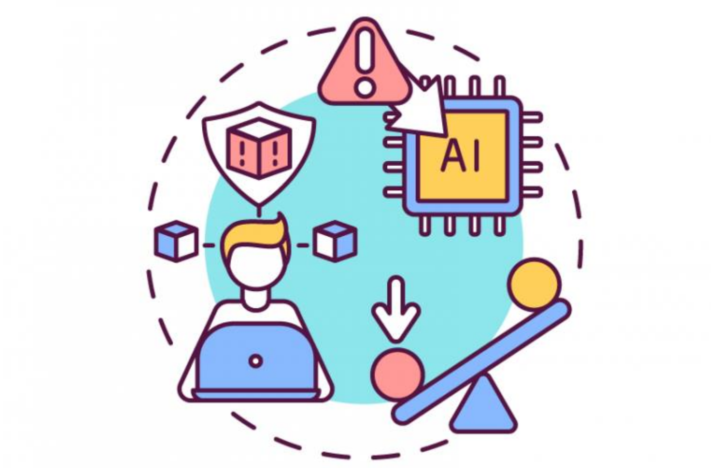
Humans, AI and Trustworthy Maintenance for Dependability

Maintaining the Future



Christos Emmanouilidis

c.emmanouilidis@rug.nl

university of groningen

University of Groningen Research & Educational Activities on Human-Centred Industrial AI

university of groningen

AIXPERT

An agentic, multi-layer, GenAI-powered backbone to make an AI system explainable, accountable, and transparent.

ai4work

ai4work.eu

humAIne

Hybrid Human-AI Decision Support for Enhanced Human Empowerment in Dynamic Situations

humaine-horizon.eu

SKILLAIBILITY

Co-designing human-centric pathways for future skills in manufacturing through augmented, empowered, inclusive, and symbiotic complementarities between AI, automation, and human tasks

https://skillaibility.eu

STAR

STAR-AI.EU

Safe and trusted human-centric AI for manufacturing

ELSA4TI: Ethical, Legal and Social Aspects of AI for Technical Industry

SMART INDUSTRY OPERATIONS course:
https://www.ai4europe.eu/education/education-catalog/smart-industry-operations

university of groningen

Maintenance you can trust

ISO/IEC TS 5723 – systems' trustworthiness

Human factors and skills

Best of both data and knowledge

Trustworthiness

university of groningen

Trustworthiness is the ability to meet stakeholders' expectations in a verifiable way

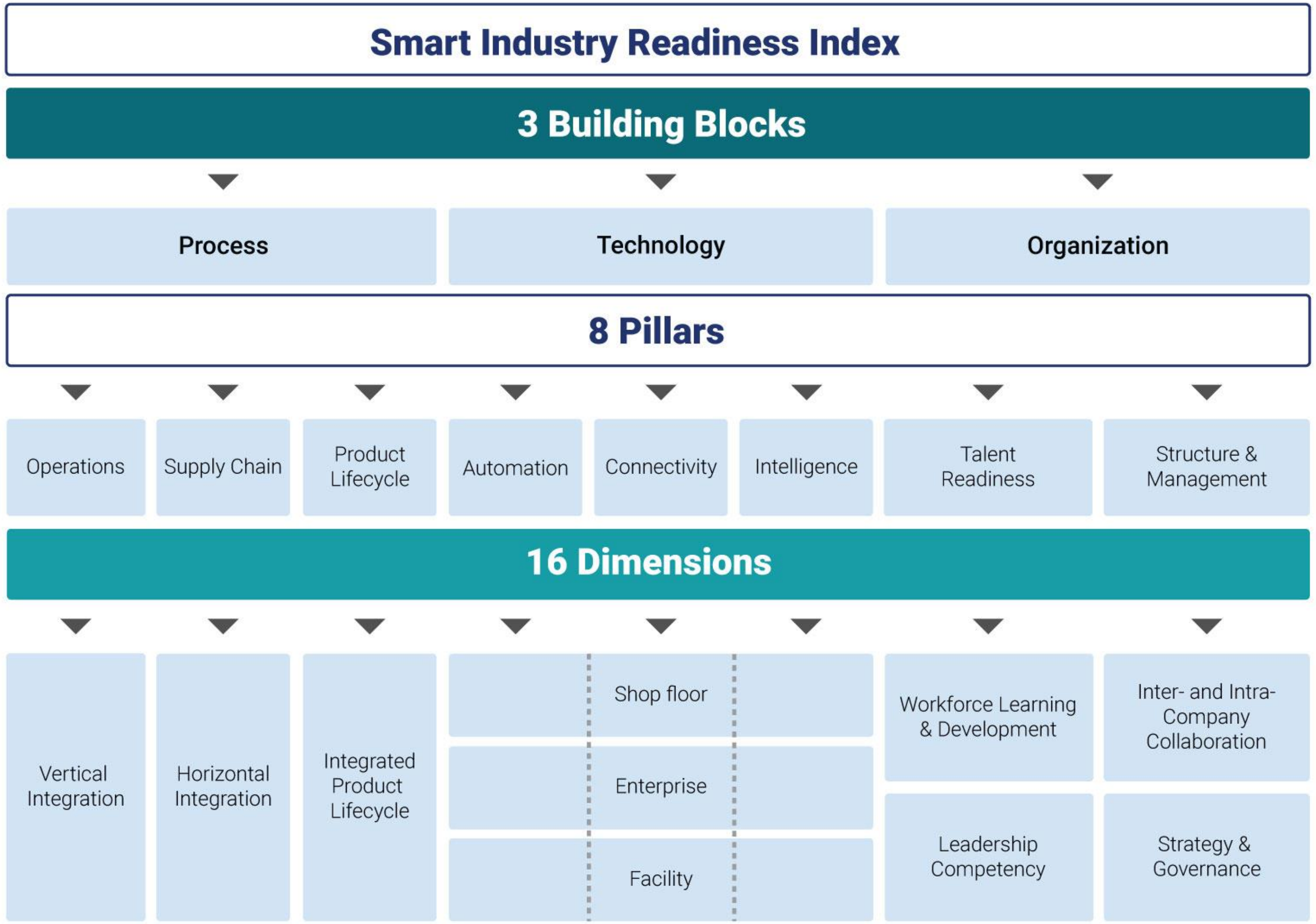
- Attribute applied to infrastructure, services, products, technology, data and information, as well as, to organizations

Trust: relational between trustor- trustee

- Misplaced trust
  - Trusting non-trustworthy suppliers for your assets
  - Trusting in non-trustworthy assets
- Failure to win trust
  - Failing to gain trust (for an asset, a process, a person), despite being trustworthy

ISO/IEC TS 5723 – trustworthiness

Smart Industry – SIRI Index

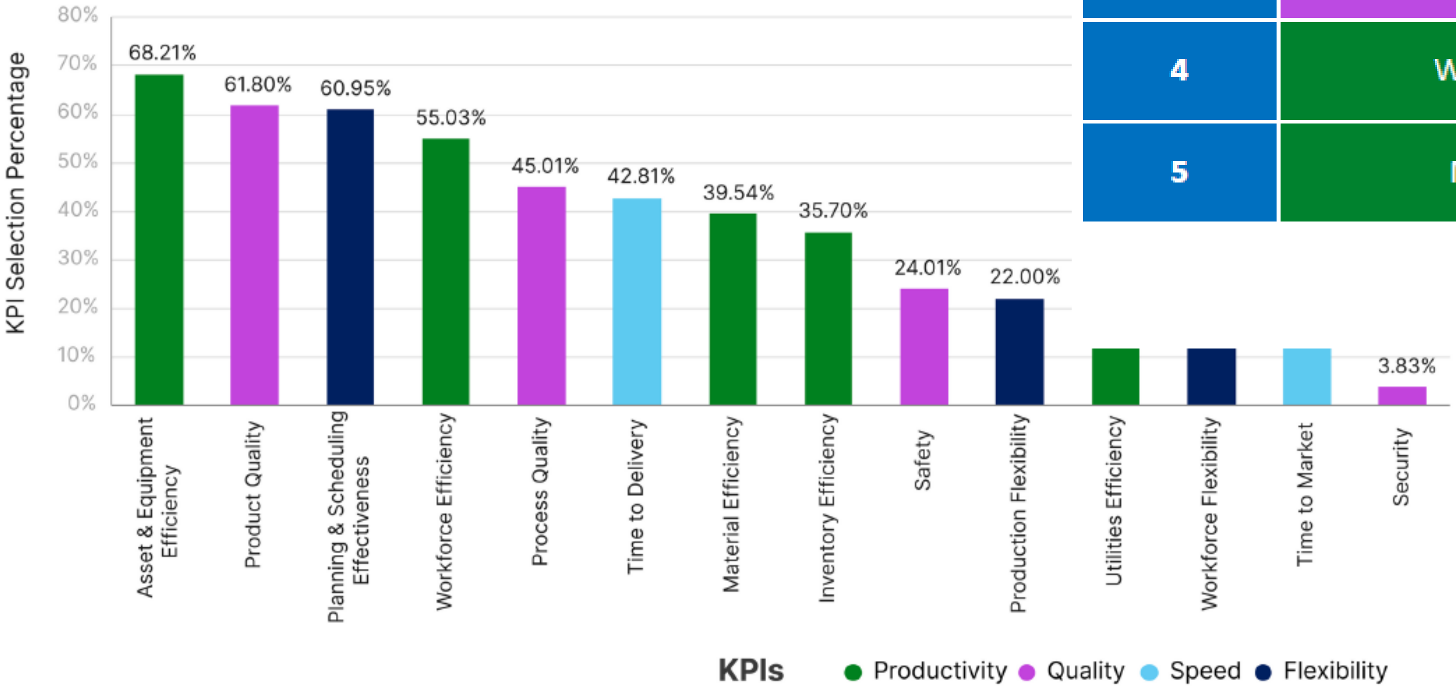


- Created by the Singapore Economic Development Board (“EDB”) in partnership with leading technology companies, consultancy firms, industry and academic experts
- SIRI was developed to address the adoption of Industry 4.0 challenges by equipping the companies with practical knowledge about:
  - What Industry 4.0 is and the tangible benefits it could yield;
  - The maturity levels of their organizations and facilities; and
  - How they can improve in a targeted and incremental manner.

Source: Smart Industry Readiness Index:
The Prioritisation Matrix, EDB Singapore, 2019 / incit.org

Trustworthiness and meeting stakeholder's expectations: what are they?

ISO 2240x and EN 62261-1:2013
Manufacturing Operations KPIs



Top KPIs	MNCs	SMEs
1	Asset & Equipment Efficiency	Asset & Equipment Efficiency
2	Planning & Scheduling Effectiveness	Product Quality
3	Product Quality	Planning & Scheduling Effectiveness
4	Workforce Efficiency	Workforce Efficiency
5	Material Efficiency	Process Quality

KPI Category Productivity Quality Speed Flexibility

Source INCIT:
The Global Smart Industry Readiness Index Initiative: Manufacturing Transformation Insights Report, 2025, DM Tech Academy, SMF, INCIT-

Trustworthiness questions

Business owner

- To what extent we can trust our operations to deliver on the business priorities?

