the purpose of detecting financial fraud*

AI ACT

High risk: all providers must ensure compliance against mandatory requirements for AI applications classified as high risk

- Establish and implement **risk management processes**
- Use **high-quality** training, validation and testing **data**
- Establish **documentation and design logging features** (traceability & auditability)
- **Record keeping**
- Ensure appropriate **certain degree of transparency** and provide users with **information**
- Ensure **human oversight**
- Ensure **robustness, accuracy and cybersecurity**

After undergoing the conformity assessment procedure, high-risk applications will have to be registered with the European Registry.

Conformity shall be ensured also by adherence to the relevant **European standards,** once defined.

Potential penalties for non-compliance (from 1 to 7% of annual worldwide turnover) will be calculated based on the actual infringement (gravity, impact, etc.).

** With the exception of applications used for the purpose of detecting financial fraud*

HUMAN – CENTRIC AI GOVERNANCE

Our approach to AI Governance: the AI Lifecycle

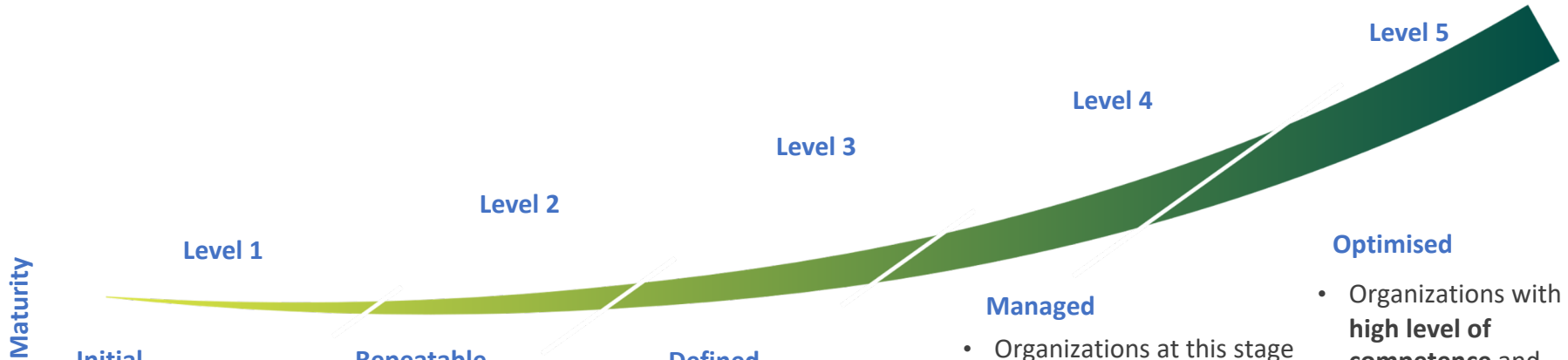
A **conceptual framework** has been introduced that divides AI management into phases, the so-called "AI Lifecycle," which facilitates the understanding of the areas involved in AI governance, in line with the upcoming legislation and the ISO standards.

AI Lifecycle Phases

Strategy	Delivery	Ecosystem	Data	Design	Implementation	Operating	Monitoring
• Corporate Strategy • Portfolio Management • Policy and regulation	• Organizational Change • Project planning • Risk management	• Vendor Management • Technology Roadmap • Standards and Practices	• Data Requirements • Data Gathering • Data Preparation	• Model Design • Verification and Tuning • Controls Framework	• Business Readiness • Deployment • Transition	• Operations • Resilience • Support	• Compliance • Cybersecurity • Performance • Maintenance

HUMAN – CENTRIC AI GOVERNANCE

Our approach to AI Governance: the Maturity Assessment



Initial

- Organizations that have **not adopted proactive measures** for AI management.

Repeatable

- Organizations at this stage are **starting to adapt** to AI management.
- They still have **only a theoretical awareness** of AI

Defined

- Organizations **sufficiently prepared** in terms of strategy, organizational structure, and data availability to implement AI.

