

# Pan-EU AI Adopters Ecosystem

## Trustworthy AI in the EU ecosystem



H2020 Innovation Action - This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N. 101017057



# AGENDA

## Pan-EU AI Adopters **Trustworthy AI** in the EU ecosystem

01

### **Welcome & Introduction**

10:00 - 10:10

Thierry Louvet, Europe and international affairs Director at Systematic Paris Region

02

### **What is “Trustworthy AI”?**

10:10 - 10:25

Sara Mancini, AI Ethics’ expert and Senior Manager at Intellera consulting

03

### **Accountable Federated Machine Learning in Bavarian municipalities**

10:25 - 10:40

Tomas Bueno Momcilovic, Scientific Researcher at Fortiss GmbH

04

### **Towards the engineering of TAI applications for critical systems**

10:40 - 10:55

Bertrand Braunschweig; Scientific coordinator of the Confiance.ai programme

05

### **AI for the people and not against them**

10:55 – 11:10

Frank Ortmeier, Managing Director at bridgefield GmbH

06

### **Q&A and final remarks**

11:10 – 11:25



# Welcome and introduction

Thierry Louvet, Europe and international affairs Director at  
Systematic Paris Region



H2020 Innovation Action - This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N. 101017057

# DIH4AI Project Introduction

‘The DIH4AI “AI on-demand platform for regional interoperable Digital Innovation Hubs Network” has clear objectives that rely on three fundamental pillars

## OBJECTIVES



Build a network of AI-on-demand innovation and collaboration platforms, **interoperable with the AIoD platform**



Supporting the **joint development and provision of services** through a sustainable network of regional AI DIHs and targeting local SMEs and GovTech agencies.



## KEY PILLARS



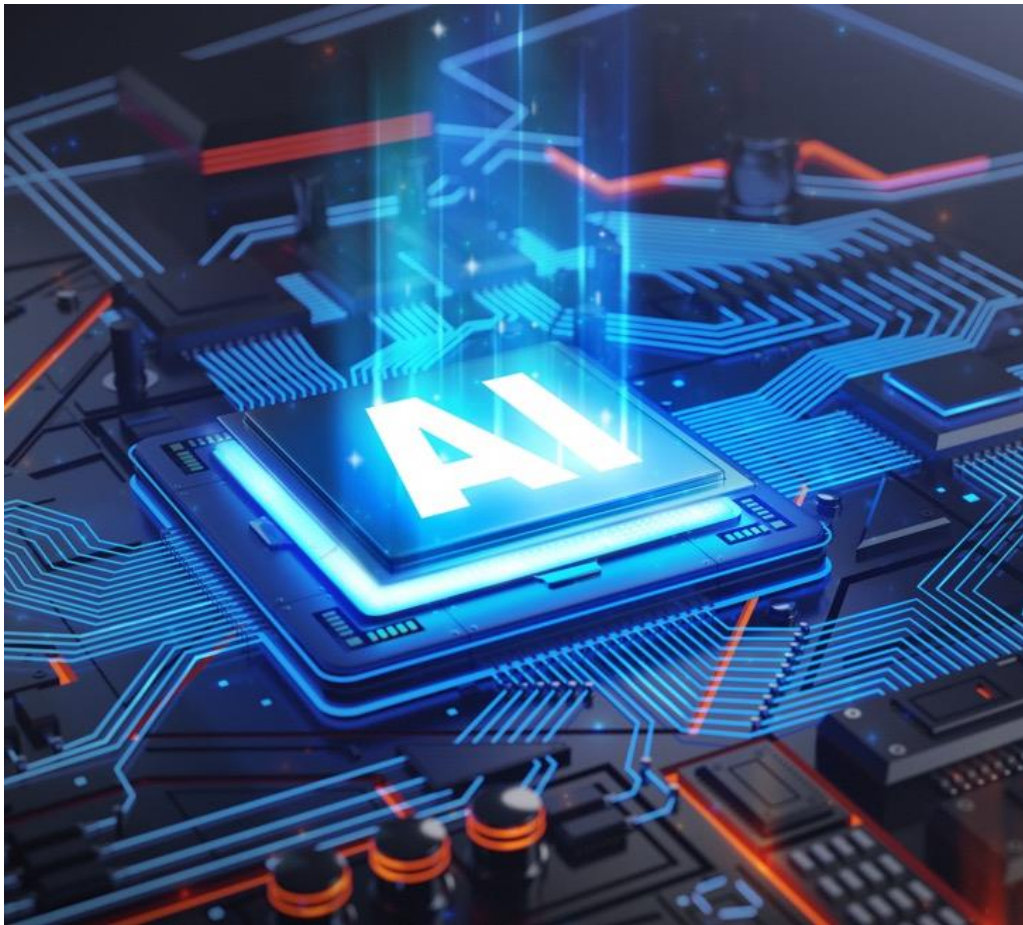
Technological Open Platform for AI DIHs



Regional and European Interoperability Framework



Methodological Framework for DIHs collaboration



# Trustworthy AI for AI adopters

This webinar is organised by **Paris-Region Digihall** and **DIH Saxony-Anhalt** with the support of **Intellera** with the aim of:

- ✓ **Raise awareness** among SMEs about ethics in AI, and help demystify the concept by giving **concrete examples**;
- ✓ **Generate ideas** among participants for potential **future collaboration**, for instance in EU-funded projects relating to AI and ethics;
- ✓ Benefits participants with **new contacts and renewed interest in the topic**.

