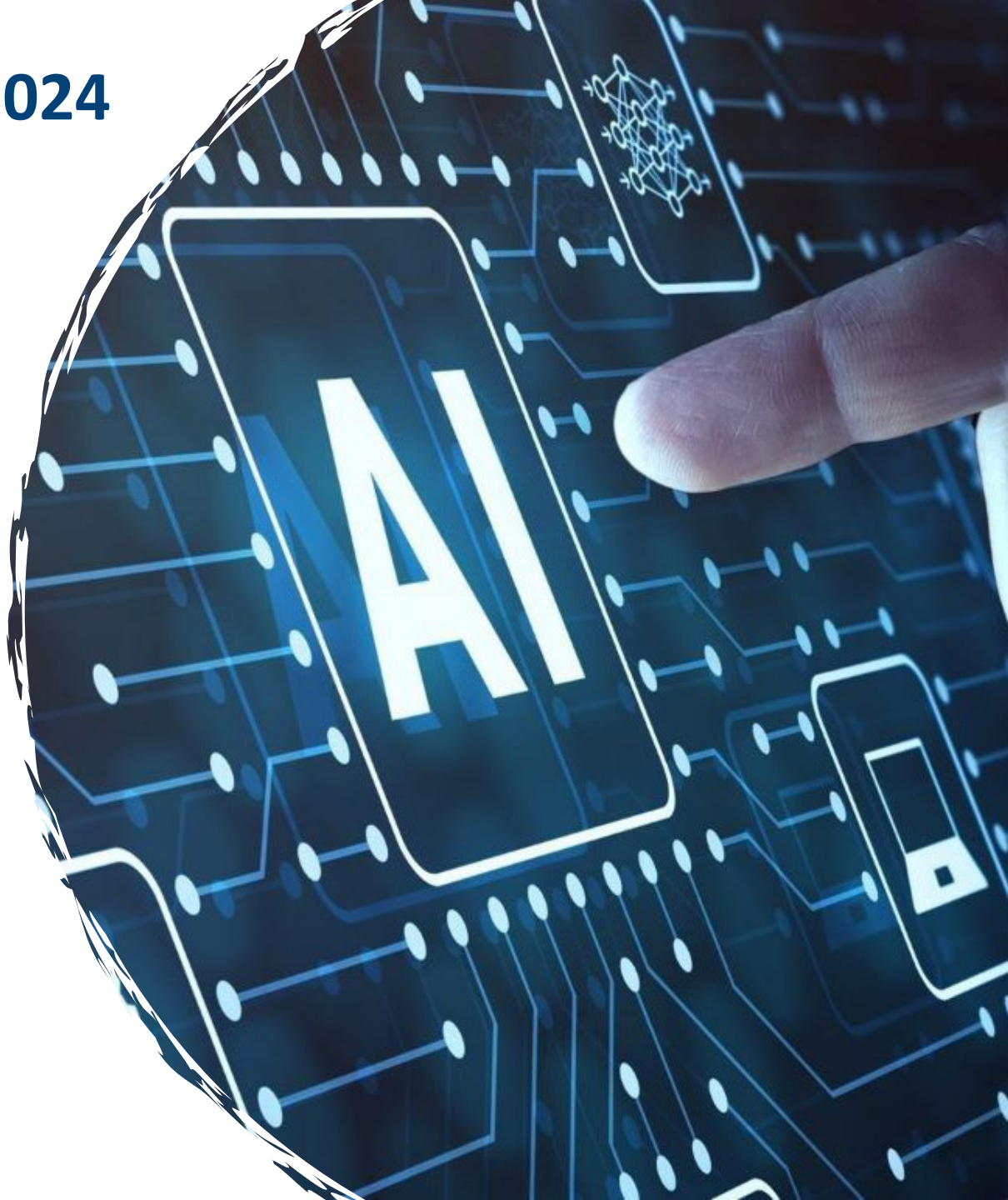
ETSI AI Conference 2024



European Approach to AI Cybersecurity – cybersecurity certification

in Session 4: Cyber Security in the Context of AI

Presented by:

