NTNU

Norwegian University of Science and Technology

TUAI kick-off meeting

Towards an Understanding of Artificial Intelligence via a transparent, open and explainable perspective

2024/10 Shen Yin

NTNU: Norges Teknisk-Naturvitenskapelige Universitet

Trondheim-Gjøvik-Ålesund

NTNU

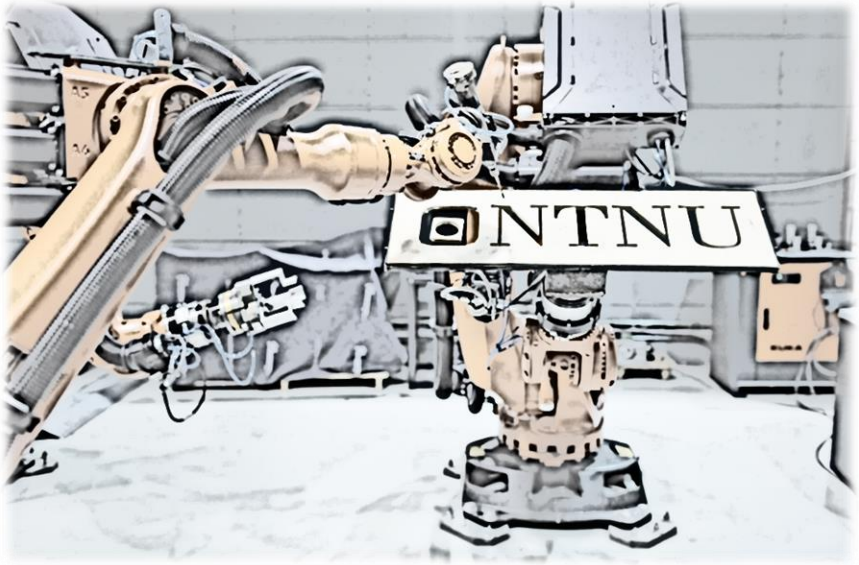
- Main profile in science and technology
- Headquarters in Trondheim with campuses in Gjøvik and Ålesund

NTNU

RAMS

NTNU

- 7 permanent faculty members cover research fields of Reliability engineering, Maintenance and management, Risk analysis, Safety and Security, etc.
- An international multidisciplinary research group
- 20+ PhDs and Postdoc
- Lead 3 Gemini Centre, lead 3 NTNU Strategic Groups
- The RAMS group is involved in industrial and scientific projects, such as SFI-SUBPRO/ZERO, FME NORTHWIND, FME HYDROGENi, BRU21, Road maintenance, 10+ EU projects, RCN, UTFORSK, NORPART etc.



Reliability, Availability, Maintainability, Safety (RAMS), Department of Mechanical and Industrial Engineering (MTP), NTNU

NTNU

NTNU RAMS

We develop theories, models, methods, and tools for the four pillar research areas:

Reliability

- System reliability and resilience assessment
- Integrity of safety-critical systems
- Reliability experiment and reliability data analysis

Risk

- Risk assessment
- Analysis of natural hazard triggered technical accidents (NATECH)
- Accident consequence analysis
- Safety barrier management

Maintenance

- Maintenance planning and optimization
- Health monitoring and asset management

Safety and Security

- Fault diagonosis, prognosis and tolerant strategy
- DT qualification
- AI assurance
- Cyber-physical and supply chain security

NTNU

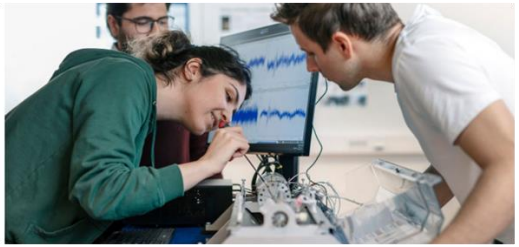
NTNU RAMS

We apply our research results in the following domains:

NTNU

NTNU RAMS

We lead two study programs:



