"CYBERSECURITY CERTIFICATION AND AI: SECURITY CONSIDERATIONS AND CHALLENGES"

Xavier Valero - Director AI & Advanced Analytics

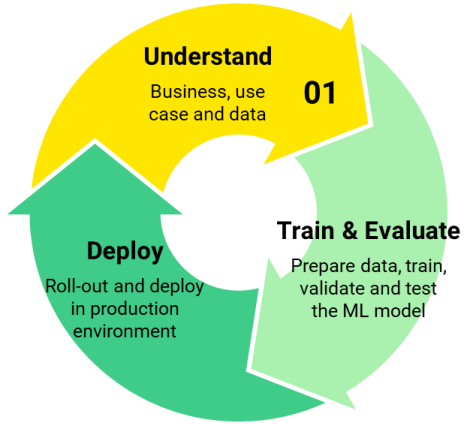
Jorge Wallace - Cybersecurity Technical Officer

Brussels, 7th June 2023

HOW TO TEST AND CERTIFY THE CYBERSECURITY OF AI?

- **Option 1) Apply general Cybersecurity certification framework**

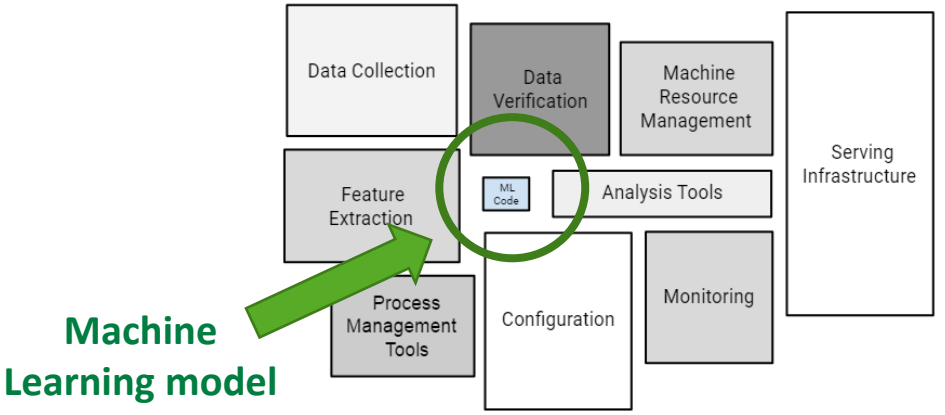
AI / Machine Learning (ML) is a very specific software task, whose output entirely depends on the input data and learning process. General Cybersecurity testing frameworks do not work here.



Machine Learning Life Cycle

- **Option 2) Create specific certification framework adapted to the Machine Learning Model**

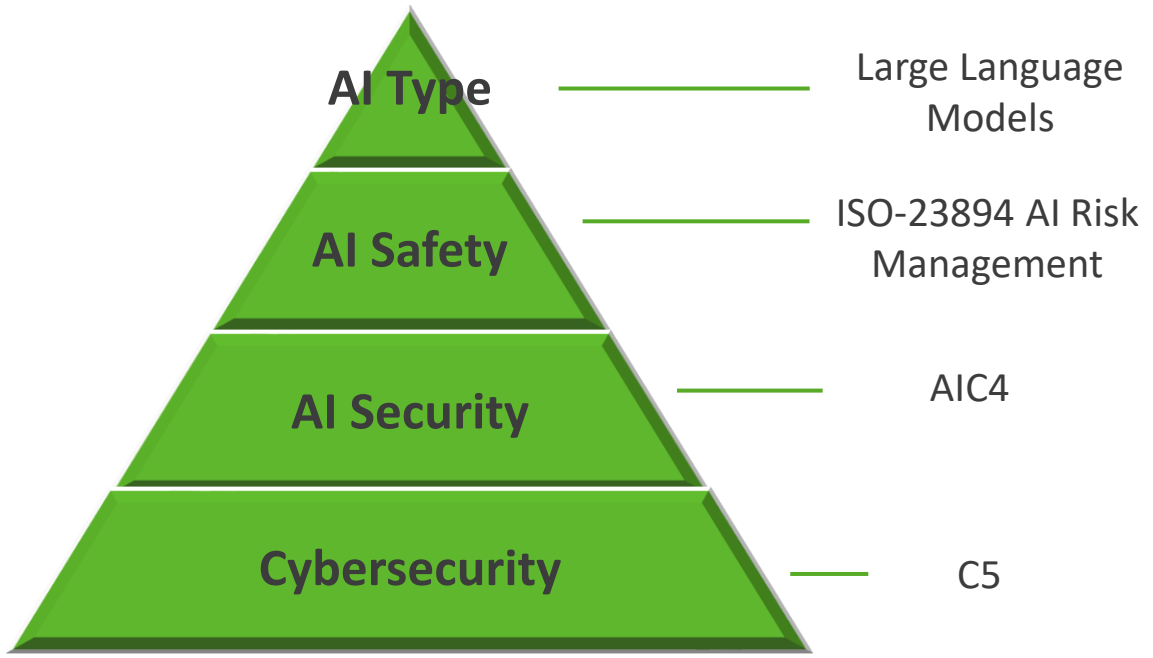
The AI / Machine Learning (ML) model is only a small part of a bigger system, regardless of the environment where is deployed. Ensuring the cybersecurity of the model does not imply having a secure AI product.