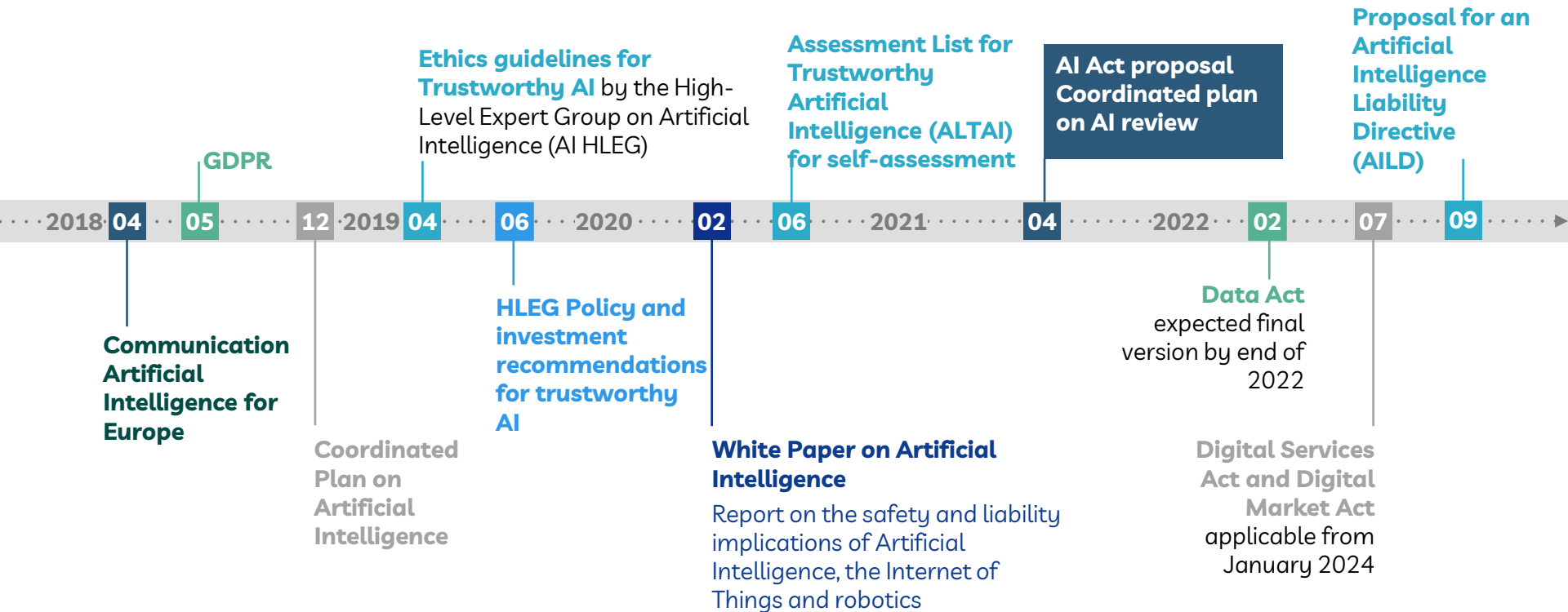


# What is Trustworthy AI?

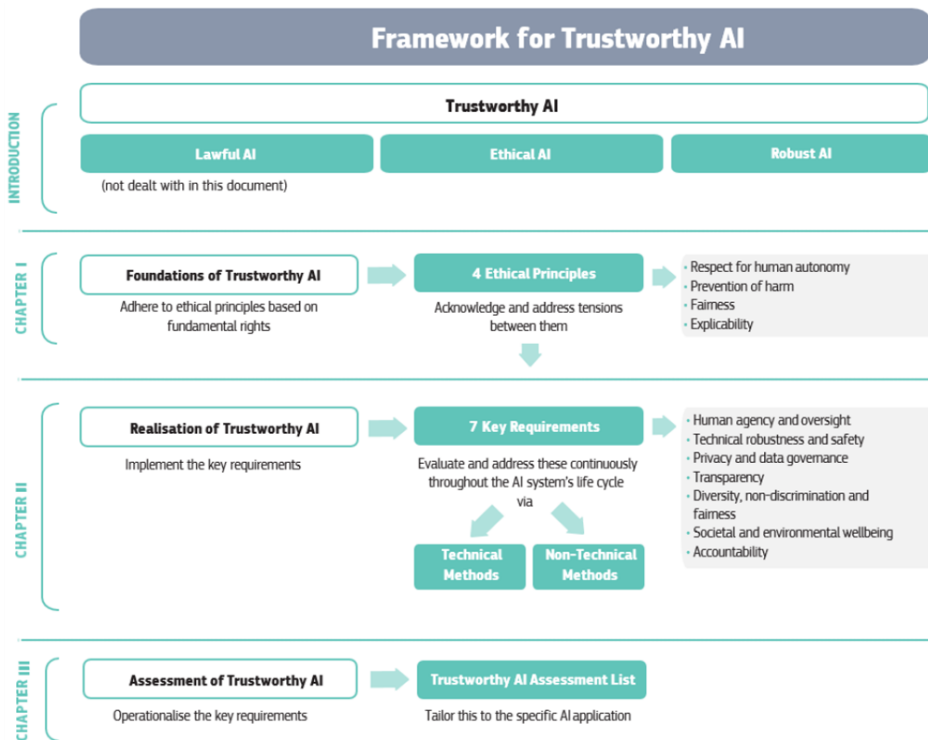
Sara Mancini, AI Ethics expert, Senior Manager, Intellera Consulting



# The European Commission has been very active with respect to AI and its risks, setting the base for the forthcoming regulation



# High-level expert group on Artificial Intelligence: Ethics Guidelines for Trustworthy AI



## Four Ethical Principles

1. Respect for human autonomy
2. Prevention of harm
3. Fairness
4. Explicability

## Seven key requirements

1. Human agency and oversight
2. Technical Robustness and safety
3. Privacy and data governance
4. Transparency
5. Diversity, non-discrimination and fairness
6. Societal and environmental well-being
7. Accountability