06/02/2024

CYBERSECURITY CERTIFICATION AND AI

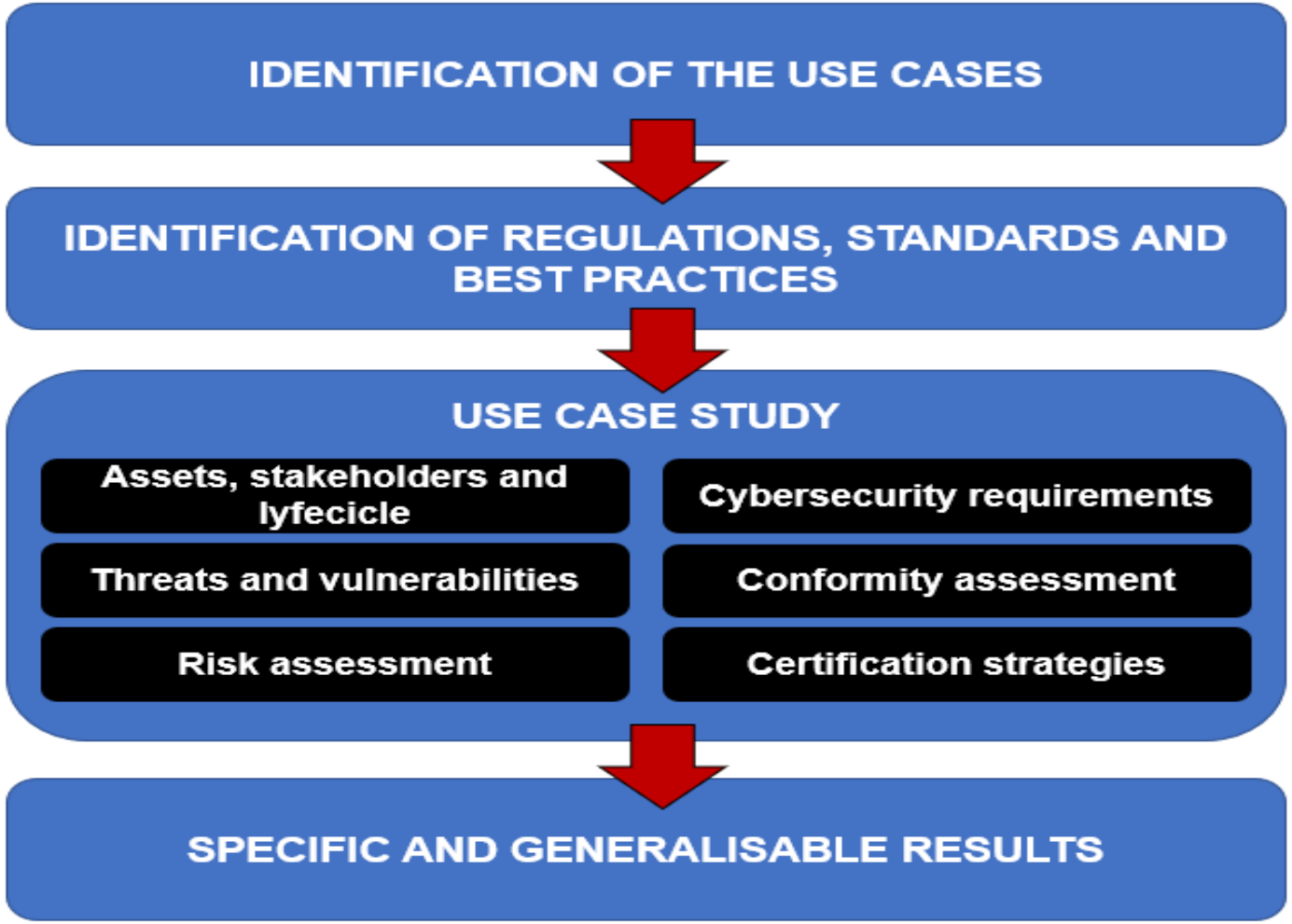
- The #1 priority is to support mitigating the cybersecurity-related risk posed by AI; Cybersecurity certification can help achieving a cyber-secure AI (certify products, solutions and services at a level that is consistent with risks to be mitigated, but also taking into account the market needs e.g. cost, time and performance to be achieved)
- Certification is also a possible means for presumption of conformity to the cybersecurity requirements of the AI Act;
- Emerging policy context e.g. AI Act, CRA, NIS2, CSA, sectorial regulations and their interplay;
- It is essential to know what an AI product, service or process shall fulfil in terms of cybersecurity requirements -> high-level cybersecurity requirements set out by the existing regulations e.g. AI Act, CRA need to be “translated” down to specific cybersecurity requirements applied to the internal architecture of the AI system