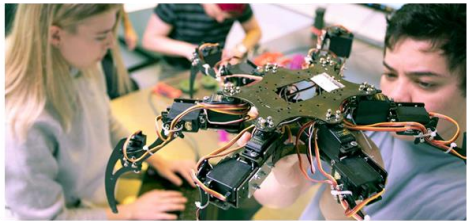
2-year International master program: Reliability, availability, maintainability, safety (RAMS)

2-year Erasmus Mundus master program: Hydrogen systems and enabling technologies (HySET)

We also involve in the following study programs:

5-year integrated master program: Machine and Energy Technology

5-year integrated master program: Engineering & ICT



3-year bachelor program: Mechatronics and product design

Background

Shen Yin, https://www.ntnu.edu/employees/shen.yin
Department of Mechanical and Industrial Engineering, NTNU

Research interests:

- Fault diagnosis, prognosis, and fault-tolerance strategies
- Safety, and security for digital (AI/DT) and technical systems
- System and control theory
- Applications in cyber-physical systems, health diagnosis, medical technology, and sustainable energy.

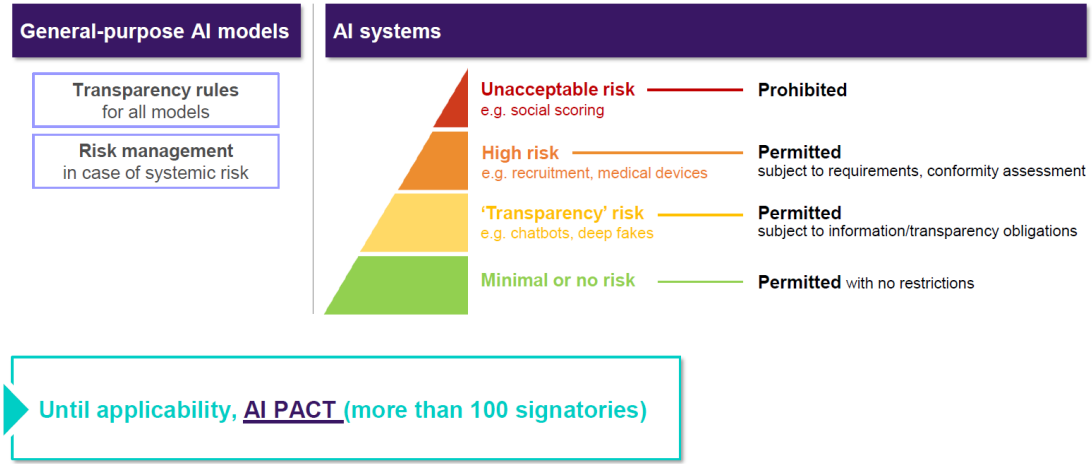
TUAI WP5

- WP5: Sustainable and Trustworthy AI Solutions for Safety-Critical Systems and Practical Applications
- The aim of WP5 is to pioneer **sustainable** and **trustworthy** AI solutions, and to target their seamless incorporation into **safety critical cyber-physical systems** and multifaceted applications. The challenges are twofold: ensuring **unshakeable trust** in the safety-critical cyber-physical platforms and facilitating **sustainable real-time anomaly detection** across a spectrum of industries. WP5 commits to navigating this complex interplay with a structured and interdisciplinary methodology.
- Shen Yin, Jia-Chun Lin, Volker Stolz, Marcin Fojcik

TUAI WP5

WP5: Sustainable and Trustworthy AI Solutions for Safety-Critical Systems and Practical Applications

AI Act provides legal certainty and is innovation friendly



Until applicability, AI PACT (more than 100 signatories)

EUROPEAN ARTIFICIAL INTELLIGENCE OFFICE

- DC11: Enhancing Trustworthy AI Integration in Safety-Critical Systems, Prof. Shen Yin
- How can we effectively develop and integrate AI into safety-critical systems in order to enhance their transparency and interpretability, ensure robustness and comply with the emerging European Union's AI safety standards, thereby significantly improving the trustworthiness and reliability of these systems?