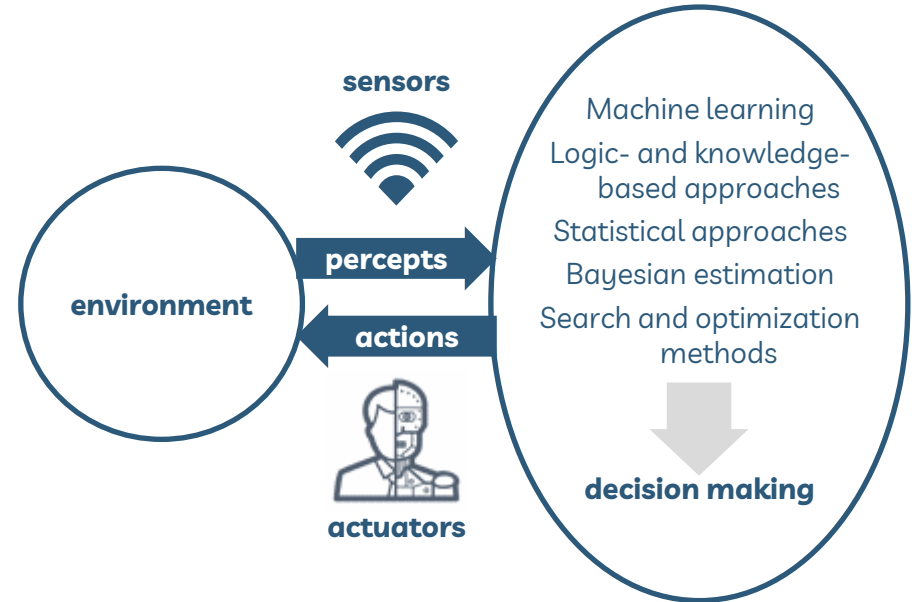


# The AI Act proposed **AI Definition**

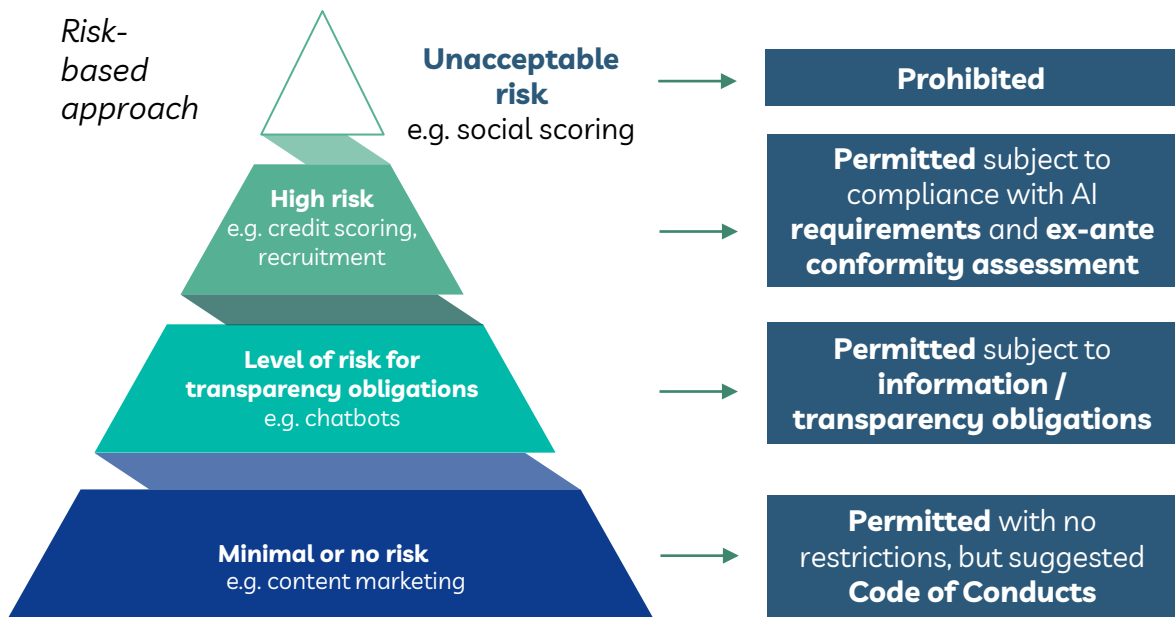
**'Artificial Intelligence system' (AI system)** means software that is developed with one or more of the following techniques and approaches and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.

Techniques and approaches:

- **Machine learning approaches**, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
- **Logic- and knowledge-based approaches**, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
- **Statistical approaches**, Bayesian estimation, search and optimization methods.



# The EC proposed the **first ever legal framework on AI** which will enter into force in a transitional period at the end of 2022, final approval by 2023



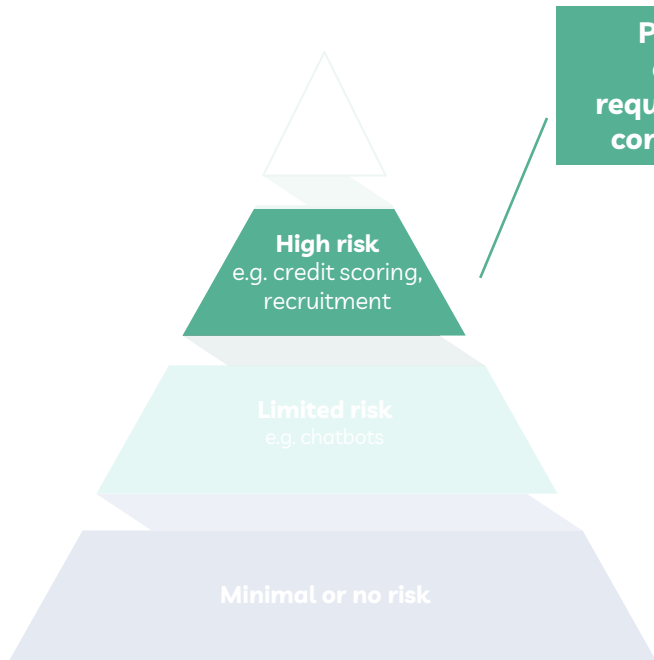
## Who is subject to the AI Act?

It will apply to any **natural or legal person** whose AI services or products reach the EU market:

- **Provider:** any natural or legal person, public authority or other body that develops an AI system, or has an AI system developed
- **User:** any natural or legal person using an AI system under its authority and places that system on the market or puts it into service

Potential **penalties** for non-compliance (from 1 to 6% of annual worldwide turnover) calculated based on the actual infringement (gravity, impact, etc.).

# The EC proposed the **first ever legal framework on AI** which will enter into force in a transitional period at the end of 2022, final approval by 2023



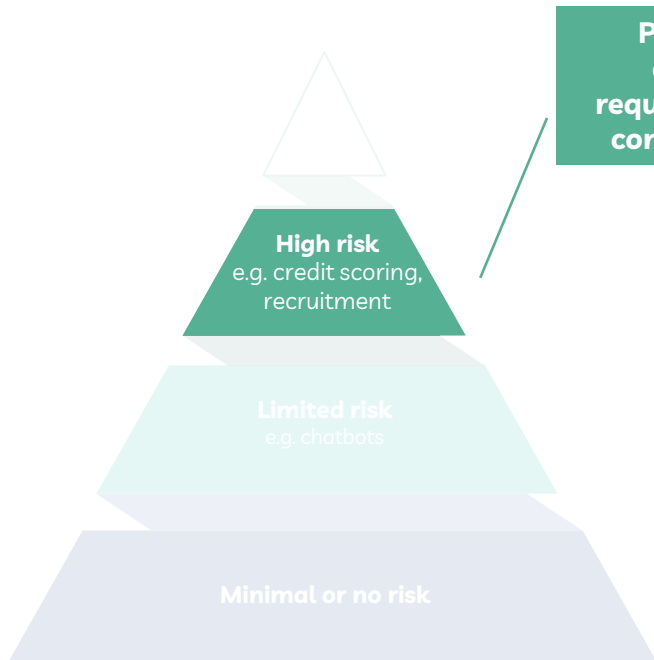
**Permitted** subject to compliance with AI requirements and **ex-ante conformity assessment**

## What AI applications are high risks?

Biometric identification and access to social benefits or credit	Management and operation of critical and digital infrastructure
01	02
03	04
Education; Employment, workers management and access to self-employment	Emergency first response, crime prediction and court use, migration / asylum & border control mgmt

- AI systems used as **safety component of products** (or that are themselves products) that are normally **subject to third-party ex-ante conformity assessment**, covered by the legislation in annex 1:
  - Annex 1 includes all NLF legislation, such as Machinery Regulation, Radio Equipment Directive, Medical Devices Regulation, In-Vitro Diagnostics Regulation
  - Annex 1 also includes sector-specific harmonisation legislation, including cars and aircrafts
- **AI systems with fundamental rights risks**, listed in annex 2 (see picture)

# The EC proposed the **first ever legal framework on AI** which will enter into force in a transitional period at the end of 2022, final approval by 2023



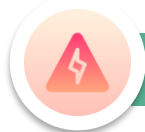
**Permitted** subject to compliance with AI requirements and **ex-ante conformity assessment**

## What AI applications are high risks?

Biometric identification and access to social benefits or credit	Management and operation of critical and digital infrastructure
01	02
03	04
Education; Employment, workers management and access to self-employment	Emergency first response, crime prediction and court use, migration / asylum & border control mgmt

- High-risk mandatory requirements**
1. Establish and implement **risk management** processes
  2. Use high-quality **training, validation and testing data**
  3. Establish **documentation** and design logging features (traceability & auditability)
  - 4. Record keeping**
  5. Ensure appropriate certain degree of **transparency** and provide users with **information**
  6. Ensure **human oversight**
  7. Ensure **robustness, accuracy** and **cybersecurity**

# AI used for credit scoring



**High risk**  
e.g. credit scoring, recruitment



*Permitted subject to compliance with AI requirements and ex-ante conformity assessment*

## **AI-powered credit worthiness assessment**

**Credit scores** can control housing decisions, the cost of taking out a loan, and even employment. The advent of **AI in financial services** poses a unique opportunity **to improve fairness** in the important arena of **credit scoring**, but it can also deepen the impact of bias. Bias can arise across within each of the different phases of the development cycle of an **AI algorithm**.