Customer support

- How to respond to a customers’ complain regarding our service?

Operations

- What is the single most important change that can improve performance (KPIs)?
- Have we followed SOPs / SMPs?

Auditors / regulators

- Have production processes and outputs been within regulatory boundaries?

Tech/Info Officers?

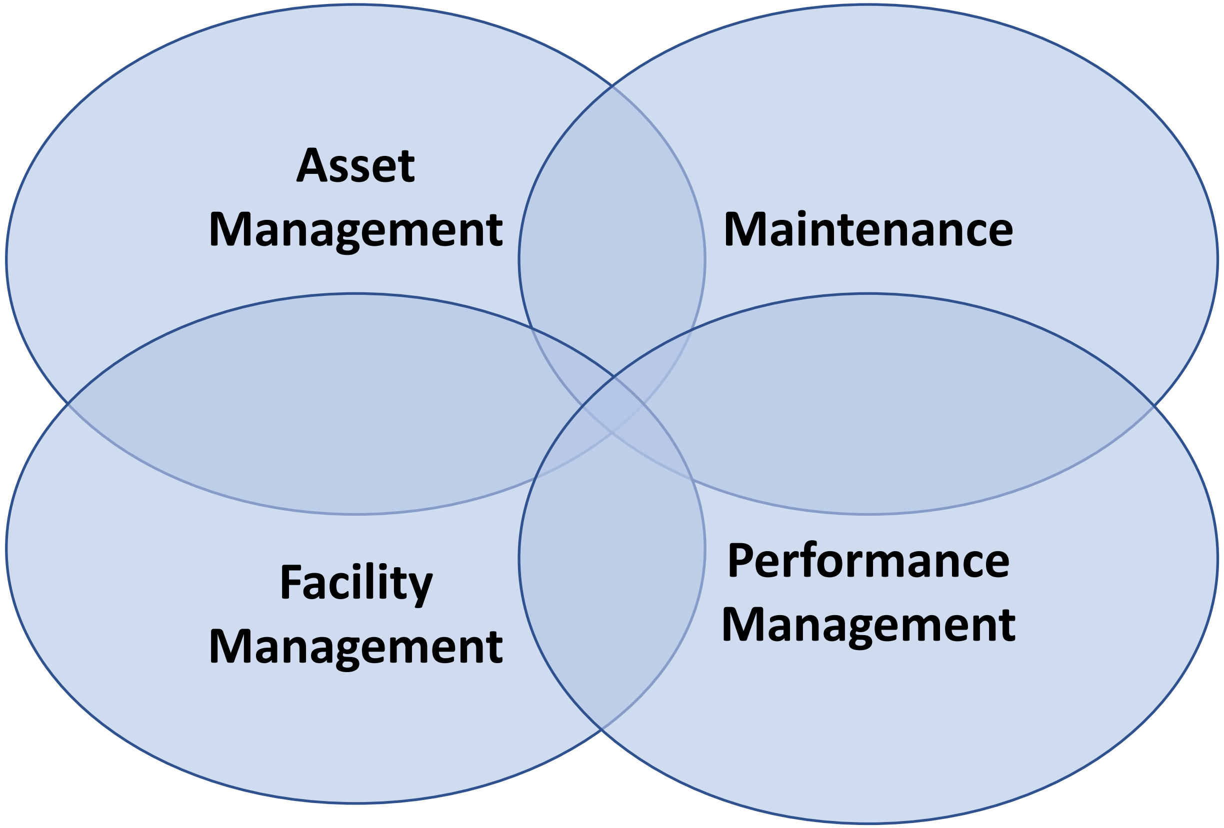
- How can we have sufficient: robustness in our controls, accuracy on data reporting and predictions; and security in our system?

Trustworthiness

ISO/IEC TS 5723 trustworthiness	

- accountability
- accuracy
- authenticity
- availability
- controllability
- integrity
- privacy
- quality
- reliability
- resilience
- robustness
- safety
- security
- transparency
- usability

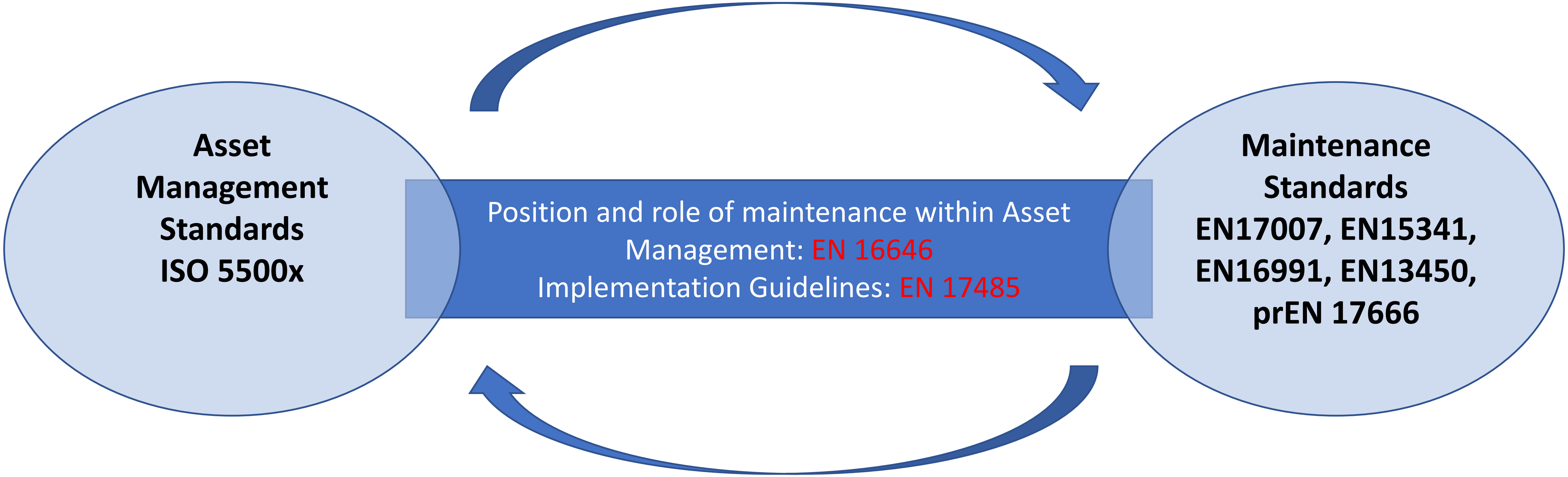
Standards in
Asset and Maintenance Management



Overview of relevant maintenance and asset management standards

Core -Processes Terminology	Management	Methods	Resources
• ISO 55000	• ISO 5500x	• EN 15341	• EN 13269
• EN 13306	• EN 16646	• EN 17385	• EN 15628
• EN 17007	• EN 17485	• EN 16991	• EN 13460
• ISO 2041	• EN 15331	• ISO 22400x	• ISO 18436x
• ISO 13372		• ISO-many CM	
		• prEN 17666	
		• EN 60300x	
		• IEC 60812	
		• EN 16602	
		• CEN TS 17385	

Maintenance within Asset Management



Data assets and decision - making

ISO 55013:2024 - Asset management — Guidance on the management of data assets

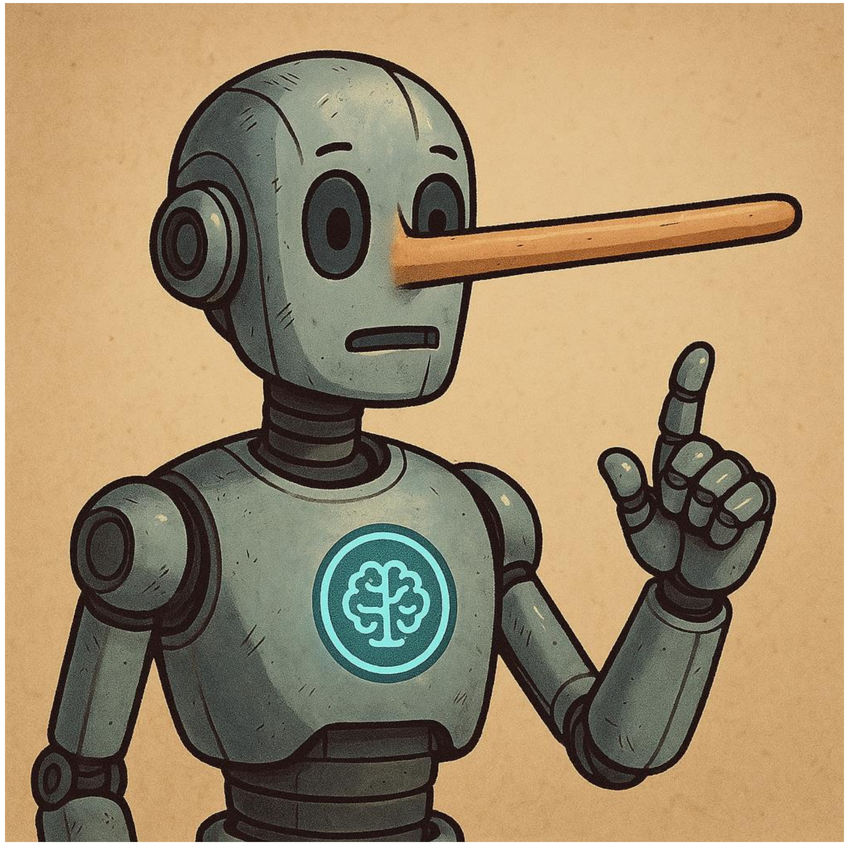
"*Guidance on the management of data assets" is an international guidance standard, which provides guidance to organizations on factors to consider in increasing and sustaining the usefulness of data assets to meet asset management objectives, and by extension, organizational objectives.*"

1. Factors and range of factors that can be generally applied to data across many types of assets in varied business contexts.
2. Definition of usefulness and guidance on how data assets become useful to an organization in relation to its objectives and in aligning asset management objectives with stated or unstated organizational objectives.
3. Definition of the types of value that data assets could hold and the types of stakeholders relevant to each type of value.
4. Alignment with other bodies of knowledge in relation to key terms such as data asset, asset data, , data quality, data governance.

Example, guidelines for:

- Delivering value from data
- Alignment of data management with organizational objectives
- Data availability and lifecycle management of data
- Interoperability and digital preservation of data assets

… and then came AI – but can we trust AI?



