European standardization landscape for Trustworthy AI & Cybersecurity

ENISA Cybersecurity Standardisation Conference 2023, Brusels, Feb 7, 2023

Dr. George Sharkov

Member of EU High Level Expert Group, Digital SME Alliance

Member ETSI TC CYBER, Vicechair ISG “Securing AI” (representing SBS)

ENISA ad-hoc group “Secure AI”, SCCG Cybersecurity Certification Schemes

Director, European Software Institute CEE & Cyber Security & Resilience Lab (Sofia, Bulgaria) – gesha@esicenter.bg

EU AI strategy and policy in practice

Technically Robust AI = Standardization

Trustworthy AI = Lawful AI + Ethically Adherent AI + Technically Robust AI

EU Ethics guidelines for trustworthy AI

- EU – High Level Expert Group < EU AI Strategy (April 2018)
- EC White paper on AI & Data strategy (Feb 2020)
  - ***Risk based approach*** *to AI requirements and recommendations*
  - *Legislative measures on data governance*
- AI Act (proposal)

- **Building Technically Robust & Trustworthy AI**
  - Transparency, Explainability, Auditability
  - Engineering perspective, requirements
  - "Securing AI" – new emerging standards – ETSI ISG SAI (since October 2019)
- **SMEs & Technically Robust AI**

AI HLEG: Assessment List -7 areas

1. **Human agency and oversight**
   - Fundamental rights
   - Human agency
   - Human oversight

     [mind self-learning, autonomous AI; "stop button"]
2. **Technical robustness and safety**
   - Resilience to attack and security
   - Fallback plan and general safety
   - Accuracy

     [required level and type of accuracy, data produced]
   - Reliability and reproducibility

     [reproduce behavior, not output]
3. **Privacy and data governance**
   - Respect for privacy and data protection
   - Quality and integrity of data

     [hidden threats, biased data for ML/DL]
   - Access to data
4. **Transparency**
   - Traceability
   - Explainability
   - Communication
5. **Diversity, non-discrimination and fairness**
   - Unfair bias avoidance
   - Accessibility and universal design
   - Stakeholder participation
6. **Societal and environmental well-being**
   - Sustainable and environmentally friendly AI
   - Social impact
   - Society and democracy
7. **Accountability**
   - Auditability
   - Minimising and reporting negative impact
   - Documenting trade-offs
   - Ability to redress

Safe and Trustworthy AI – AI/ML specific standards

EC standardization request to ESOs, Deadline: 31/10/2024

The European Committee for Standardisation (CEN), the European Committee for Electrotechnical Standardisation (Cenelec) and the European Telecommunications Standards Institute (ETSI) are requested to draft the new European standards or European standardisation deliverables listed in Table 1 of Annex I in support of safe and trustworthy artificial intelligence.

1. **Risk management system** for AI systems – ensure health, safety and fundamental rights (cover entire lifecycle, integrated to RM of overall product)
2. **Data and data governance** (quality of datasets used to build AI systems, train, validate and test, data biases)
3. **Record keeping** through built-in **logging capabilities** in AI systems (traceability, monitoring)
4. **Transparency** and information to the users of AI systems (understand and use, or interpret properly the output, capabilities and limitations, for different user (professional) profiles), explainability/explicability (?)
5. **Human oversight** of AI systems (high-risk AI systems, developer-provider-user, specific requirements / e.g. remote biometric identification)
6. **Accuracy** specifications for AI systems (metrics, levels)
7. **Robustness** specifications for AI systems (errors, faults, inconsistencies, continuous learning)
8. **Cybersecurity** specifications for AI systems (specific assets and vulnerabilities, data and models, underlying ICT infrastructure)
9. **Quality management system** for providers of AI system, including post-market monitoring process (all aspects 2-8, integrated to QMS of manufacturer, SMEs applicable)
10. **Conformity assessment** for AI systems (points 1-8, QMS by the provider, self and third-party assessment, testing)

- Horizontal - especially for high-risk
- Vertical - intended for certain specific AI systems (use cases)
- Scope/type - technology-, process- or methodology-based technical specifications

Trustworthy & Robust AI – the engineering perspective (specific ICT/SW systems)

Quality of AI (Robust AI) =

Quality of "knowledge"

\+ Quality of Data (learning – ML/DL, operation)

\+ Quality of technology (standard + specific for AI)