Bank of Italy\* (2022) conducted a survey and found that among 10 financial intermediaries, both banks and non-banks:

- the use of AI methods in the **assessment of credit risk is not yet widespread but is growing**
- in almost all cases, the **model scores are provided to support the assessment of creditworthiness** by analysts, who are responsible for the final decision. However, some respondents have declared their intention to gradually **reduce human intervention in the lending process in the future**



### **Why it is high risk?**

- The **AI Act Proposal** considers as high-risk AI systems intended to be used to evaluate the **creditworthiness** of natural persons or establish their credit score

### **Critical aspects**

- There's a **risk of biased assessment**. Credit scoring systems can present some biases (such as incorrect risk **differentiation and discrimination against individuals** or social groups) and such biases also tend to generate a dangerous feedback loop in which they can be confirmed and amplified.

# Trustworthy AI within the DIH4AI project



Within the DIH4AI project we have three main outputs/activities where we bring and apply Trustworthy AI aspects: the DIH4AI experiments, the L-BEST taxonomy and the Trustworthy AI Working group



DIH4AI  
experiments

A set of AI experiments developed by Digital Innovation Hubs, that are categorized according to a set of services. Specific experiments present Trustworthy AI aspects.



L-BEST Taxonomy of  
services

The L-BEST taxonomy is a 3-level categorization for AI DIHs services. The objective is to provide to DIHs a standard framework to describe their services.



Trustworthy AI  
Working Group

The definition of Cross-DIHs collaboration scenarios to ensure the joint provision and development services, and the joint matchmaking of complementary competencies

# DIH4AI experiments for **Trustworthy AI services**



## [X-MUC-2] Platform-as-a-service for accountable evidential transactions

Leading DIH: fortiss

Partner: TNO

- ❑ The platform under development (**PIANAI**) is a tool to enable transparency and accountability for the collaborative development of AI solutions. This platform is itself **a tool for trustworthy AI**;
- ❑ On top of that, the working group is discussing the **idea of integrating some AI standards** into the platform in order to ensure that the development of solutions on the PIANAI platform is also compliant with the standard.



*Next presentation*

## [X-PAR-1] Pan-EU AI Adopters Ecosystem

Leading DIH: DIGIHALL

Partner: Fortiss, Fraunhofer IFF, TNO

- ❑ DIGIHALL and DIH Saxony-Anhalt, together with Intellera, will organize **workshops** dedicated to the respective SMEs' networks;
- ❑ One of the main objectives of the workshops will be to raise awareness on the topic of Trustworthy AI, and the impact of possible new regulation or standards for AI.



*Today's Webinar*

## [I-PAR-2] AI Ethical Assessment

Leading DIH: DIGIHALL

Partner: CEA+ Systematic

- ❑ DIGIHALL and Intellera will support an **assessment of social and ethical issues** of an AI application of a SME from DIGIHALL network based on the outcome and assets of the ETAPAS project;
- ❑ In particular, the **ETAPAS RDT Framework** and its methodology will be used to assess the most important ethical principles, social and ethical risks associated with the adoption of such AI application;
- ❑ In addition, the Framework will be used to identify mitigation indicators to reduce the identified risks and monitor the adoption.

# L-BEST service portfolio – Level 2

## L

### Legal

- LEGAL AND IPR ASSISTANCE
- ETHICAL AI ORGANISATIONAL SUPPORT
- ETHICAL AI LIFE CYCLE ASSISTANCE & ASSESSMENT

## B

### Business

- INCUBATION ACCELERATION SUPPORT
- ACCESS TO FINANCE
- OFFERING HOUSING
- BUSINESS TRAINING AND EDUCATION
- PROJECT DEVELOPMENT

## E

### Ecosystem

- COMMUNITY BUILDING
- DIH INNOVATION DEVELOPMENT
- ECOSYSTEM GOVERNANCE

## S

### Skills

- PROCESS & ORGANIZATIONAL MATURITY
- HUMAN CAPABILITY MATURITY
- SKILLS IMPROVEMENT

## T

### Technology

- IDEAS MANAGEMENT & MATERIALIZATION
- CONTRACT RESEARCH
- PROVISION OF INFRASTRUCTURE
- TECHNICAL SUPPORT ON SCALE UP
- VERIFICATION AND VALIDATION
- DATA MANAGEMENT



# ICT49 Trustworthy AI Working Group



With a view to **strengthen synergies across ICT-49 projects and enhance the EU AI on demand platform DIH4AI** initiated a collaboration initiative on **Trustworthy AI** involving all the six ICT49 projects.

DIH4AI newly developed the L(egal-ethical) category of services of the L-BEST taxonomy and is **validating and improving it with external stakeholders** as part of the activity of the **ICT49 TAI WG**

## Legal and ethical services taxonomy

### LEGAL AND IPR ASSISTANCE

- Legal advice and support
- IPR assistance & management
- IPR management tool
- Model agreements & assistance
- Regulatory Sandboxes

### ETHICAL AI ORGANISATIONAL SUPPORT

- Support definition of internal AI Code of Conduct
- Ethics-related organizational measures
- Training on Ethical & Legal AI
- Ethics Expert on-demand
- Technical support and tooling for ethical AI

### ETHICAL AI LIFE CYCLE ASSISTANCE & ASSESSMENT

- Ethical AI Committee as a Service
- Ethical risk assessment
- Support the development of ethically-aware AI solutions
- Conformity assessment / certification of AI solutions
- AI solution independent audit

## Key objective 1

- Sharing across ICT-49 projects a **common approach** (L-BEST) ✓
- **Revision and extension of the L services' catalogue** ✓
- **Mapping of the L services** ✓
- Definition of a **strategy to integrate it with the AloD Platform** 🔄

# ICT49 Common Approach to TAI assessment



ICT 49

The TAI WG is drafting a **common methodology document, based on ALTAI** and aiming at integrating it with further **needed guidelines and recommendations** for the trustworthy assessment of AI applications. The methodology document will be **validated with external stakeholders**.