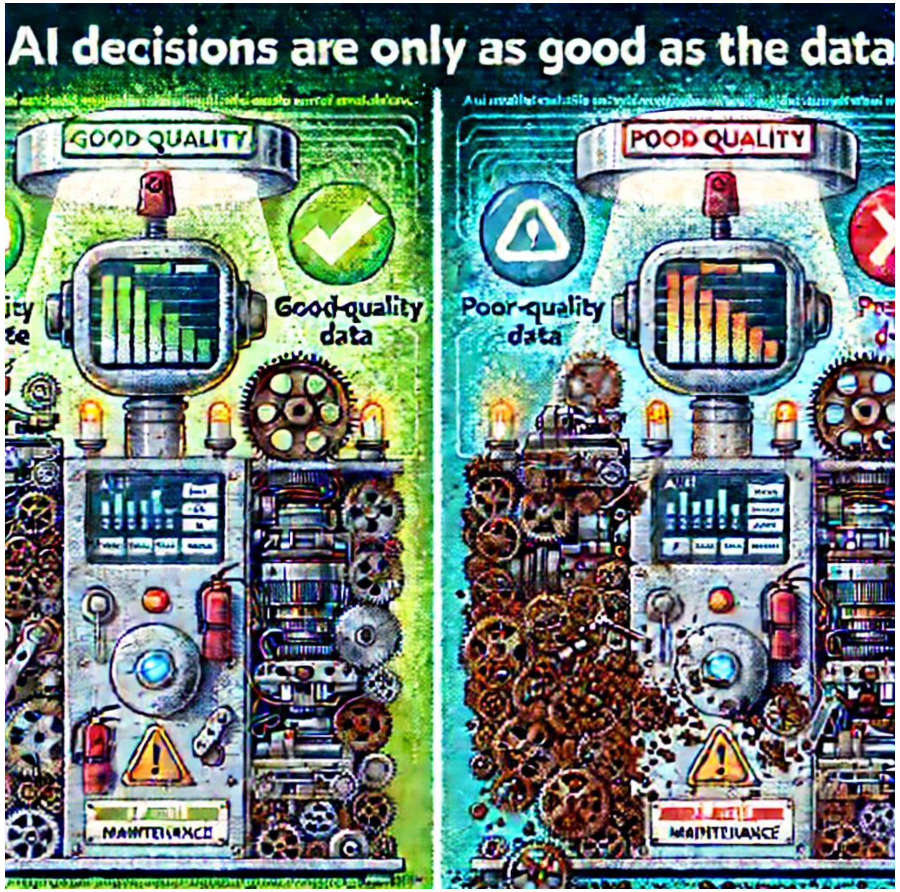
- Yes, if we use eXplainable AI (XAI)
- Do you agree?
- We used to say that you can prove anything with statistics
- We may now say that you can explain anything with XAI
- Risk: trusting non-trustworthy AI-enabled systems !
- We need more than explanations => Trust/Trustworthiness

Source: Raza et al., (2025): VLDBench: Vision Language Models Disinformation Detection Benchmark, https://doi.org/10.48550/arXiv.2502.11361

Is every application of AI in maintenance sound?

Challenges

- Data obtained from maintenance activities are scarce: not enough to cover all contexts of maintenance decisions and actions
- Data may also be of poor quality or not properly annotated
- Technical complexity of systems increases
- Knowledge from experienced operators/workers becomes elusive
- Human roles / tasks are being transformed



AI decision-making and Controllability of AI-enabled systems

ISO/IEC TS 8200: 2024

Information technology — Artificial intelligence — Controllability of automated AI systems

Controllability in engineering: the monitoring of system output to compare with expected output and taking corrective action when the actual output does not match the expected output

"Controllability is a technical prerequisite of human oversight of an AI system, so that the human-machine interface can be technically feasible and enabled. The design and implementation of controllability should be considered and practiced by stakeholders of an AI system that can impact users, the environment and societies."

Differences between generic risk management and AI system impact assessment

ISO/IEC 42005:2025

AI system impact assessment, as opposed to general risk management, addresses reasonably foreseeable impacts of AI systems:

AI Impact Assessment

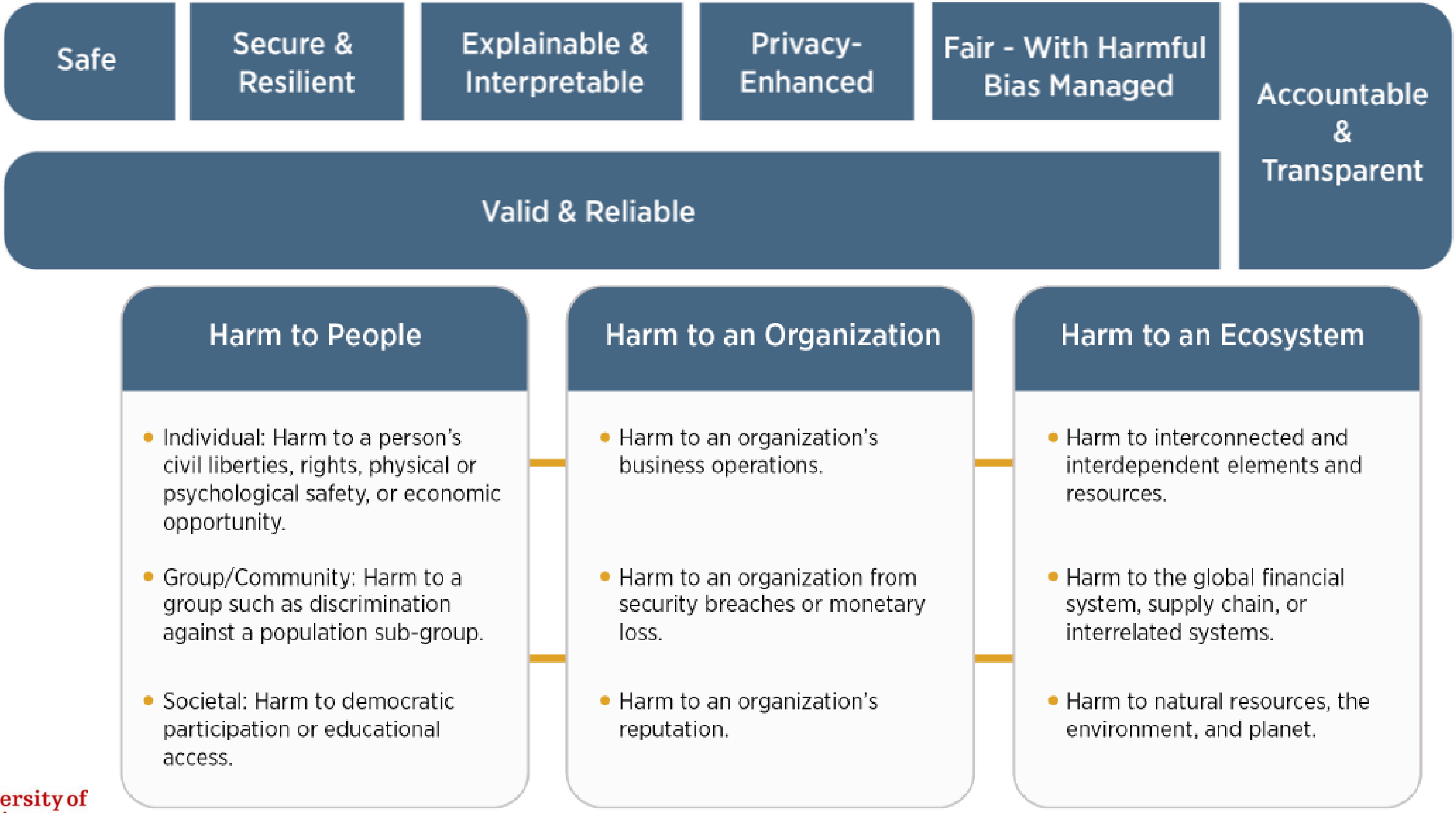
to a restricted set of parties: individuals, groups of individuals, societies and environment;

to uses of AI systems;

not to overall governance/management: does not replace overall risk management

it becomes part of overall risk management

NIST AI Risk Management Framework: Characteristics of Trustworthy AI Systems

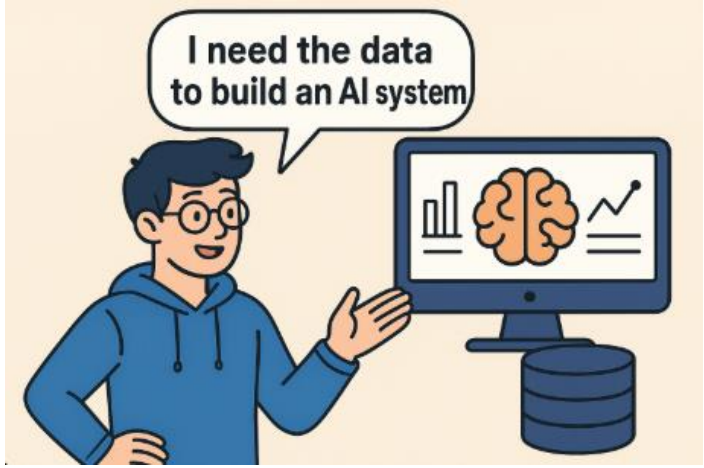


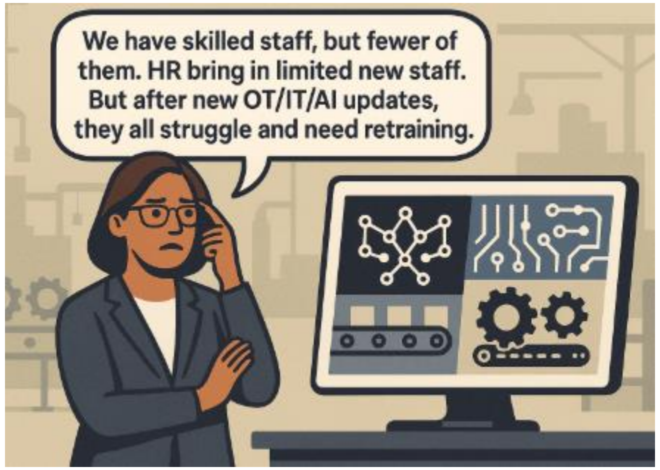
Trustworthiness and Risk

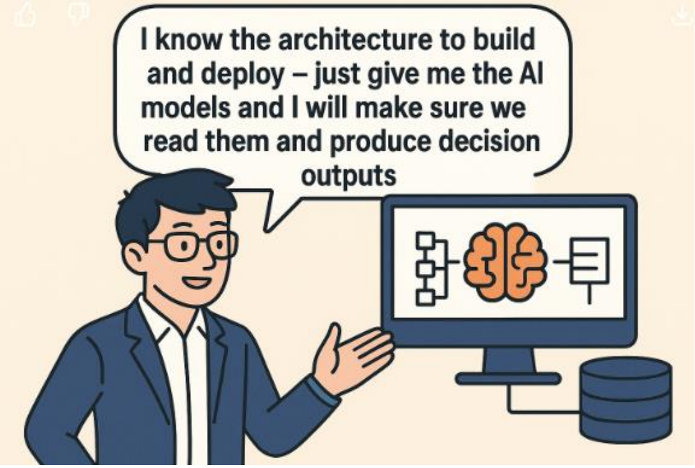
ISO/IEC 23894: **Information technology — Artificial intelligence — Guidance on risk management**