TUAI WP5

WP5: Sustainable and Trustworthy AI Solutions for Safety-Critical Systems and Practical Applications

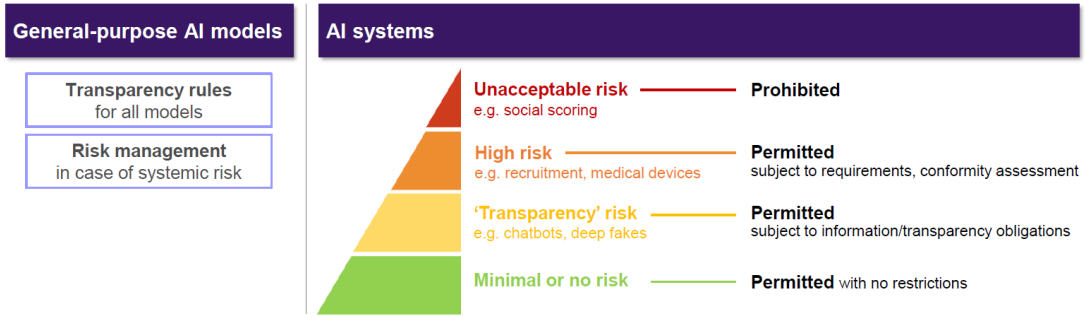
DC11: Enhancing Trustworthy AI Integration in Safety-Critical Systems, Prof. Shen Yin

Fellow *(e.g. researcher1)*	Host institution	PhD enrolment*	Start date *(e.g. Month 6)*	Duration *(e.g. 36 months)*	Deliverables *(refer to numbers in table 3.1b)*
DC11	**NTNU**	**NTNU**	**9**	**36**	
Project:***Enhancing Trustworthy AI Integration in Safety-Critical Systems*** /WP5/ **PhD Supervisor:** Prof. Shen Yin (ER11/NTNU); **Auxiliary supervisors:** David Camacho (ER1/UPM); Volker Stolz (ER13/HVL), Francesco Piccialli (ER8/UNINA); **R&D cooperation:** Hueseyin Erdogan (CONTI)					
Objectives: The primary objectives of this research are to develop AI algorithms that instill a high degree of trustworthiness within safety-critical systems in order to ensure reliability and predictability. Emphasis will be placed on improving the transparency and interpretability of the AI algorithms, thereby enabling operators and decision-makers to comprehend and trust their decision-making processes. Additionally, the research will attempt to create AI algorithms with robust fault tolerance mechanisms to maintain system stability and safety during abnormal scenarios. Furthermore, it will aim to ensure that the design and implementation of the AI algorithms align with the requirements of functional safety standards including estimating performance and meeting safety specifications.					
Expected Results: The DC11 will deliver highly trustworthy AI algorithms for safety-critical systems, which will help to accelerate the green transition. The outcomes will include practical solutions for transparent AI, which will foster trust among operators and decision-makers. Additionally, robust fault-tolerant AI algorithms will ensure system stability. The research will also provide performance assessment tools and compliance guidelines, which will improve system reliability for support during the green transition. The research result of this PhD project will be applied to the cyber-physical electricity systems in the National Smart Grid Laboratory at NTNU/SINTEF lab in collaboration with the NTNU Cyber Range. **Applied research:** Prototypes will be developed for and evaluated with cooperation with CONTI, to verify of correct operation in production conditions, improve the ongoing processes and testing procedures.					
Planned secondment(s): UPM(4 months); HVL(4 months); UNINA (4 months)				**Enrolment in Doctoral degree(s):** NTNU	

TUAI WP5

WP5: Sustainable and Trustworthy AI Solutions for Safety-Critical Systems and Practical Applications

AI Act provides legal certainty and is innovation friendly



Until applicability, AI PACT (more than 100 signatories)

EUROPEAN ARTIFICIAL INTELLIGENCE OFFICE

- DC11: Enhancing Trustworthy AI Integration in Safety-Critical Systems, Prof. Shen Yin
- DC12 Project: Sustainable Real-time Anomaly Detection for Practical Applications, Prof. Jia-Chun Lin
- DC13 Project: Trustworthy and Reliable Cyber-physical Systems, Prof. Volker Stolz