## Key objective 2

- Perform a **Literature Review** on practical use of ALTAI
- Collect **direct feedback** on ALTAI from ALTAI-users among **ICT49 project Consortia** to catch **needs and requirements**
- Define a **draft common methodology**, based on ALTAI, for assessing the trustworthiness of AI applications
- **Validate the methodology with DIHs and SMEs** engaged through our projects, as well as with the European Digital SMEs Alliance, the pool of experts of the network of excellence, ICT48 and other projects
- **Define a strategy to integrate it into the AIoD platform**
- **Lesson learned and comments on the ALTAI** could be provided to the European Commission by the working group as a result of this activity



### Integrating ALTAI by:



- providing concrete guidelines with examples
- providing specific feedback in areas where improvement is needed due to low scores
- integrating examples and guidelines on the specific sectors and types of AI technologies covered by the ICT49 projects
- Identifying which indicators are relevant to each development stage

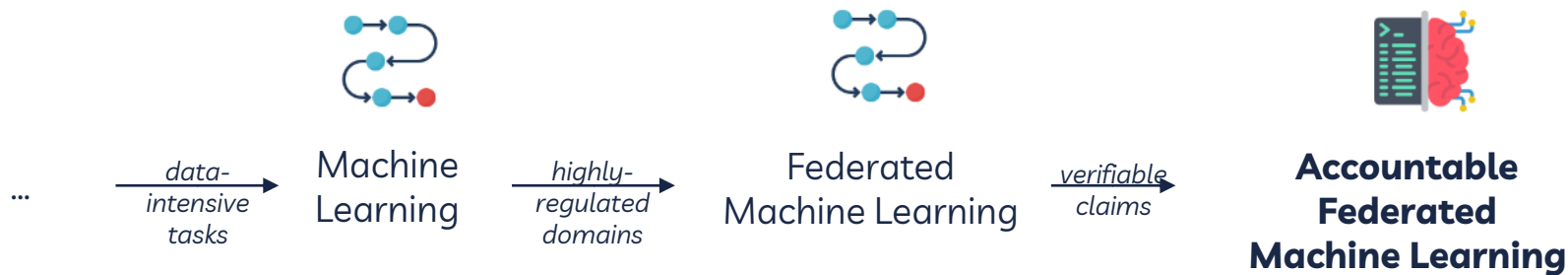


# Accountable Federated Machine Learning in Bavarian municipalities

Tomas Bueno Momcilovic, Scientific Researcher, Fortiss GmbH



# Defining the **learning landscape**

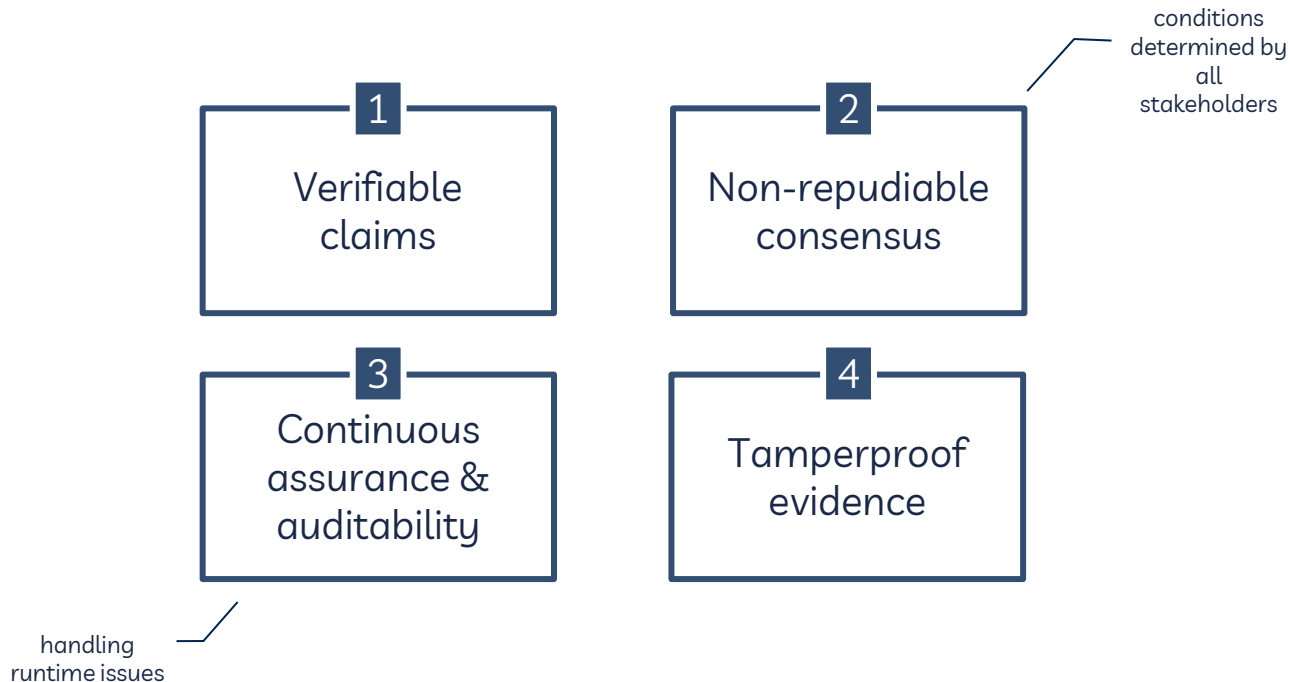


# The need for **accountability**

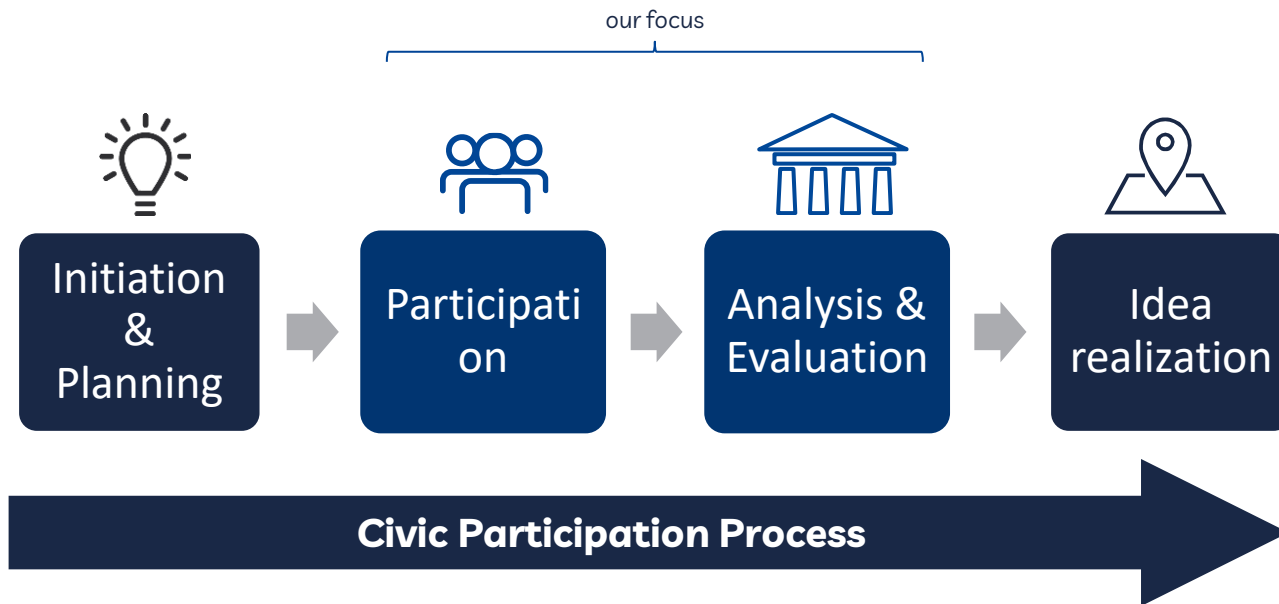
- ▶ ML models might not be trusted
  - How can we introduce measures of **trustworthiness**?
  - Bias, adversarial attacks, explainability, etc.
- ▶ Insufficient data & data sharing not feasible
  - How to develop models under **compliance** restrictions for data sharing? (e.g., GDPR)
  - „Data lakes“ practically non-existent, privacy & security concerns, etc.
- ▶ Data & model provenance in a distributed setup
  - How to account for ML development & **operation** process?
  - Factsheets, continuous integration of data & models, etc.



