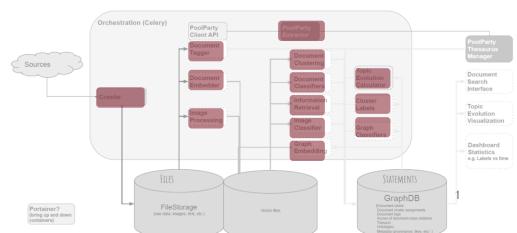
RESPONSIBLE AI

AUSTRIAN INSTITUTE OF TECHNOLOGY

Role and Responsibility of an applied research institution

- Projects with public authorities
- Decission support systems
- Systems are getting more complex
 - Past: One-project one-model / One-problem one-model
 - **Presence / Future:** One-problem many-models / One-project a lot of models
- Challenge:
 - Integrated consortial projects

Al Models - Project RAIDAR (Detect Hate Speech & Radicalization)

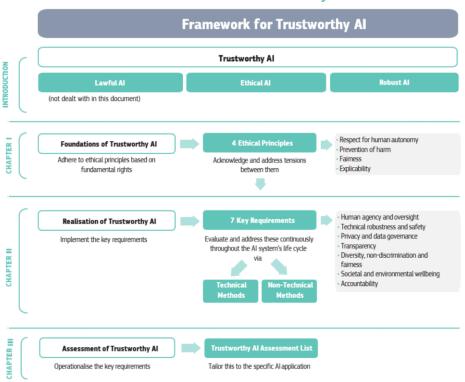


TRUSTWORTHY AI

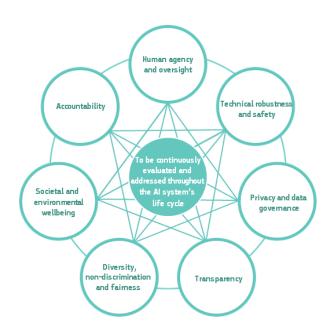
EU Ethics Guidelines for trustworthy Al



Framework for Trustworthy Al



7 Requirements of Trustworthy Al



TRUSTWORTHY AI @ AIT

Ethical Al

• AIT AI-Ethics Lab (P. Biegelbauer)

Robust Al

- Data Literacy Principles
- Software / Model Quality (ML-Ops)
- Implementing Standards

Legal Al

- Data Protection Officer (GDPR compliance)
- Research Projects
 - RAIDAR: Lawfulness by Design (by Research Institute)

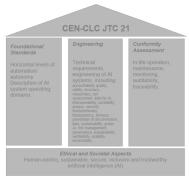
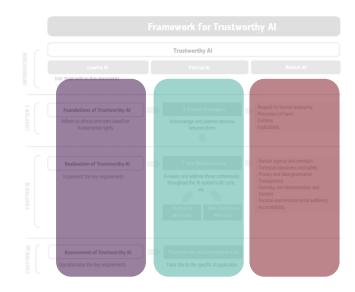
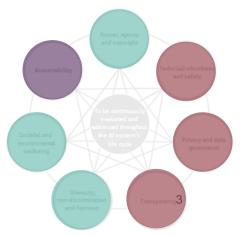


Figure 3: Overview of JTC 21 framework





Ethics guidelines for trustworthy Al

STANDARDIZATION

and implementing standards



- accountability quality
- Safety
- accuracy
- robustness
- Risk assessment / management
- Interoperability
- portability
- Privacy
- security
- Transparency
- fairness
- prevention of discrimination

- Bias
- Sustainability / green Al
- Governance
- explainability
- Verifiability
- usability
- Accessibility

Implemented (company/center level)

Projects

e.g., STARLIGHT, Green AI projects, DeepRUL, RAIDAR, defalsif-ai

Thematic Focus / Strategy 22/23



CEN-CLC JTC 21

Foundational Standards

Horizontal levels of automation/ autonomy. Description of Al system operating domains

Engineering

Technical requirements, engineering of Al systems, including: accountability quality, safety, accuracy, robustness, risk assessment, data for Al, interoperability, portability, privacy, security, trustworthiness, transparency, fairness, prevention of discrimination bias, sustainability, green Al, risk management, governance, explainability, verifiability, usability, accessibility

Conformity Assessment

n-life operation maintenance, monitoring, auditability, traceability.

Ethical and Societal Aspects

Human-centric, sustainable, secure, inclusive and trustworthy artificial intelligence (AI)