Real-world production ML system, Overview of ML Pipelines, Google.

VISION ON AI CYBERSECURITY CERTIFICATION

- **High-risk AI** applications: special focus on applications classified as high-risk according to the EU AI Act
- **Data-centric ML**: data is key in successfully training, validating and testing ML models. Hence, it becomes critical to assess data quality and protect it against attacks in the different ML life cycle stages.
- **AI Type-specific** certifications could also be developed, depending on the ML algorithm nature (e.g. LLMs, CV nets, statistical-based, etc.,)
- **Holistic approach** addressing general Cybersecurity, data quality and ML models and including all AI trustworthy aspects.



Vision of a multi-layer AI product cybersecurity certification

HOW CERTIFICATION CAN HELP ENSURING THE CYBERSECURITY OF AI SYSTEMS?

1. Defining methodologies to **audit processes** conducted by AI users and suppliers to build, test and operate AI models, identifying and **assessing vulnerabilities** and applying corrective measures.
2. Specifying methods and test tools to **measure the robustness of ML** models against adversarial attacks such as poisoning, evasion, model inversion, etc.
3. Enforcing **Risk Assessments,** both Cybersecurity and AI-specific, to identify, analyze and evaluate vulnerabilities risk and impact, and drafting corresponding risk mitigation plans.
4. Establishing **AI Management Systems** as well as defining methodologies to audit them along the entire AI Life Cycle.

NEEDS AND CHALLENGES OF TESTING LABS

Main needs:

- **Harmonized standards**, addressing holistically Cybersecurity and AI/ML models
- **Assessment methodology**, with metrics (e.g. to measure model robustness) and acceptance/rejection criteria
- **Software and evaluation tools** to conduct the tests e.g. ML model robustness against poisoning, evasion or model inversion attacks
- **Testing Sandbox**: physical secured space where to run the test, mimicking real operation conditions (e,g, Cloud, Edge)
- Harmonized **testing methodology** and assessment to ensure comparability between test labs
- **Qualified testers** and **auditors**
- **AI Cybersecurity Certification Scheme**