# Our view of accountability



# The **civic participation** case



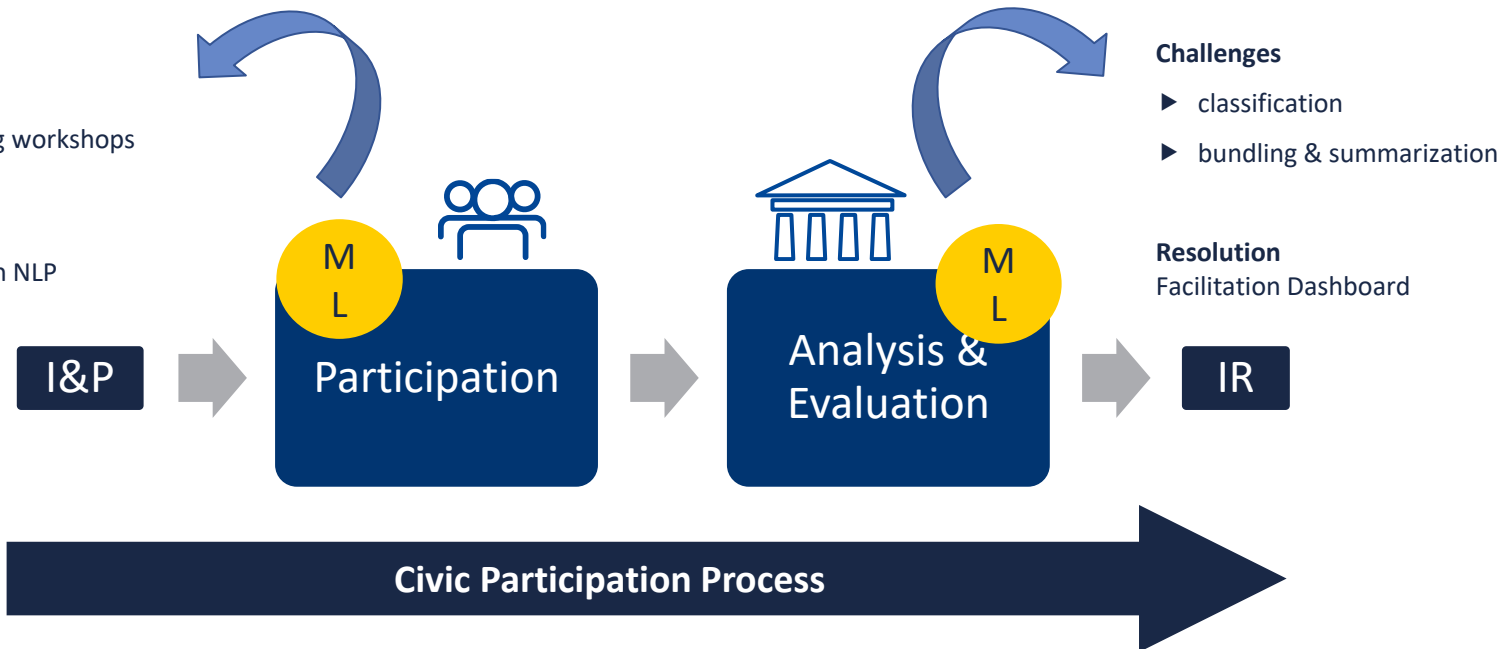
# ML challenges and resolution

## Challenges

- ▶ idea quality
- ▶ time consuming workshops

## Resolution

Idea Facilitator with NLP





# Overcoming ML challenges



## Tasks

- Involved parties
- Purpose definition
- Data assessment
- Legal assessment for data processing

## Examples

- Categorize user feedback for an efficient user response
- Private data exists in the data set (name, e-mail, street address, ..)

## Tasks

- Data cleaning/filtering
- Feature extraction
- Vectorizing/transformation
- Agree on a test dataset for the training
- {anonymization techniques}

## Examples

- Remove too short text
- Column naming in Excel
- Use dictionary X for character mapping

## Tasks

- Anonymization techniques
- Model optimization
- Fine-Tuning

## Examples

- Hyper-parameters
- Optimisation algorithm
- Model architecture
- ...