Managed

- Organizations at this stage have a reasonable level of **practical experience** and understanding of how to proceed with AI management.
- However, there are still **gaps** and **limitations**.

Optimised

- Organizations with a **high level of competence** and experience in AI.
- They have a proven **track record in AI management**, which also includes reliable AI practices

HUMAN – CENTRIC AI GOVERNANCE

Trustworthy AI strategy

- AI strategy in line with the EU Regulation
- AI and governance maturity assessment
- Roadmap definition & Action plan in line with the EU Regulation
- Code of Conduct definition and adoption

Organisational governance

Process review* concerning:

- Data and models governance & management
- AI development
- AI procurement
- Risk management and Audit

Training & Change management

- Training and upskilling programmes
- Change management strategy and implementation

Trustworthy AI Governance platform

- Support in platform setup and adoption
- Access to the platform (SaaS)
- Ad-hoc training sessions

Technology enablement

AI development process review concerning:

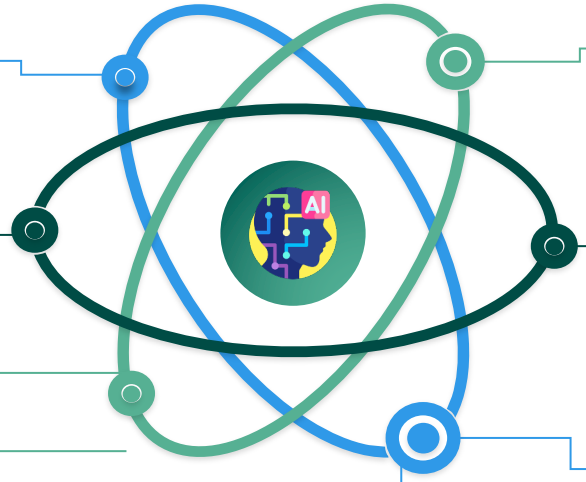
- Implementation of AI governance techniques
- Development of unbiased, interpretable, transparent, **robust and secure AI** solutions

*Approach to the operating model through four main dimensions - **POTI**: processes, organizational model, technology, and information.

FOCUS ON CYBERSECURITY FOR AI

FOCUS ON CYBERSECURITY FOR AI

Types of threats in the AI domain, according to ENISA



Nefarious activity/abuse

These are intentional actions aimed at ICT systems, infrastructure, and networks with the goal of stealing, altering, or destroying a specific target.

Physical attack

These are actions aimed at destroying, exposing, altering, disabling, stealing or gaining unauthorized access to physical assets like infrastructure, hardware, or interconnection.

Disaster

This refers to a sudden accident or a natural catastrophe that causes great damage or loss of life.

Unintentional Damage

These are unintentional actions that cause destruction, harm, or injury to property or persons resulting in a failure or reduction in usefulness.

Failure/malfunction

These refer to partial or full insufficient functioning of an asset (hardware or software).

Legal

These are legal actions of third parties (contracting or otherwise) in order to prohibit actions or compensate for loss based on applicable law.

Outages

These are unexpected disruptions of service or decrease in quality falling below a required level.