CYBERSECURITY CERTIFICATION AND AI – HOW FAR CAN STANDARDS TAKE US?

- Standards help mitigate risks: there are existing general-purpose standards that are readily available for information security and quality management in the context of AI; some other need adaptation for AI – guidance needed as to how existing standards related to the cybersecurity of software should be applied to AI;
- AI-specific requirements: deriving from the domain of application; standards to cover aspects specific to AI, such as the traceability of data and testing procedures;
- ML vs AI and reflection of risk mitigation - inherent features of ML in AI (risk mitigation in particular should be considered by associating hardware/software components to AI; reliable metrics; and testing procedures)

https://www.enisa.europa.eu/publications/cybersecurity-of-ai-and-standardisation

GENERAL OVERVIEW OF THE STUDY



Source: ENISA Analysis of Use cases – AI feasibility study

USE CASE-BASED REASONING

Example: Use Case 1 – Medical Imaging

Description: A private clinic buys access to a cloud-based AI platform that allows training a ML model on patients' medical imagines (X-rays) and on data related to age, gender and body mass. The private clinic develops a ML-based tool to detect the presence or absence of osteoporosis in patients. The private clinic sells the tool to other private clinics.

Regulatory aspects: MDR / AI Act. The tool put on the market by the private clinic could be considered a medical device requiring third-party conformity assessment AI systems that are products under this Regulation also fall in scope of AIA, but not in scope of CRA. AI platform: In principle, the AI platform does not fall under the AI Act.

Cybersecurity requirements: AIA and MDR

Certification options: EUCS for AI cloud-based services

Challenges: Training / Re-training; Interplay of sectorial regulation and AI act; Supply chain

USE CASE-BASED REASONING - EXAMPLE

ASSETS

- Machine learning algorithms
- Machine learning framework
- Machine learning service
- Foundational model
- Model using CNN
- Web server
- Model server
- Database Management System Server
- Training data
- Production data
- Evaluation data
- Input data
- Scanner
- X-ray computer aided diagnostic system
- Integrated development environment
- Version control system
- Network communications
- Additional assets for AI security

STAKEHOLDERS

- AI system owner
- Clinic / Hospital
- Radiologist / medical practice
- AI Solutions provider
- Patients
- CSP
- Data scientist
- Developers
- Data engineers
- Data labeller
- Cloud service architect
- Importer
- Distributor
- National competent authorities
- Notified body

LIFECYCLE

- Plan and design
- Data collection
- Data cleaning
- Data pre-processing
- Model training
- Model evaluation
- Software development
- Instructions & training
- Verification and validation
- User validation
- Clinical evaluation
- Deployment
- Post-market surveillance
- Decommissioning

AI THREATS

Data poisoning	Model inversion
Adversarial attacks	Membership inference
Model misuse	IT security threats
Scaling attacks	Privacy threats
Model stealing	

SECURITY CONTROLS

Data validation and sanitization	Data augmentation and randomization
Data anomaly identification	Input validation
Model validation	Input distortion
Model ensemble	Adversarial detection
Data augmentation	Model encryption
Data quality monitoring	Secure coding
Web filtering	Data minimization
Secure development lifecycle	Limited output of data
Data anonymization and de-identification	Privacy-preserving ML technique