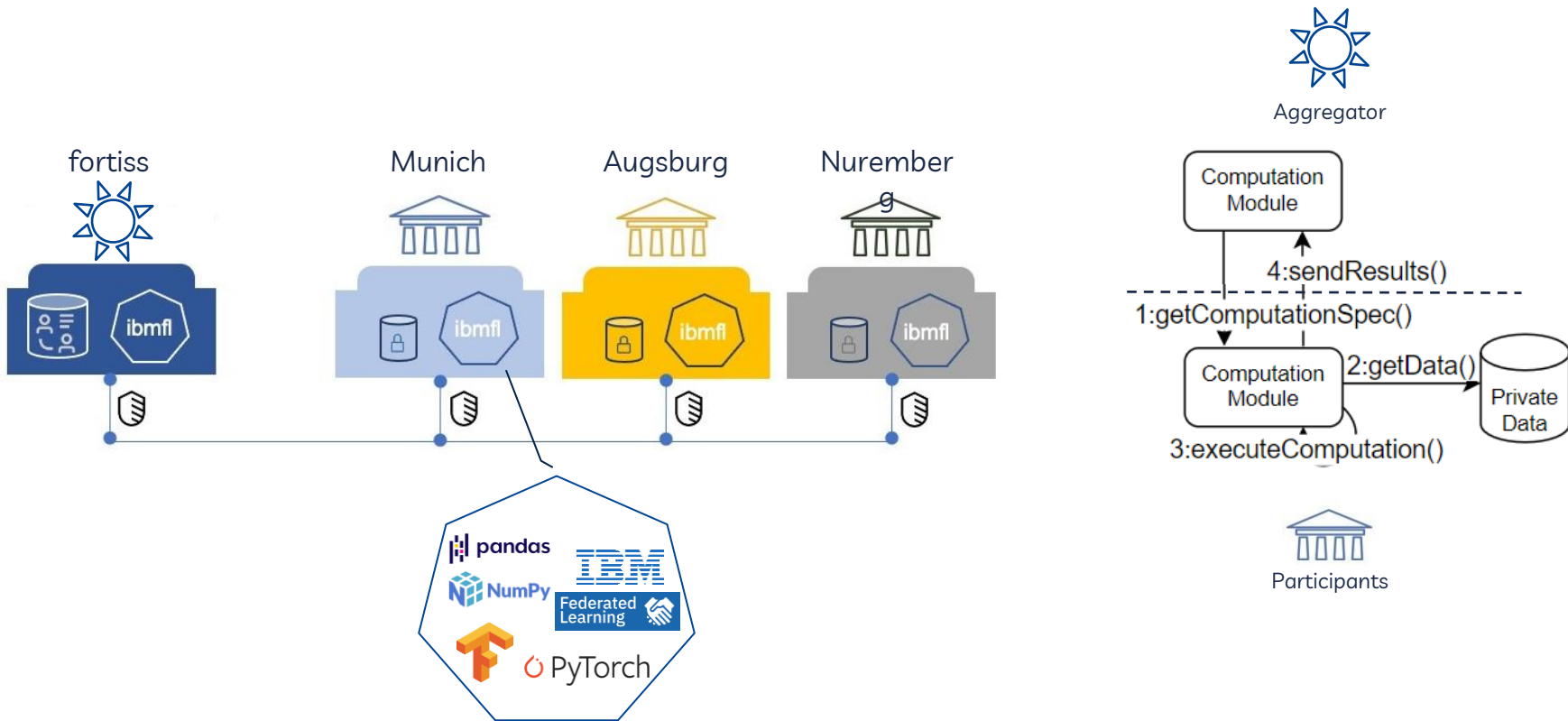
## Tasks

- Anonymization techniques
- Model aggregation
- Robustness check
- Bias check
- Metrics computation

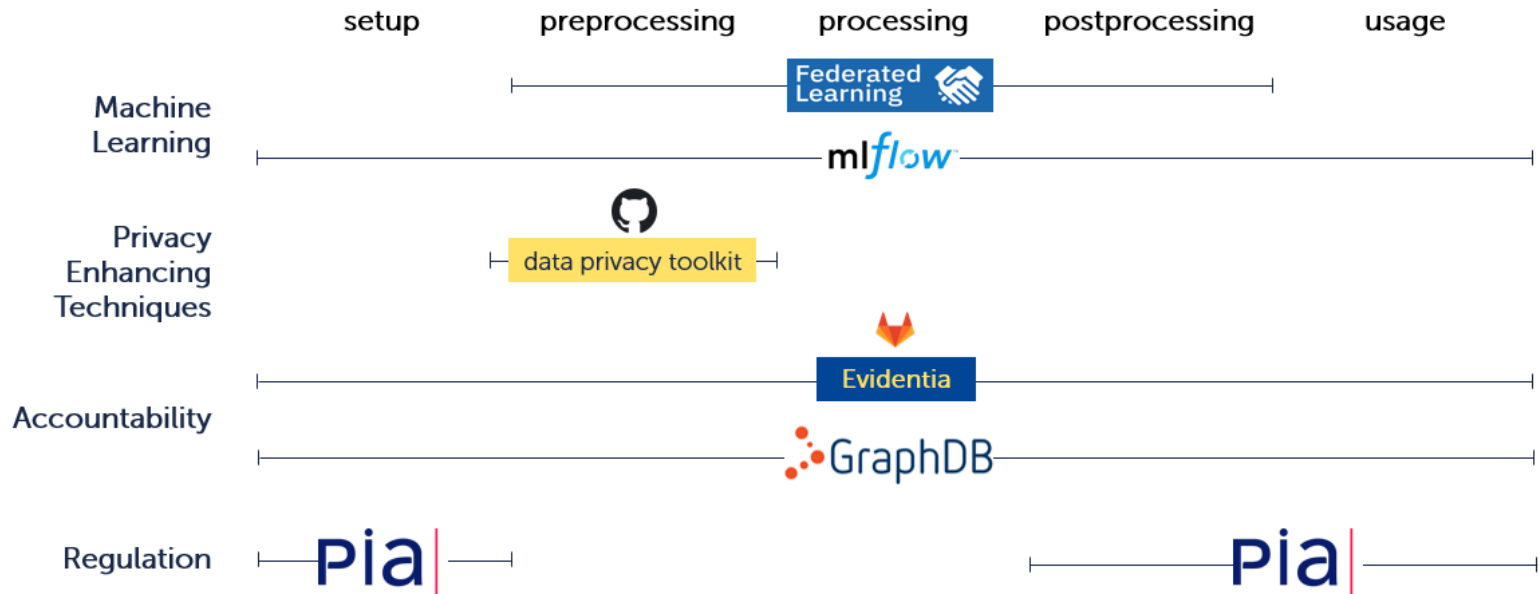
## Examples

- Accuracy/F1-score
- Differential privacy applied to the model
- ...