Eavesdropping

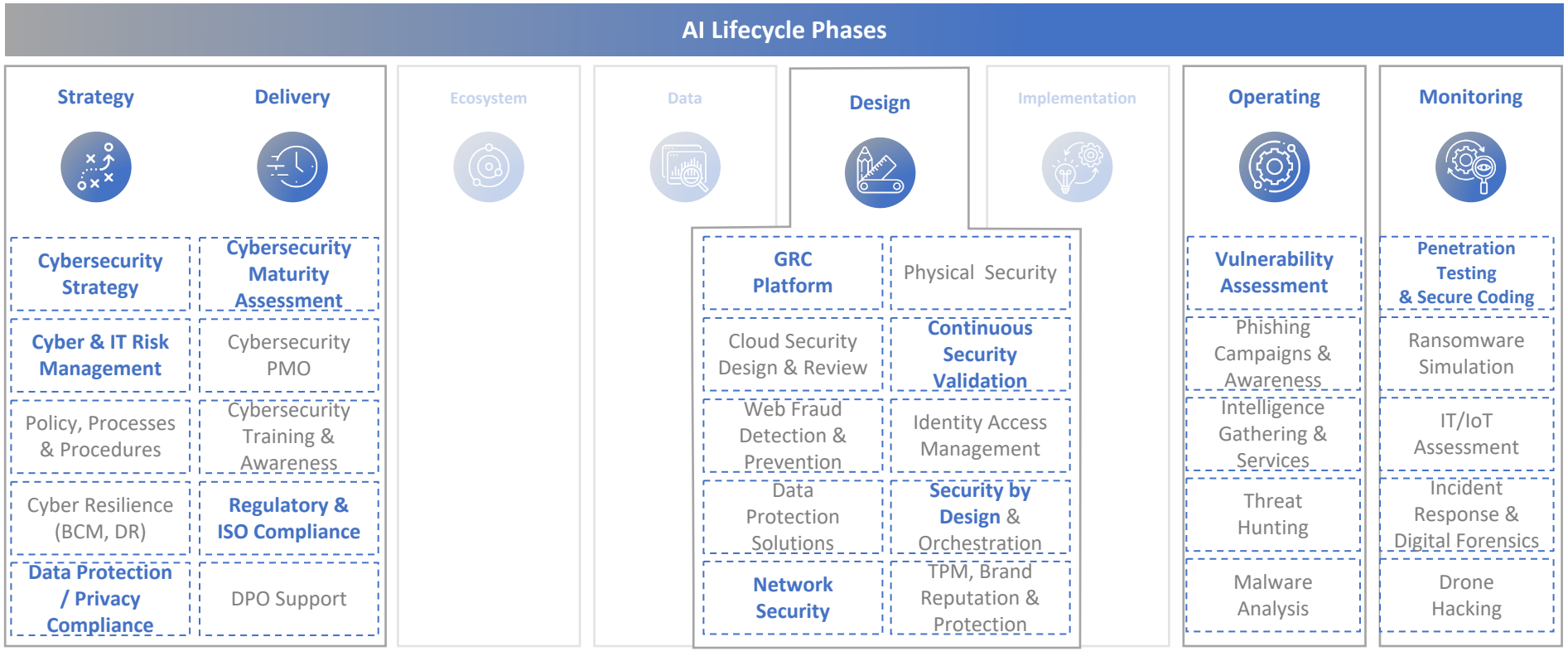
These actions aim to listen, interrupt, or seize control of a third-party communication without consent.