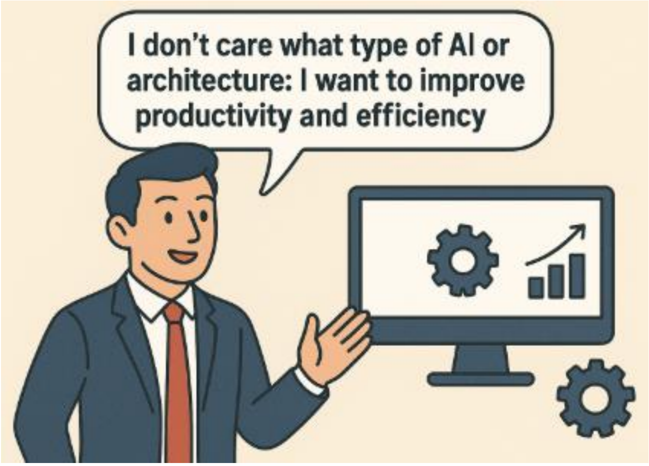
External/internal context
(potentially affected stakeholders)
— the organization (itself);
— customers, partners and third parties;
— suppliers;
— end users;
— regulators;
— civil organizations;
— individuals;
— societies.

Anyone can design, build, and operate AI-enabled maintenance systems (?)

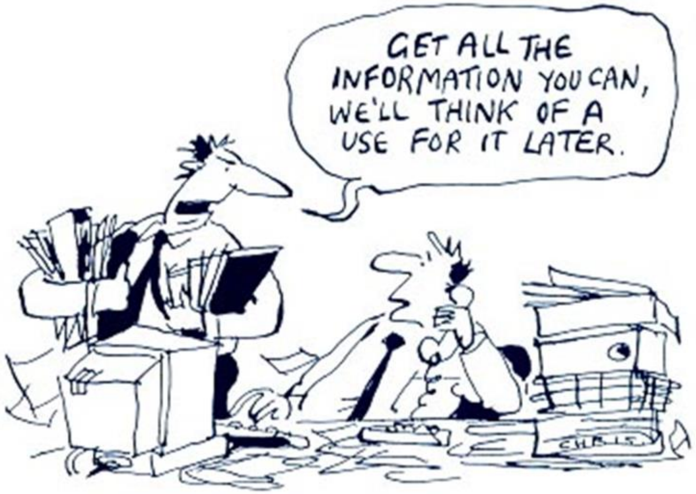








Data-driven maintenance?

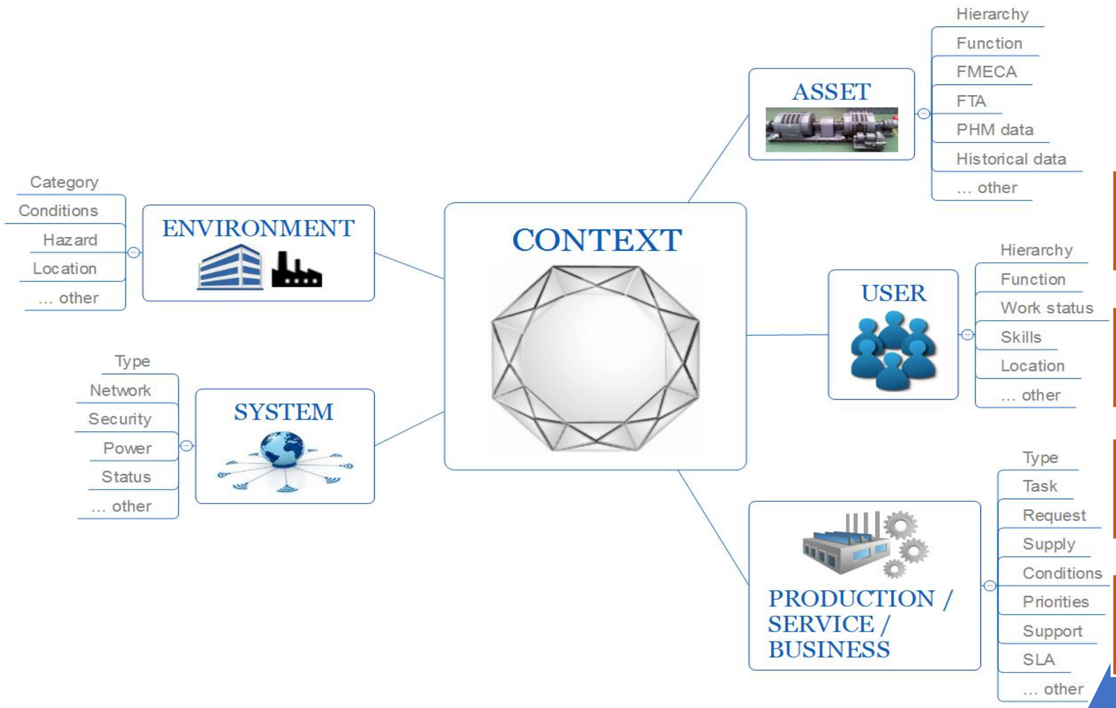


10	BML1 210(	7	280.96	3.436667	-0.48667	-0.71333	0	15.84249	-7.58333	0	-1.89333	3.82	46.2	303.5367
11	BML1 210(	7	281.17	3.25	-0.15667	-0.29333	0	15.84249	-7.35333	0	-2.50667	3.873333	45.89333	304.6399
12	BML1 210(	7	281.39	2.986667	0.193333	0.153333	0	15.84995	-7.63667	0	-3.02667	3.906667	45.6	305.7115
13	BML1 210(	7	281.61	2.706667	0.6	0.623333	0	15.84995	-7.79333	0	-3.36	3.96	45.30667	306.764
14	BML1 210(	7	281.83	2.463333	1.01	1.096667	0	15.84995	-7.30333	0	-3.41333	4.013333	45	307.7784
15	BML1 210(	7	282.05	2.206667	1.353333	1.54	0	15.85741	-6.46	0	-3.24	4.02	44.70667	308.7669

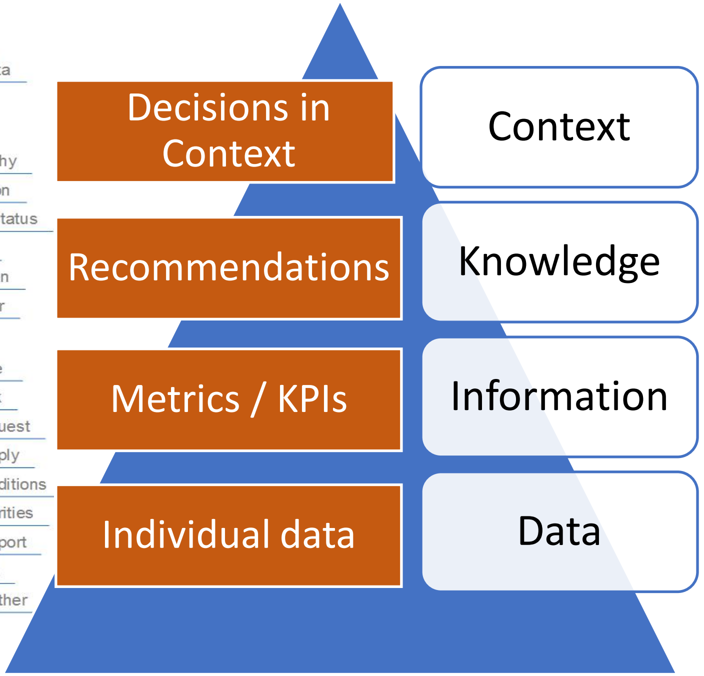
76.23291	7.5683594	0.9765625	496	432	421	505	500
100.09766	0.9765625	0.9765625	545	471	457	505	500
82.6416	2.6855469	0.9765625	607	518	452	491	500
102.35596	1.2207031	10.498047	622	416	474	476	500
85.38818	3.6621094	4.8828125	453	448	473	470	500
96.13037	1.2207031	0.9765625	530	421	601	518	500
102.7832	0.9765625	6.1035156	659	540	427	532	500
73.791504	0.9765625	0.9765625	548	452	491	453	500
110.16846	5.6152344	6.591797	487	425	466	522	500
82.27539	62.98828	0.9765625	615	480	479	489	500
112.854004	0.9765625	3.1738281	480	490	506	573	411
83.55713	4.8828125	5.859375	658	423	480	475	411

Sources: Network Rail, Xerox, [Autonom (Network Rail), Data-driven service innovation for large production-grade digital printers (Xerox International, Cranfield University)]

Maintenance and data in context



Enhanced value



Track quality affects performance; entering summer period with higher temperatures and increased transport load. We are performance-driven: taking into account forecasts, resources, optimal decision to plan maintenance of specific track section the second week of June 2019

Replace part of track within 6 months

Data indicates track quality deteriorates faster than expected. Send inspections team within a month for a non-routine inspection

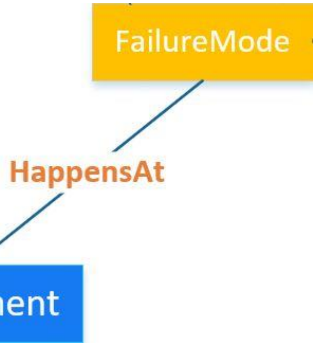
Data taken on 25/5/2019 from the 4th 1/8th of mile of the track Waterloo to Southampton

10	BML1 210(	7	280.96	3.436667	-0.48667	-0.71333	0	15.84249	-7.58333	0	-1.89333	3.82	46.2	303.5367
11	BML1 210(	7	281.17	3.25	-0.15667	-0.29333	0	15.84249	-7.35333	0	-2.50667	3.873333	45.89333	304.6399
12	BML1 210(	7	281.39	2.986667	0.193333	0.153333	0	15.84995	-7.63667	0	-3.02667	3.906667	45.6	305.7115
13	BML1 210(	7	281.61	2.706667	0.6	0.623333	0	15.84995	-7.79333	0	-3.36	3.96	45.30667	306.764
14	BML1 210(	7	281.83	2.463333	1.01	1.096667	0	15.84995	-7.30333	0	-3.41333	4.013333	45	307.7784
15	BML1 210(	7	282.05	2.206667	1.353333	1.54	0	15.85741	-6.46	0	-3.24	4.02	44.70667	308.7669