\+ Quality of software / hardware

\+ (Cyber) security for AI

+ new business models and processes – ethics guidelines

+ new compliances – standards

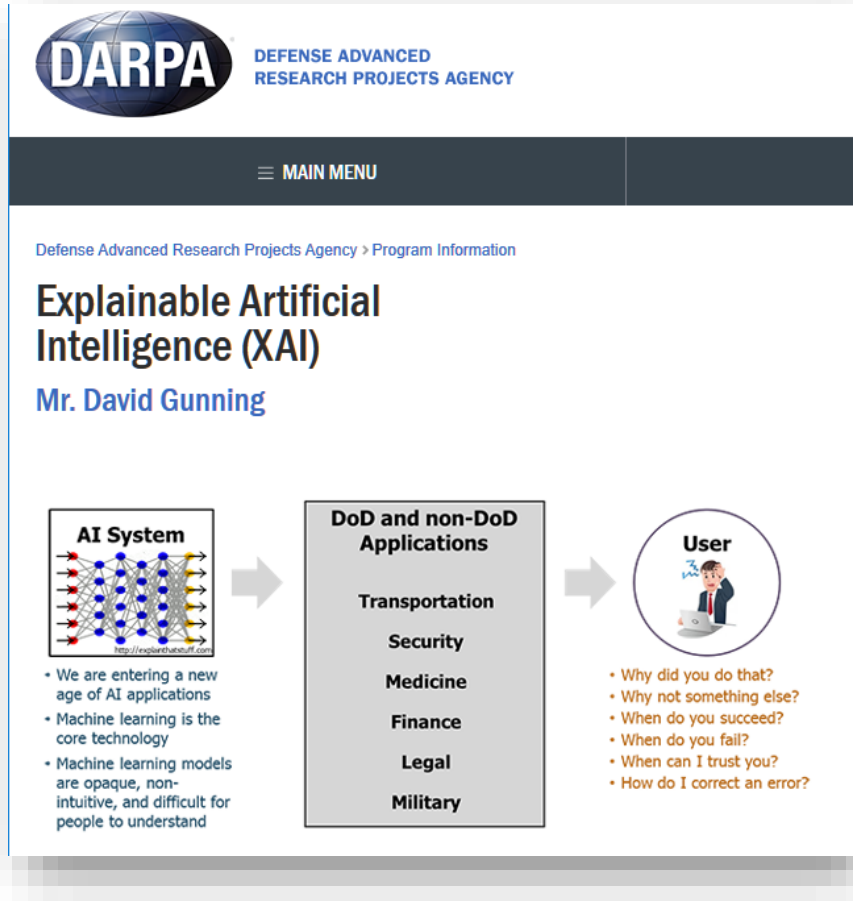
+ new (adapted) legal base

Examples:

AI systems & safety = "supervising" any ICT / SW systems (e.g. SCADA, ICS)

AI systems and autonomous defense/weapon systems:

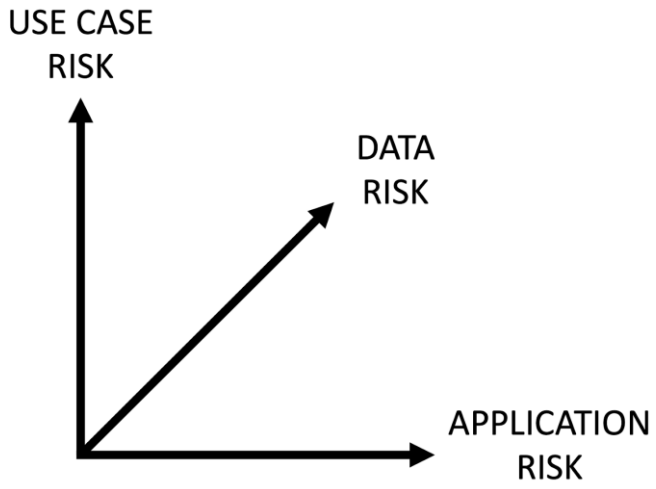
need for Explicable/Explainable AI



Problem Statement

- AI ACT is horizontal regulation, applying across many sectors, because:
  - cross-cutting nature of AI technology
  - need to ensure regulatory consistency and equal expectations across domains
  - need to make legislative burden and liability issues manageable
- So: standards need to look at the risks and the technical mitigation common across the various ("high-risk") AI application areas
  - **How to define such common standards and conformity tests?**
- But: some sector-specific ("vertical") considerations may be needed ?
  - **How to define application-specific risk analysis and mitigation?**

Cybersecurity and AI – synergy and complementary with other (EU) standardization landscape (rolling plan)