FOCUS ON CYBERSECURITY FOR AI

Intellera's proposed solutions to the threats

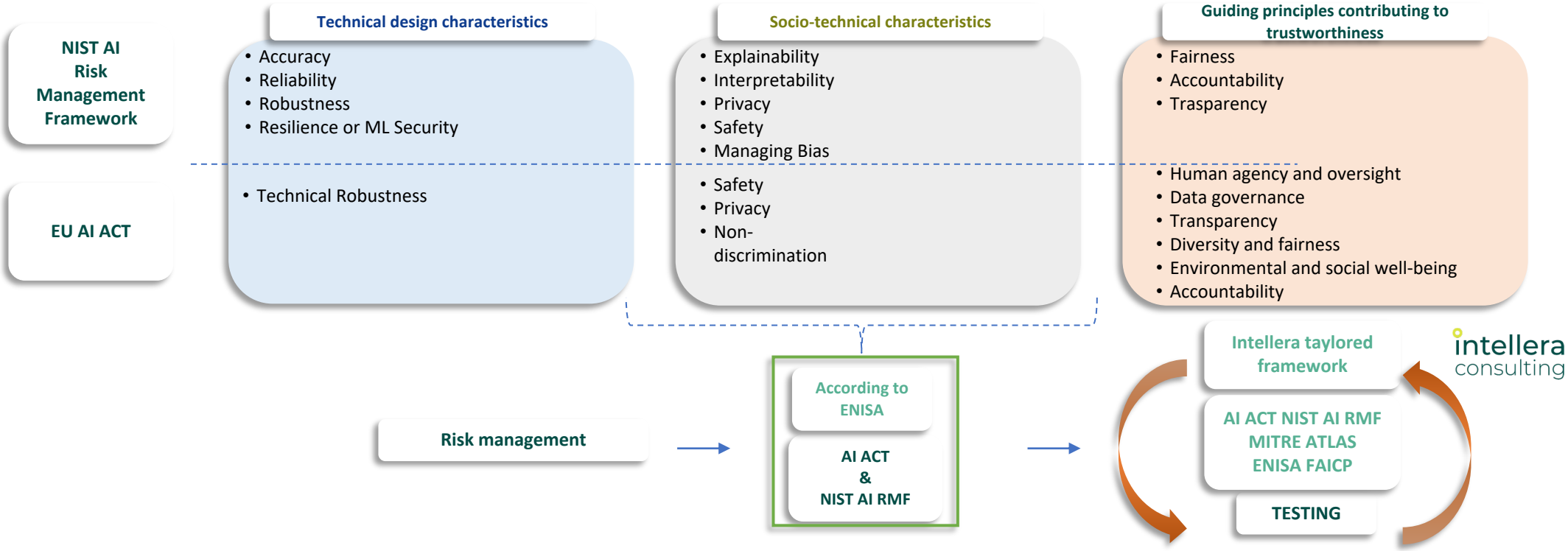
Threat Taxonomy (ENISA)		Proposed solution
Nefarious activity/abuse	Malicious acts is to steal, alter, or destroy a specific target (e.g.: Data poisoning, Malware, Adversarial examples, etc.)	Risk Assessment against adversarial attacks, Threat Intelligence
Physical attack	Unauthorized attempts to damage, expose, modify, disable, gain unauthorized access (e.g.: Theft, Sabotage, Tampering, etc.)	Risk assessment, information security policy, cyber hygiene best practices
Disaster	Natural disasters and environmental phenomena (e.g., fire, flood, storm, etc.)	Disaster recovery plan outlines steps for service restoration
Failure/malfunctions	Partial or full insufficient functioning of an asset (e.g.: Third-party provider failure, compromise of model frameworks, etc.)	Vulnerability management, a regular process of identifying, assessing, reporting and remediating cyber vulnerabilities across endpoints, workloads, AI systems
Legal	Legal actions of third parties, to prohibit actions for loss based on applicable law (e.g.: Data protection, Cybersecurity regulation, etc.)	Regular audits ensure compliance with new AI data privacy laws
Outages	Unexpected disruptions of service or decrease in quality falling below a required level (e.g.: Service disruption, Service termination, Service degradation, etc.)	Integrating AI systems and data models with cloud security and DDoS Plan
Unintentional damages	Misconfiguration, reducing data accuracy, compromising ML training augmented data (e.g.: Human error, software bug, etc.)	Employee training can help recognize and report threats like phishing
Eavesdropping/Hijacking	Unauthorized attempts to control a third-party communication (e.g.: Man-in-the-middle attack, Wiretapping, etc.)	Data protection policy: SSL/TLS and AES encryption can secure data in transit and at rest, making intercepted data unreadable

FOCUS ON CYBERSECURITY FOR AI



FOCUS ON CYBERSECURITY FOR AI

Our approach: based on ENISA and Intellera's multiframework asset



FOCUS ON CYBERSECURITY FOR AI

Security testing process on AI models

Given the nature of AI, and its being a rapidly evolving technology, there are currently no frameworks on AI testing, but preparation of some is underway
Some of these are:

- **ISO/IEC 5338**
- **ETSI GR SAI 003**
- **ISO/IEC AWI TS 12791**-Treatment of unwanted bias in classification and regression machine learning tasks

According to the ETSI working group on AI published draft of Security Testing of AI, some of the methods and techniques are presented below along with solutions to the proposed elements:

Security testing approaches for AI: AI security testing involves creating test cases to identify potential vulnerabilities. It tests the AI's response to various data inputs, its ability to handle large requests, and its resilience against manipulation attempts.