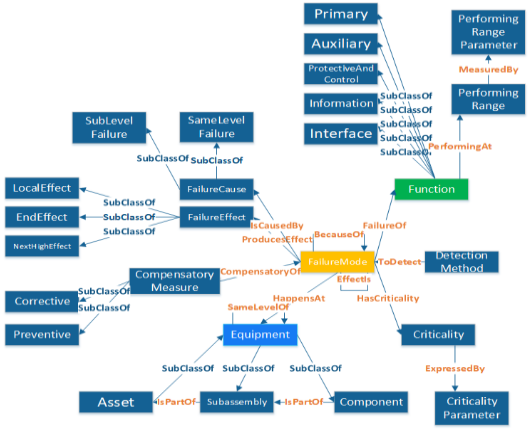
Source: Emmanouilidis, C., et al., (2019), Enabling the Human in the Loop: Linked Data and Knowledge in Industrial Cyber-Physical Systems, Annual Reviews in Control

Linking Data and Knowledge via Maintenance Ontologies

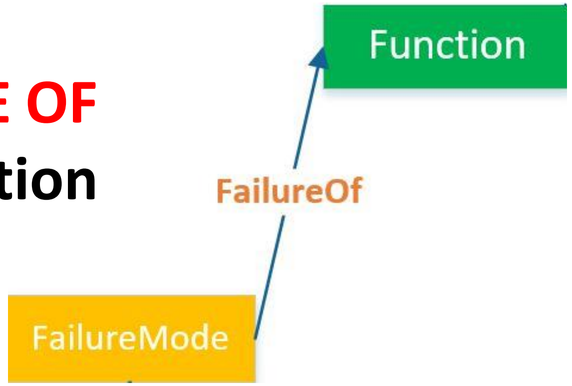
- ONTOLOGY model expresses the RELATIONS between Functions, Failures, Causes, Effects, Compensatory Measures, etc



A specific **Failure Mode HAPPENS AT one particular Equipment (Asset)**



A specific **Failure Mode IS FAILURE OF one particular Function**



e.g:

FUSING HOT NOT READY HAPPENS AT FUSER and is a **FAILURE OF FUSING**

Example of context modelling on Maintenance Services for Xerox large scale production printers
Source: Gregori, M., (2019), MSc Thesis, Cranfield University

Asset management competences

Business domain

Business domain awareness
Demand forecasting
Asset requirements planning
Financial analysis
Formulating a business case
Asset budgeting

Technical domain

Technical knowledge of specific industry
Asset information systems
Awareness of current technical developments

Asset domain

Acquisitions management
Life cycle planning
Logistics
Risk, Reliability, Maintenance
Change

Projects / People domain

Project management
Negotiation
Communication
Team work
Personnel development

GFMAM View on Asset Management

39 Asset Management Subjects

Strategy and Planning	Asset Management Policy
	Asset Management Strategy & Objectives
	Demand Analysis
	Strategic Planning
	Asset Management Planning
Asset Management Decision Making	Capital Investment DecisionMaking
	Operations & Maintenance Decision-Making
	Lifecycle Value Realisation
	Resourcing Strategy
	Shutdowns & Outage Strategy
Lifecycle Delivery	Technical Standards & Legislation
	Asset Creation & Acquisition
	Systems Engineering
	Configuration Management
	Maintenance Delivery
	Reliability Engineering
	Asset Operations
	Resource Management
	Shutdown & Outage Management
	Fault & Incident Response
	Asset Decommissioning and Disposal

Asset Information	Asset Information Strategy
	Asset Information Standards
	Asset Information Systems
	Data & Information Management
Organisation and People	Procurement & Supply Chain Management
	Asset Management Leadership
	Organizational Structure
	Organizational Culture
	Competence Management
Risk and Review	Risk Assessment and Management
	Contingency Planning & Resilience Analysis
	Sustainable Development
	Management of Change
	Assets Performance & Health Monitoring
	Asset Management System Monitoring
	Management Review, Audit & Assurance
	Asset Costing & Valuation
	Stakeholder Engagement

Human – AI synergies and action affordances

Engineering the Human – Systems Interaction: the application of principles, models, and techniques to system design, taking into account human capabilities and limitations (but this is a System Engineering perspective)

Context and Situation Awareness: human actors can be more effective when they act upon a shared context ("situational awareness") of work activities.

Human-AI synergy can be part of a "**Collective activity**" view of work environments. **Emergent outcomes** of Human + AI interaction can go beyond "augmentation/assistance" and impact on the action affordances.

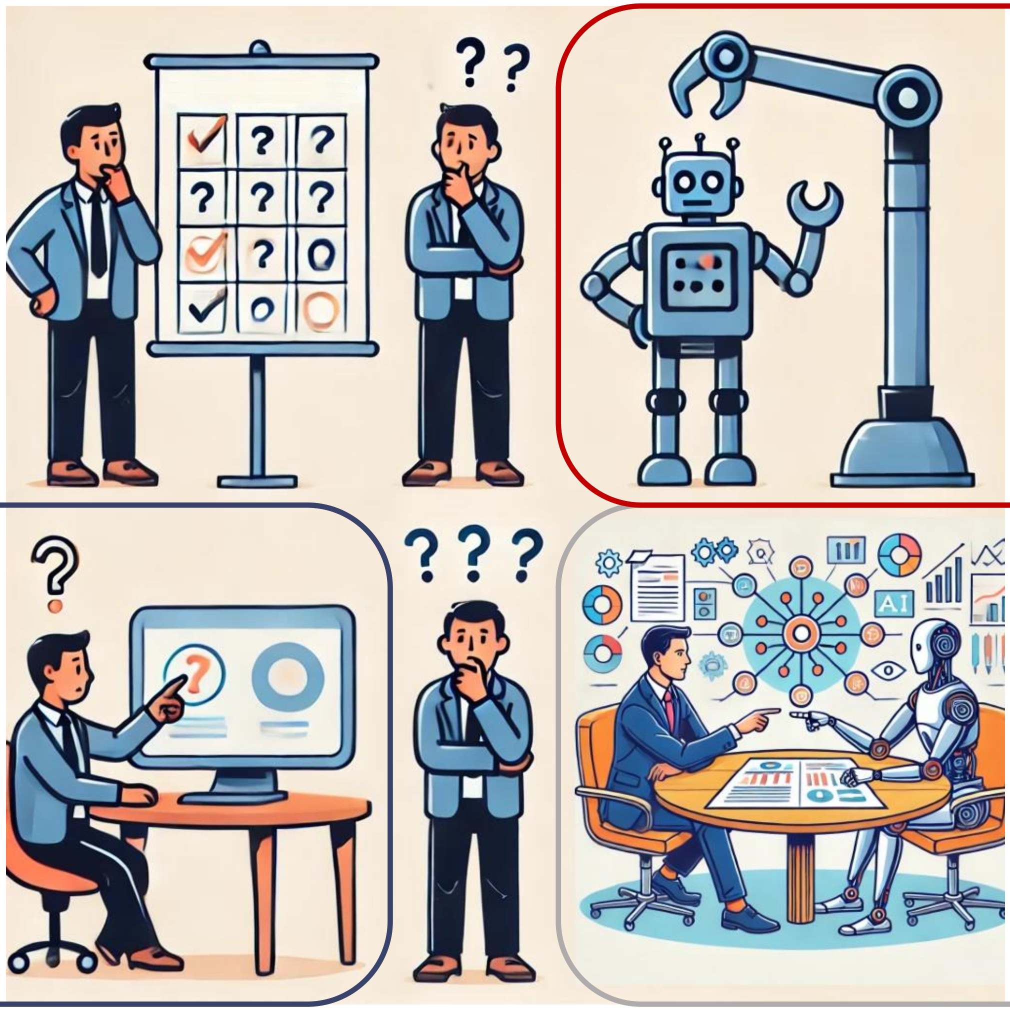
Affordance: denotes action possibilities provided to the actor by the environment

Human – AI teaming reshapes the action affordances of both humans and non-human (AI-enabled) actors

Source: Emmanouilidis, C. et al., (2021). Human in the AI Loop in Production Environments. Artificial Intelligence for Sustainable and Resilient Production Systems. IFIP APMS 2021. IFIP AICT, vol 633. Springer, Cham. https://doi.org/10.1007/978-3-030-85910-7_35