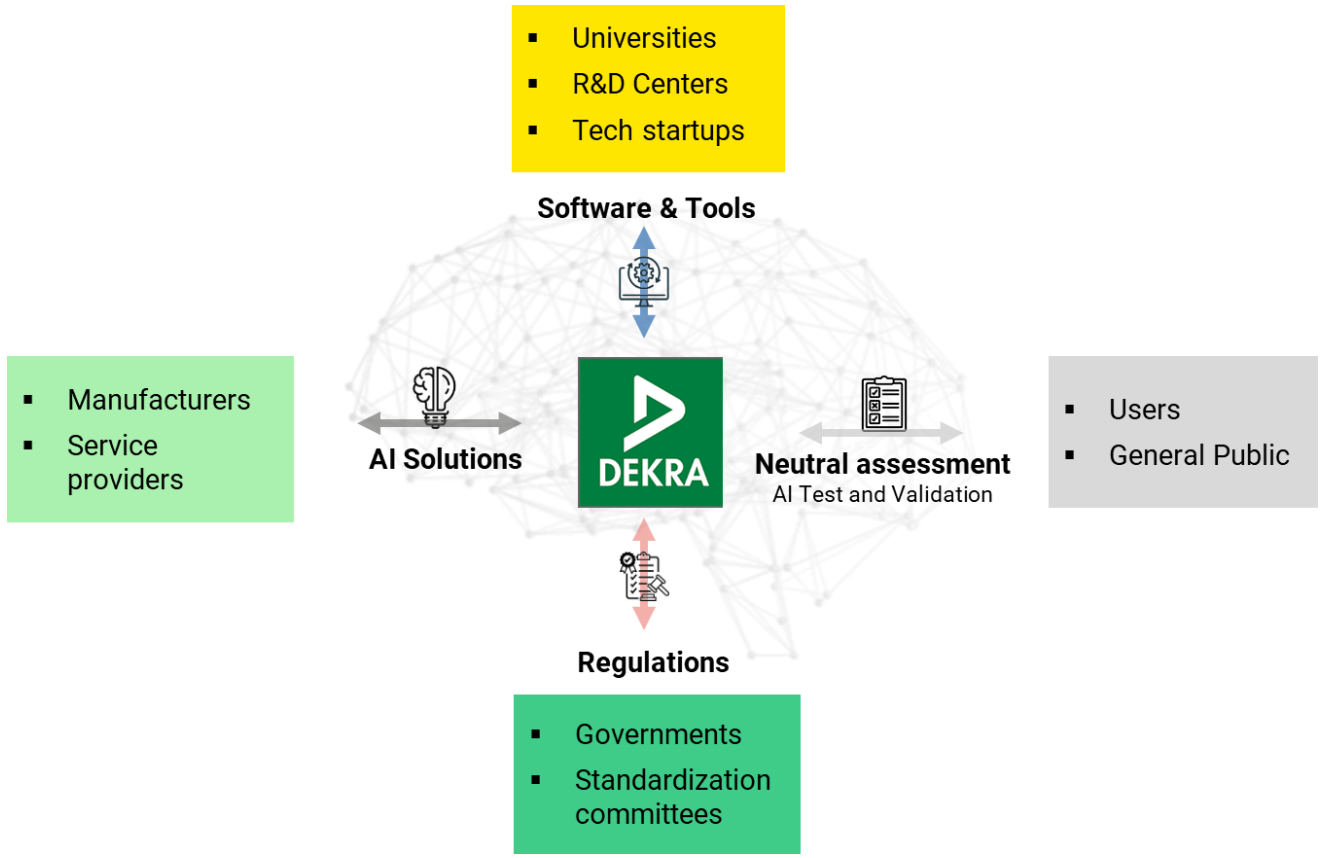
Main challenges:

- **Technical**: how to react against new attacks i.e. arms race?
- **Human**: qualified workforce with expertise both on AI and Cybersecurity

CLOSING REMARKS

- Testing and **certification** would help ensuring secure AI applications but also **building trust** of users and consumers **in AI technology**.
- A **Certification Scheme** would be highly beneficial to **implement** certifications, focusing on **AI high-risk** applications as defined by the EU AI Act.
- Technical and human challenges are still present. To overcome them, **cooperation between all stakeholders is key**: SDOs, governments, test labs, universities. R&D centers, manufacturers and users.
- We need to start NOW:

  Safety and **cybersecurity of AI cannot wait!**



Thank you!

Dr. Xavier Valero
Director AI & Advanced Analytics

xavier.valero@dekra.com

linkedin.com/in/xvalero

Digital & Product Solutions

innovating safety & security