Security test oracles for AI: A test oracle determines if a test has passed or failed. It checks AI outputs against expected results, monitors AI behavior for anomalies during testing, and validates its performance under various conditions.

Definition of test adequacy criteria for security testing of AI: criteria for AI security testing help determine when testing is sufficient. They consider factors like the number of vulnerabilities identified, and alignment of AI behavior with expected behavior during testing. These criteria can track progress in cybersecurity and decide when to stop a security test.

Security testing tools mentioned in "Multilayer framework for good cybersecurity practices for AI":

OECD

GuardAI

ATLAS

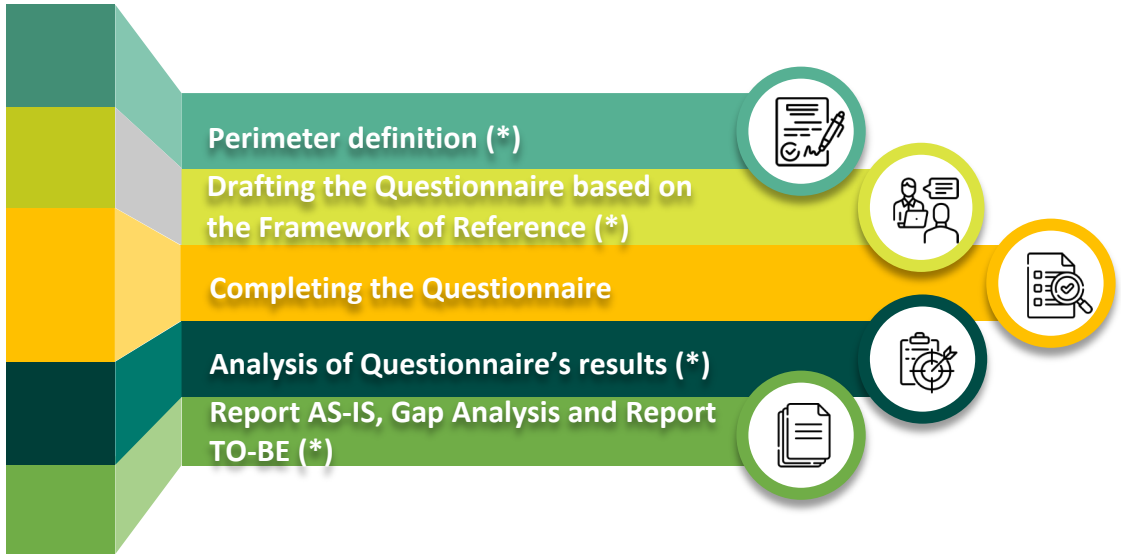
Counterfit

futurium.ec.europa.eu

OWASP AI Security & Privacy guide

FOCUS ON CYBERSECURITY FOR AI

Roadmap of interventions aimed at strengthening AI models

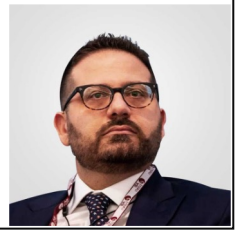


- Determine the **Perimeter** for Assessment in accordance with specific industry guidelines such as:
  - **NIST AI RMF**
  - **MITRE ATLAS**
  - **AI ACT**
  - **ENISA FAICP**
- Questionnaire design based on the controls of the chosen framework.
- **Completion of the Questionnaire** is the crucial step in the evaluation process.
- After completing the Questionnaire, you can proceed with its **analysis**, focusing on the items of greatest interest.
- The final stage involves the preparation of **AS-IS** and **TO-BE** Reports containing insights into current and target security posture, respectively, including **Gap Analysis**.

(*)Intellera is able to identify the best Technology Partner to automate and execute the tasks related to the roadmap and meet the client's needs

Marco Tulliani
Partner
M + 39 331-6703155

marco.tulliani@intelleraconsulting.com
www.intelleraconsulting.com



GRAZIE PER L'ATTENZIONE