Humans and AI

Human acts without AI aid



Automated AI-enabled actions

Human acts with AI aid

Human-AI teaming

Example of human-AI teaming

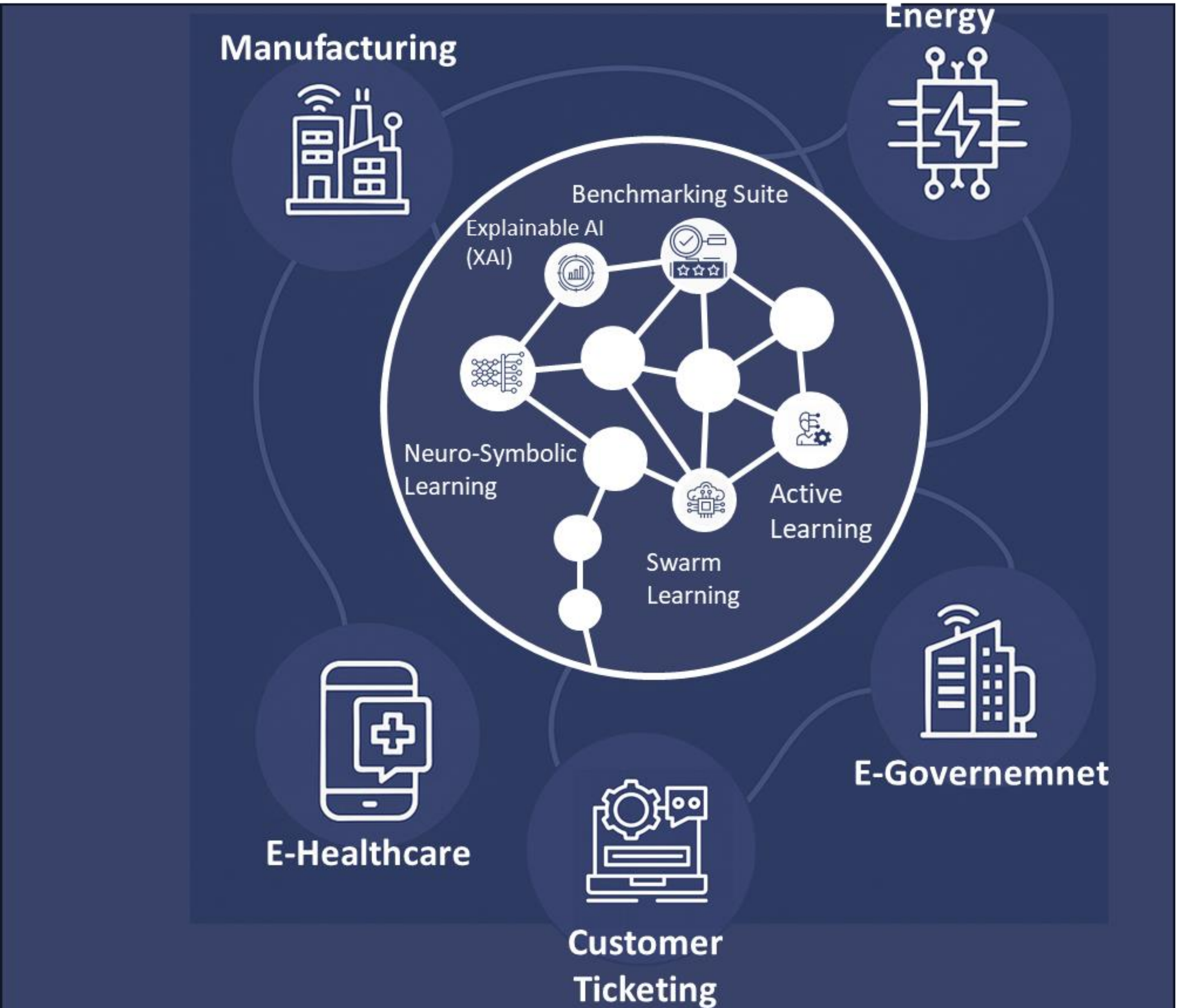
Human

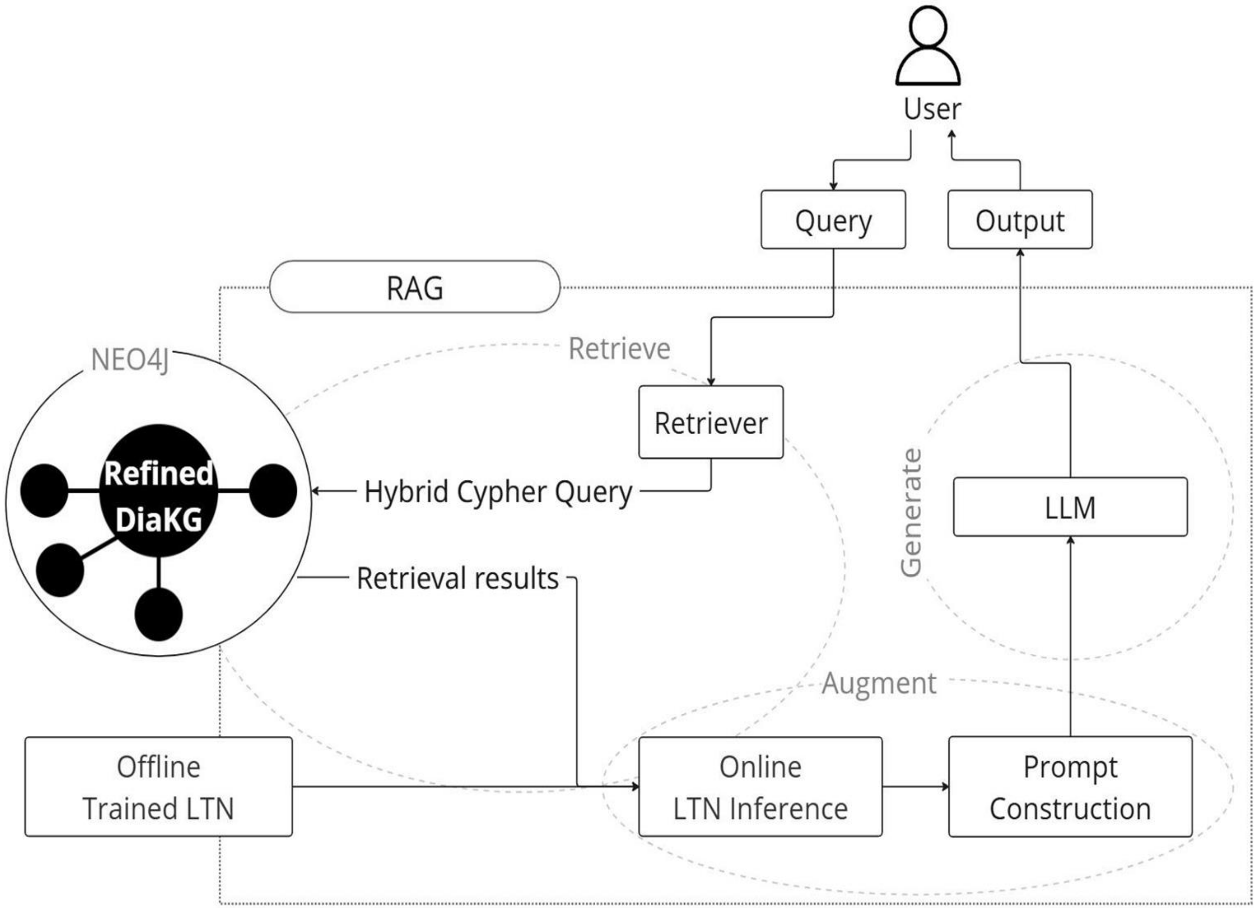
decides on the optimization objectives of maintenance (and priorities)

AI

creates plans and schedules with given objectives (and priorities)

Human – AI teaming: HumAIne project https://humaine-horizon.eu





Examples from RuG courses: RAG (Retrieval Augmented Generation)

university of groningen

A) Create Knowledge Base

Step 1: Authenticate with OpenAI

Search, drag and drop **OpenAI Authenticator**, add the flow variable connection from the Credential Configuration node. Configure the node by selecting "credentials", then execute.

Step 2: Select embedding model to create vectors

Search, drag and drop **OpenAI Embedding Connector** node and execute with default settings. This will pick the AI embedding model "text-embedding-ada-002" to create vectors from the input sentences.

Step 3: Divide PDF text into sentences

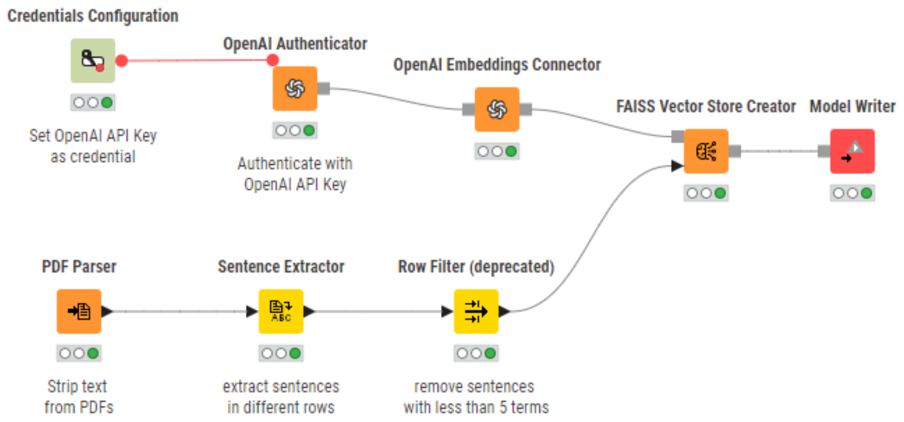
Search, drag and drop **Sentence Extractor** node and execute on the column "Document" from the "PDF Parser" node. This will split the document cell in multiple rows: one row for each sentence. Then use a **Row Filter** node to remove all sentences below 5 terms.

Step 4: Create the vector store

Search, drag and drop **FAISS Vector Store Creator** node, connect to the **OpenAI Embeddings Connector** and your string sections output. Execute the node on the column with the strings to create the vector store.

Step 5: Save the vector store

Save the vector store by adding a **Model Writer** node. To save properly you can use a relative path: selected "Relative to" then "Workflow Data Area" and specify the name of the vector store such as "vector_store.model"



B) Use an LLM for a Completion Task via the Knowledge Base

Step 1: Authenticate with OpenAI and select an LLM model

Similarly to the past exercise use the **OpenAI Authenticator** with the Credentials Configuration flow variable input.Then add an **OpenAI LLM Connector** and select "gpt-3.5-turbo".

Step 2: LLM Hyperparameter Tuning

Inside the **OpenAI LLM Connector** customize the parameters "max tokens" and "temperature".

Step 3: Adopt the Knowledge Base from the Vector Store

The Model Reader node reads from the workflow data area the vector store you previously created. Reading again from disk the vector store is good practice to learn how to deploy with a fixed knowledge base. You can also use the output of the FAISS Vector Store Creator from the past exercise.

Step 4: Connect to the Vector Store Retriever and the LLM Prompter

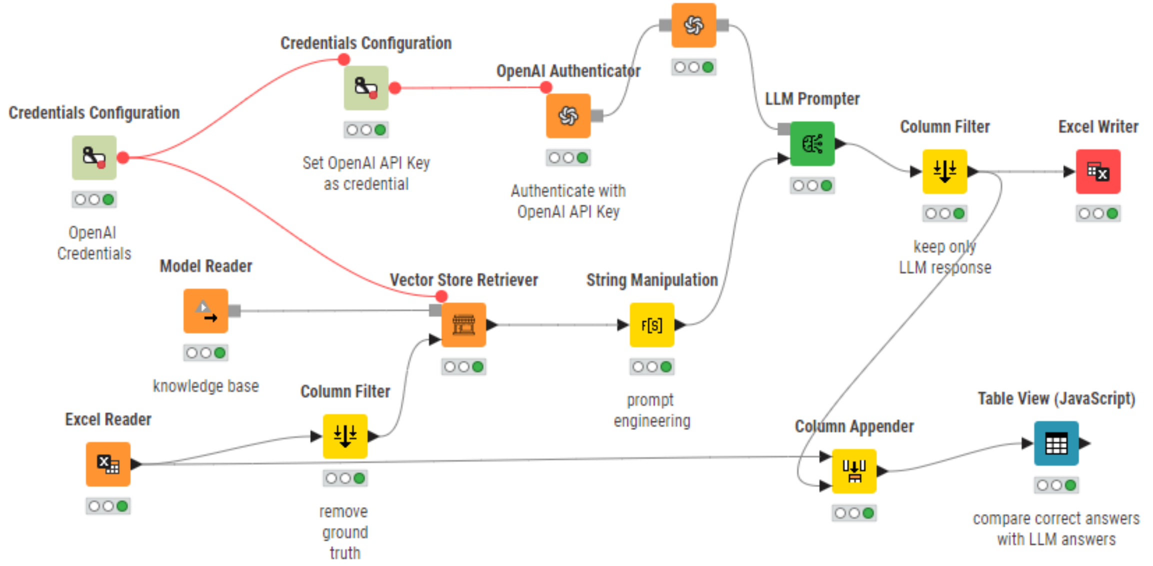
Drag in the **Vector Store Retriever** node and the **LLM Prompter** node, in-between add a **String Manipulation** node for Prompt Engineering.