The holistic view would require then a three-dimensional risk-based approach, considering:

- Use cases risks – the high-risk as defined in the AI Act
- Data risks – adapted provisions of GDPR, the new Data Act and Data Governance Act
- The applications (technology related) risks – mainly addressed by standards

PLEASE VISIT ETSI: *https://portal.etsi.org/TB-SiteMap/OCG/OCG-AI-Co-ordination*

ETSI has work in multiple domains (data from June 2020)

◔ = considering ◑ = started ◕ = strong	3GPP	EP eHEALTH	ISG ARF	ISG CIM	ISG ENI	ISG MEC	ISG NFV	ISG SAI	ISG ZSM	oneM2M	SC EMTEL	TC CYBER	TC INT AFI WG	TC SmartM2M	TC MTS	ISG PDL
Terminology					◕			◕	◕	◔			◕	◔	◕	
Use cases	◔	◔			◕	◑	◑	◑	◕	◔	◔		◕	◔	◕	
Impact of EU ethics guidelines		◕						◑		◔				◔	◔	
Trustworthiness & Explainability		◑						◑	◔	◔				◑	◔	◔
Security/privacy		◕		◔	◔			◑	◕	◔		◕	◔	◔	◔	◑
Acrchitectures and RPs			◔		◕	◑	◑	◑	◕	◔			◕	◕	◔	
Management of AIs					◔			◑	◕	◔			◕	◑		
Dataset requirements and quality		◕		◑	◕		◑	◑	◔	◔			◑	◑	◔	◑
Interoperability		◑			◕		◑	◑	◕	◔			◕	◑	◕	
Test methodology and systems					◔		◕	◑					◕		◕	
KPIs and conformance					◔						◔		◔		◔	
System maturity assessment			◔		◔								◔		◕	

ETSI aims to handle specific needs for AI:

- to harness AI for optimization of ICT networks,
- to include ethical requirements in AI usage e.g. for eHealth, privacy/security
- to ensure reliability through appropriate testing of systems using AI,
- to overcome some AI-related security issues, and
- to better manage and characterize data, including from IoT systems, that is used by AI.

https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp34_Artificial_Intelligence_and_future_directions_for_ETSI.pdf

EU Cybersecurity and other regulations & AI

EU Cybersecurity Strategy . The importance of cybersecurity for AI was explicitly addressed in the strategy:
"Cybersecurity must be integrated into all these digital investments, ***particularly key technologies like Artificial Intelligence (AI)*** *, encryption and quantum computing, using incentives, obligations and benchmarks."*
NIS2 expands the scope of application and strengthens the obligations of management bodies.

EU AI Act

Recitals 48-51: high-risk AI systems should be ***"resilient against risks connected to the limitations of the system (e.g. errors, faults, inconsistencies, unexpected situations)*** *as well as against malicious actions that may compromise the security of the AI system…"*
ENISA is currently examining the main considerations involved for developing a **cybersecurity certification scheme for AI systems** and is expected to publish a report this month (?)
The main cybersecurity-specific obligations of the Act are set out in **Article 15** , with corresponding transparency obligations in Article 13. The Act requires that high-risk AI systems have **appropriate levels of robustness, accuracy and cybersecurity** which must be maintained throughout the entire lifecycle. The exact technical solutions to be employed will depend on the circumstances and risks. These requirements overlap with existing legislation, namely the certification process as set out in Regulation 2019/881 (CSA).

Data Governance Act (DGA)
creates a framework for data sharing by strengthening mechanisms to both increase data availability and overcome obstacles to the reuse of data. Unlike GDPR, it is not solely concerned with personal data. Data intermediaries (AI/ML ?) are required to meet licence conditions designed to ensure their independence and restrict their re-use of data and metadata.

Other specific areas of AI & Cybersecurity ETSI ISG SAI Scope

Autonomous mechanical and computing entities may make decisions that act against the users/parties either by design or as a result of malicious intent. The conventional cycle of risk analysis and countermeasure deployment represented by the Identify-Protect-Detect-Respond cycle needs to be re-assessed when an autonomous machine is involved.

ISG SAI addresses 3 aspects of AI in standards domain:

1. **Securing AI from attack** e.g. where AI is a component in the system that needs defending.
2. **Mitigating against AI** e.g. where AI is the 'problem' (or used to improve and enhance other more conventional attack vectors)
3. **Using AI to enhance security** measures against attack from other things e.g. AI is part of the 'solution' (or used to improve and enhance more conventional countermeasures).