VULNERABILITIES

Not integrated transparency	Insufficient monitoring
Insufficient training	ICT security not adapted
Insufficient explainability	Verification & validation absence
Insufficient control mechanisms	ANN evaluation absence
Absence of analysis	Model easy to poison
AI bias legal requirements	Evasion not considered
Techniques for reducing bias	Vulnerable AI components
Not enough testing	Non-trustable data sources
PII not anonymised	Uncontrolled data
Adversarial ML detection	Unprotected sensitive data
Insecure ML model	

CYBERSECURITY REQUIREMENTS

- EU Artificial Intelligence Act **(High-risk AI system)**
- Medical Devices Regulation **(Class IIb medical device)**
- Cybersecurity Act

EVALUATION CRITERIA

- Data poisoning
- Adversarial attacks
- Model stealing
- Model inversion

CERTIFICATION STRATEGIES

CERTIFICATION OPTIONS

- **EUCC + Protection Profile**
- **EUCS + Security Profile**

PROCEDURE AND PATH FOR CERTIFICATION

RISK ASSESSMENT

MRC (max)

MRC3

APL (max)

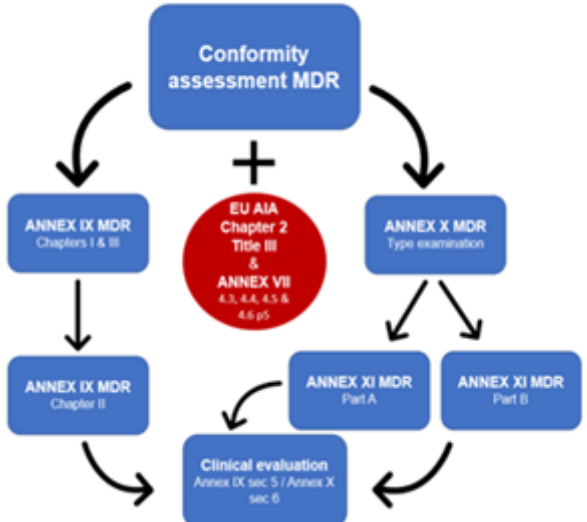
APL2

CAR (max)

CAR3

CSA ASSURANCE LEVEL

Cert. Scheme dependent



AI RISK ASSESSMENT

- There are different risk assessment methodologies currently proposed (ex. NIST, ISO, etc). In our study, we have considered ENISA Sectoral Cyber Security Assessment (SCSA) https://www.enisa.europa.eu/publications/methodology-for-a-sectoral-cybersecurity-assessment/@@download/fullReport (on ballot in JTC 13 WG 3 to become a standard)
- Differences from SCSA:
  - Identifying sectoral context: asset-based approach (ISO 27005:2022)
  - Risk scenarios: ISO/IEC 27090 for AI assets and threats, together with the ISO/IEC 27005 for those specific threats and vulnerabilities associated to IT

  E.g. data poisoning, adversarial attacks(dataset), model inversion attacks/backdoor attacks/model (ML model)
- Assessment of consequences and likelihood and attack potential levels: we followed SCSA

STAKEHOLDERS' ENGAGEMENT

A dedicated AI Thematic Group was created

- With members of existing ENISA AHWGs (EUCC, EUCS, EU5G, TGVH)
- Including EC, CEN CLC, ETSI
- Including MSs

THE FINE LINE…

- Not an easy task to implement: complex ecosystems, complexity in AI supply chain

-> need to *assess the market that is driving this ecosystem*

- AI risks vs “traditional” security risks

->how much we need/should/could take into consideration for cybersecurity

- #1 Priority is cybersecurity. But the relationship between the non-cybersecurity aspects and the “traditional” CIA needs to be assessed such as data quality, interpretability, explainability,
- How far can standards take us towards secure AI?

https://www.enisa.europa.eu/publications/cybersecurity-of-ai-and-standardization.

THANK YOU FOR YOUR ATTENTION

Agamemnonos 14, Chalandri 15231
Attiki, Greece

+111 123 456 789

info@enisa.europa.eu

www.enisa.europa.eu