Prompt Engineering Hint

double click to open

Step 5: Export Results to an Excel File

You can save the table via an **Excel Writer** node. Optionally you can compare with a **Table View** node the answers by the LLM and the ones we imported with the questions for reference.



Sliding autonomy – Sliding Work Sharing AI4Work project, https://ai4work.eu

university of groningen

KUKA

Funded by the European Union

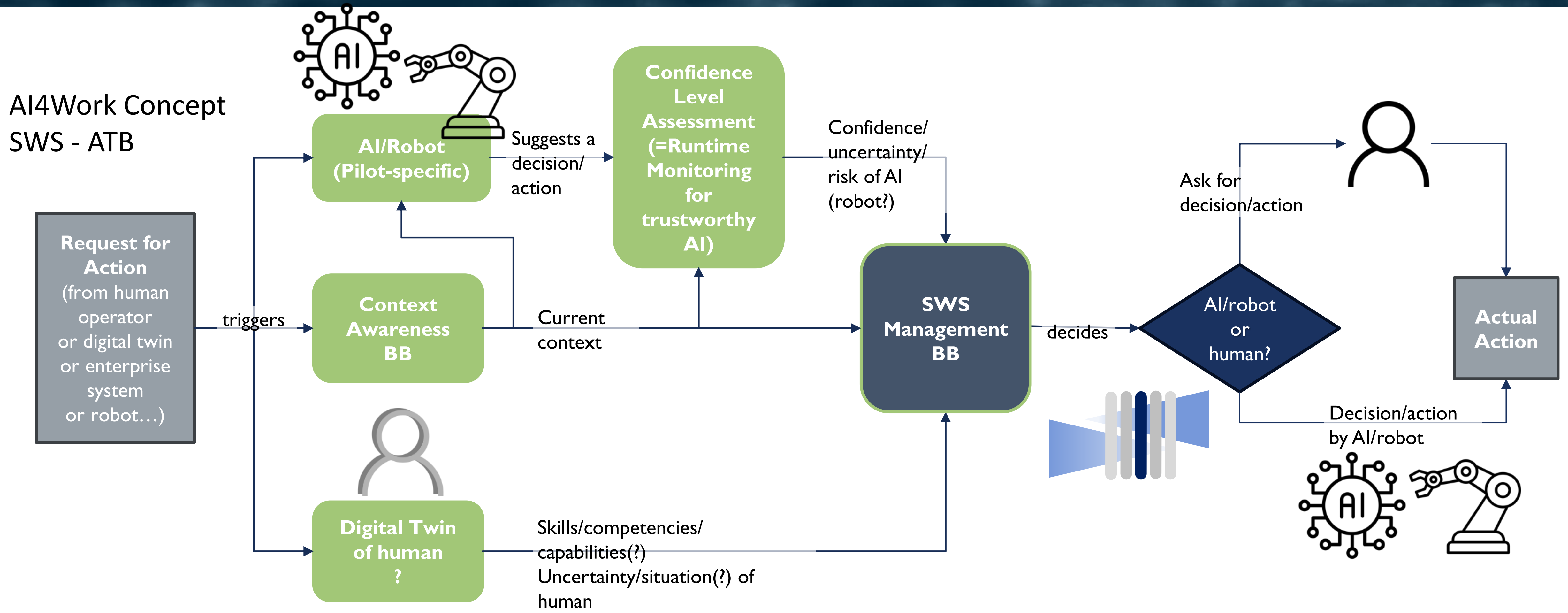
Beginning

- Plattform
- Lift
- Caterpillar
- Batteries
- Control Unit with HMI
- Robotic arm

KUKA

Endeffectors

Sliding autonomy – Sliding Work Sharing
AI4Work project, https://ai4work.eu

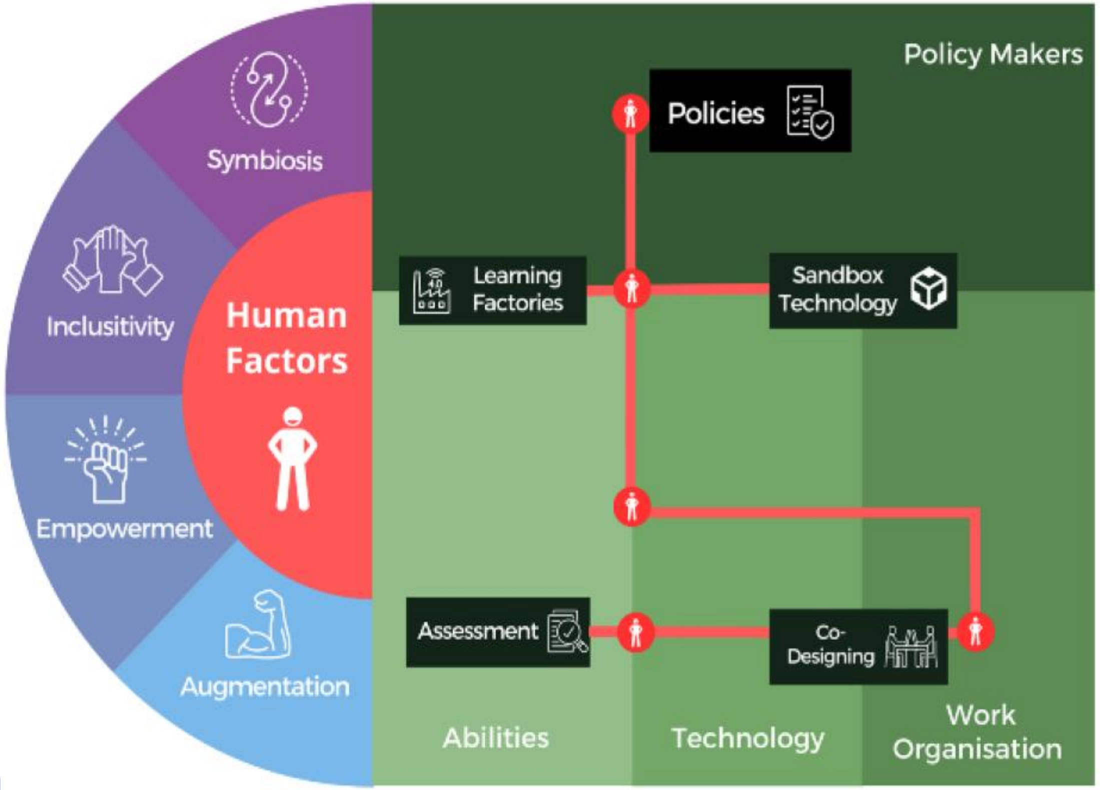


Karapanagiotis, P., Kottagaha W.M., K., Rovere, D., Bokhorst, J. A. C., Valdata, A., & Emmanouilidis, C. (2025). Enabling interoperable human–AI teaming for automation in construction and manufacturing via Digital Twins and Sliding Work Sharing ontologies. *Journal of Industrial Information Integration*, *48*, 100962. https://doi.org/10.1016/j.jii.2025.100962

SkillAIbility Project -https://skillaibility.eu

<u>Digital technologies and AI</u> do not simply create or destroy jobs: they change what people do on the job, and how they do it, shaping job profiles.

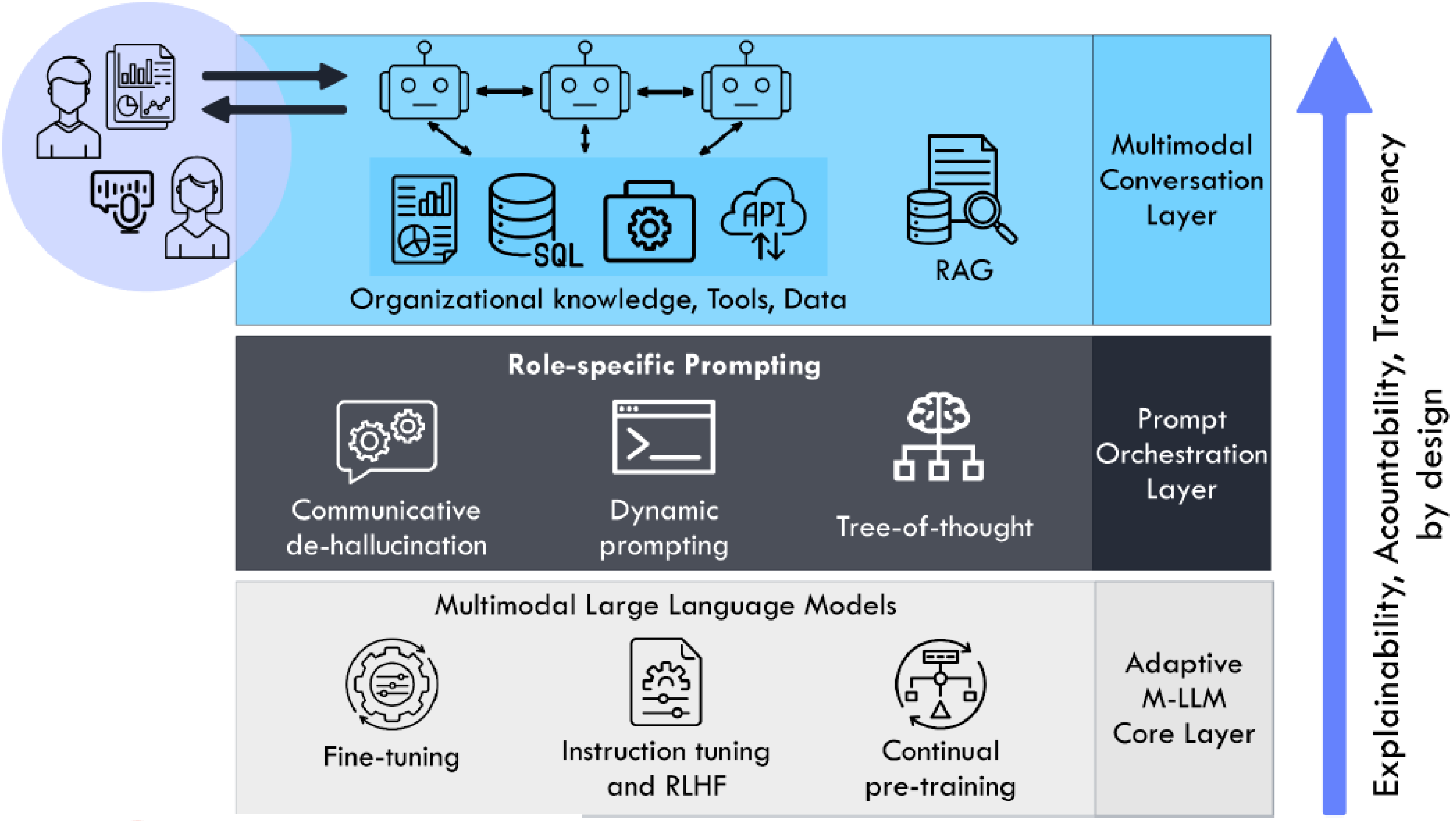
<u>Educational programs must adapt their methodologies to the evolving needs for skills</u>, preventing/reversing inequalities, and promoting equal opportunities and inclusion for all.



Funded by the European Union

AIXpert project: Agentic AI - aixpert-project.eu

university of groningen



AIXPERT

Auditing and evaluation framework for human-centric AI

Of Interest

- Customer satisfaction and warranties
- Product and production safety and security / liabilities
- Workers' safety / wellbeing
- Operators and workers assistance on the spot / training (SOP/SMP)
- Data sovereignty / governance

Solution options

- Grounding LLMs to domain knowledge
- Usage of Knowledge Graphs, capturing domain knowledge
- Retrieval – Augmented Generation (RAG) to optimise LLM output referencing domain-specific knowledge

ELSA4TI: ELSA Lab for Technical Industry

Four research projects launched for NGF-Call AiNed ELSA Labs

28 March 2025

Responsible AI requires handling its ethical (E), legal (L), societal and economic (S) aspects (A

ELSA4TI will target challenges for AI adoption in Dutch Technical Industry due to complex ELSA considerations.

ELSA4TI aims at providing practical guidance and recommendations for industry and policy-making, with a clear ELSA landscape vision.

AI-driven autonomy and humans in production and technical asset lifecycle management

AI-supported manufacturing

AI in Manufacturing and Decision Support

Funded by the European Union NextGenerationEU

Broader relevant standards

university of groningen

Manufacturing Operations

Asset/Maintenance Management

ISO 22400x
EN 62261-1
ISO 550xx
EN 13306
EN 16646, …

Human Factors

ISO 9241x
PD ISO 45001
ISO 6385

BS 8611
ISO 10075-x
ISO 27500

AI / Automation

ISO/IEC TS4213
ISO/IEC 25058-9
ISO/IEC TS4202x
ISO/IEC 38507
ISO/IEC 23894
ISO/IEC TS 8200
ISO/IEC 42005

ISO/IEC 5338, 12781
ISO/IEC 42001
ISO/IEC TR 5469
ISO/IEC TR 24027
ISO/IEC TR 24028
ISO/IEC TR 24030
IEEE P700x, IEEEP339x

Quality and Trust

ISO/IEC 250xx
ISO/IEC TS 25058 (AI-specific)

ISO/IEC TS 273

Assessing benefits of standards at organization level

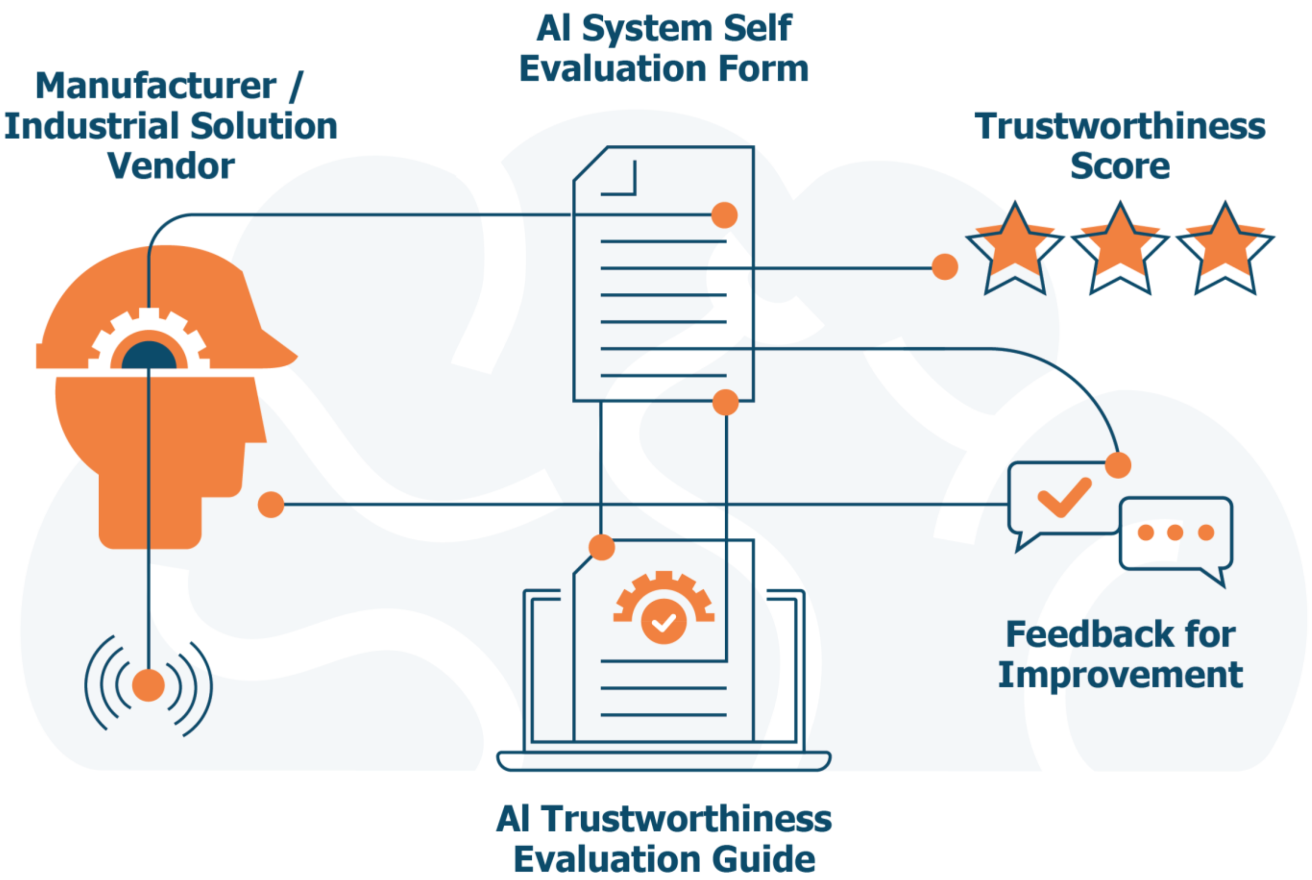
Understand Value Chain

Identify Impacts of Standards

Analyse Value Drivers and Operational Indicators

Measure Impacts of Standards

AI Solution Spaces in Industrial Maintenance



https://star-ai.eu/star-auditing-framework-trustworthy-ai

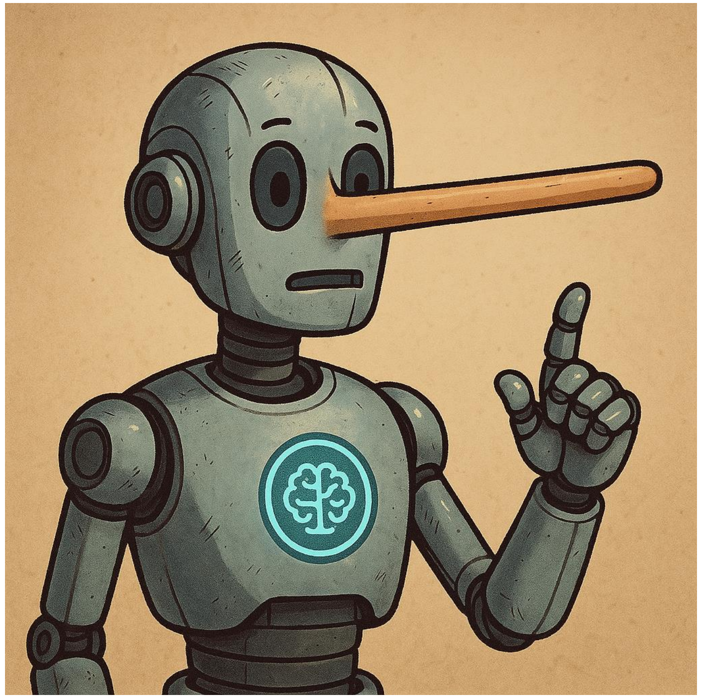
ELSA dimension

Operational dimension

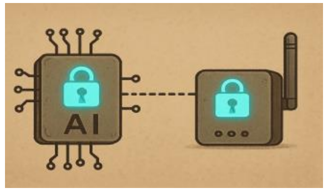
Technical dimension

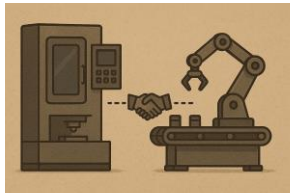
Source: https://doi.org/10.1016/j.ifacol.2023.10.1891

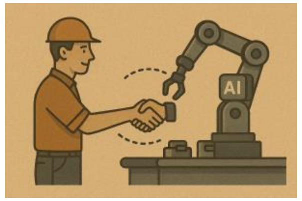
Preparing organisations and individuals for the AI skills of the present and near future



- Know your goals: can we target simultaneously operational, technical, and human-values oriented AI-system design spaces?
- Learn to work with AI and AI agents in your team
- Controllability of AI solutions and human agency
- Sliding autonomy (between humans and non-human agents)
- Safeguards to shield trustworthiness of AI outcomes
- How (not) to trust (non-) trustworthy AI systems and deliver value



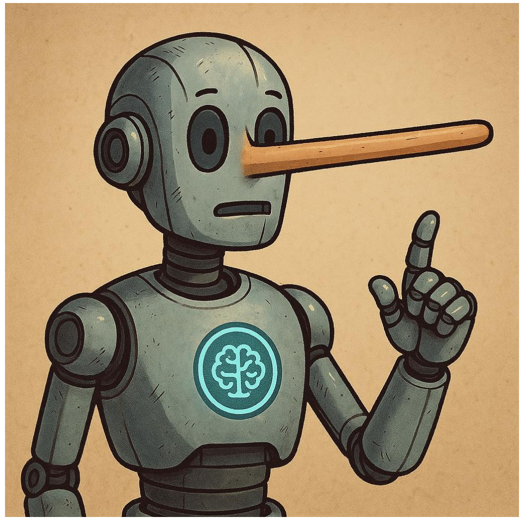




Humans, AI and Trustworthy Maintenance for Dependability

Design, Evaluation, Auditing, and Lifecycle Management for Human-Centric AI and Human-AI Teaming in Industry

- Co-creation design patterns for human–AI teaming
- How to achieve interoperable human-AI teaming
- How to empower people to work on the job with trustworthy AI
- Tripple-pillar evaluation for Human-Centric AI (Operational, Technological, Social)
- Trust, risk, and security management in AI and LLM-based Agentic AI
- Responsible AI reasoning and agentic AI
- Benchmarking Agentic AI
- Smart Industry Readiness (incl AI Readiness)



Humans, AI and Trustworthy Maintenance for Dependability

Maintaining the Future

Christos Emmanouilidis

c.emmanouilidis@rug.nl

university of groningen