# Use case in **Bavaria**

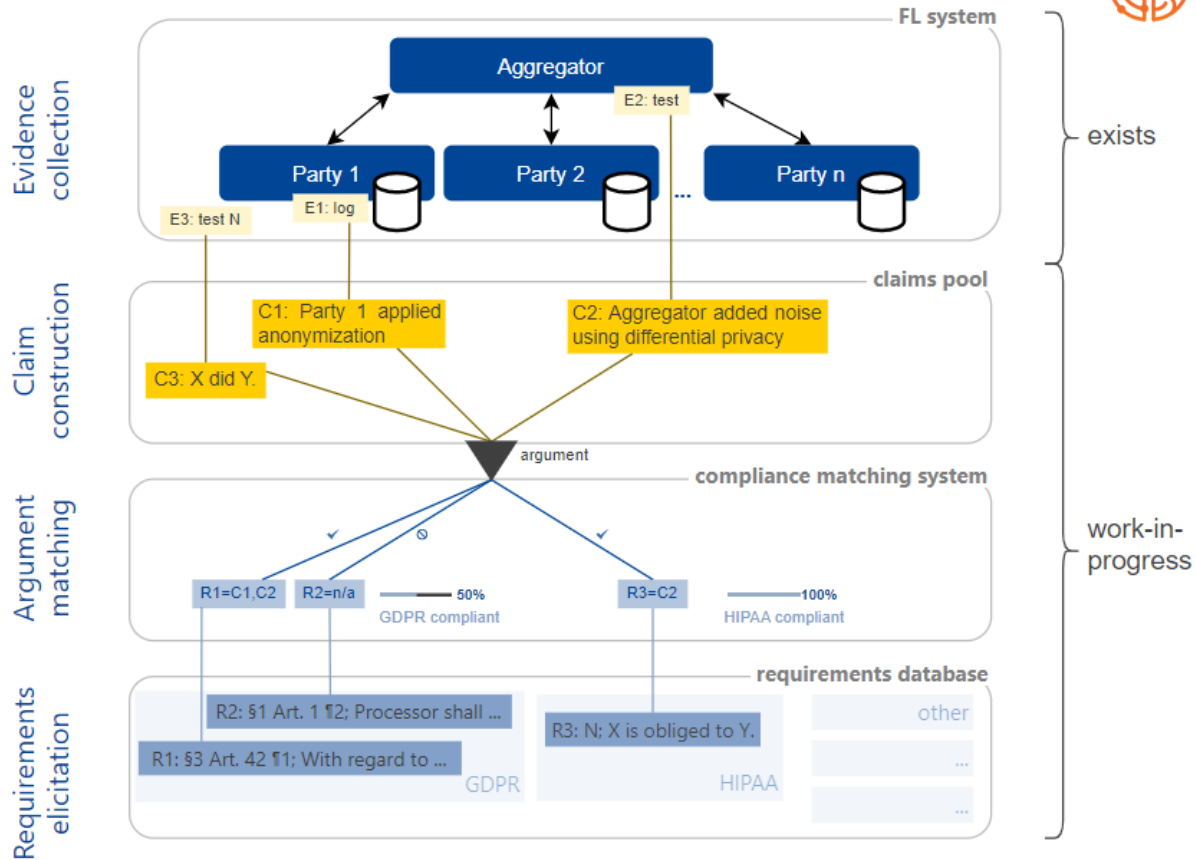


# Architecture overview



# Next steps

- ▶ Assurance of regulatory compliance in AFML
- ▶ Improving auditability in a multifaceted approach



# Contact

---

fortiss GmbH  
Guerickestraße 25  
80805 Munich  
GERMANY  
[www.fortiss.org](http://www.fortiss.org)  
[info@fortiss.org](mailto:info@fortiss.org)



# Confiance



AIRBUS

 Air Liquide

Atos



Inria

NAVAL  
GROUP

Renault  
Group

 SAFRAN



sopra  steria

SystemX

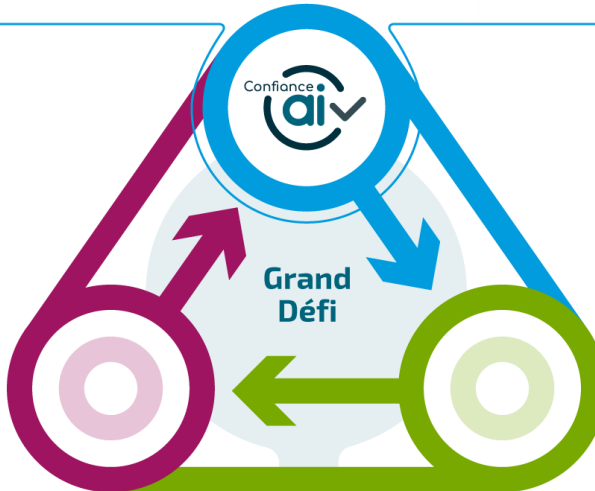
THALES  
Building a better world of flight

Valeo

# Making France one of the leading countries in artificial intelligence for industry

**NORMS PILLAR**  
Norm, standard  
and regulation  
environment toward  
certification

**afnor**  
GROUPE



*Securing, certifying and enhancing  
the reliability of AI-based systems*

**SystemX**  
INSTITUTE OF RESEARCH  
AND TECHNOLOGY

## **TECHNOLOGICAL PILLAR**

Design, deploy and maintain  
AI-based critical systems

## **APPLICATION EVALUATION PILLAR**

Ensure the right  
operational exploitation

- Industry strongly involved in programs, especially AI Manifesto members.
- Cooperation with French basic research Initiatives, such as Aniti or DataIA, and academic research.



**A unique French community  
to design and industrialise  
thrustworthy AI-based critical systems**



# The three faces of trustworthy AI

## TECHNOLOGY



- Reliability, robustness
- Lawfulness & Compliance
- Accuracy
- Security
- Safety

## INTERACTIONS



- Transparency
- Explainability
- Accountability
- Oversight & control


## ETHICS



- Fairness
- Privacy
- Diversity & inclusion
- Sustainability

A close-up, blue-tinted photograph of a precision industrial machine, likely a CNC lathe, with a cutting tool positioned over a workpiece. The lighting is dramatic, highlighting the metallic surfaces and the intricate details of the machinery.

## AI Engineering: a national strategy at the service of French industry

A network diagram consisting of numerous small blue dots connected by thin, light blue lines, forming a complex web of connections. The dots and lines are arranged in a way that suggests a large-scale, interconnected system or data network.

Provide industrial companies  
with solutions that enable them to  
develop new critical products and  
services based on trustworthy AI

# Key figures



**4**

YEARS

Duration



**9**

Large industrial groups



**4**

Research centers



**7**

Thematic projects



**+300**

FTP involved over 4 years



**2**

Sites: Paris-Saclay and Toulouse



**45**

M€

Budget



**43**

Associate partners (laboratories, SMIs, startups)

# A unique French community

A group of major French academic and industrial players pooling their cutting-edge scientific and technological skills



Defence



Transport



Manufacturing industry



Energy

AIRBUS

Air Liquide

Atos

cea

Inria

NAVAL GROUP

Renault Group

SAFRAN

SAINT EXUPÉRY

sopra steria

SystemX

THALES

Valeo

# Program structuring

FEDERATIVE ENVIRONMENT, METHODS, TOOLS AND USE CASES

Open / Interoperable / Maintained

Data  
and knowledge

Design

Evaluation

Embedded  
systems

QUALITY ASSURANCE – ENGINEERING

# Focus on the scientific challenges

## TRUST AND SYSTEMS ENGINEERING WITH AI COMPONENTS



- Global Approach to Trustworthy AI Components
- Building Trustworthy AI Components
- Qualifying AI-based components and systems
- Embeddability of AI

## TRUST AND HUMAN INTERACTION



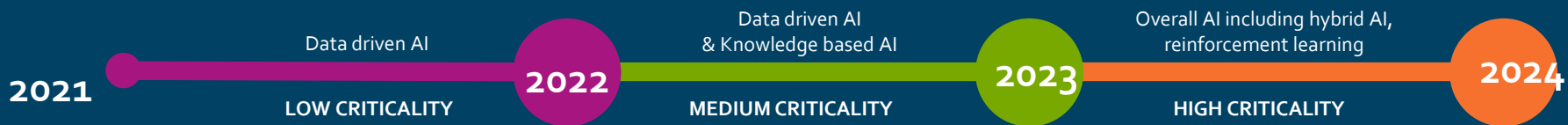
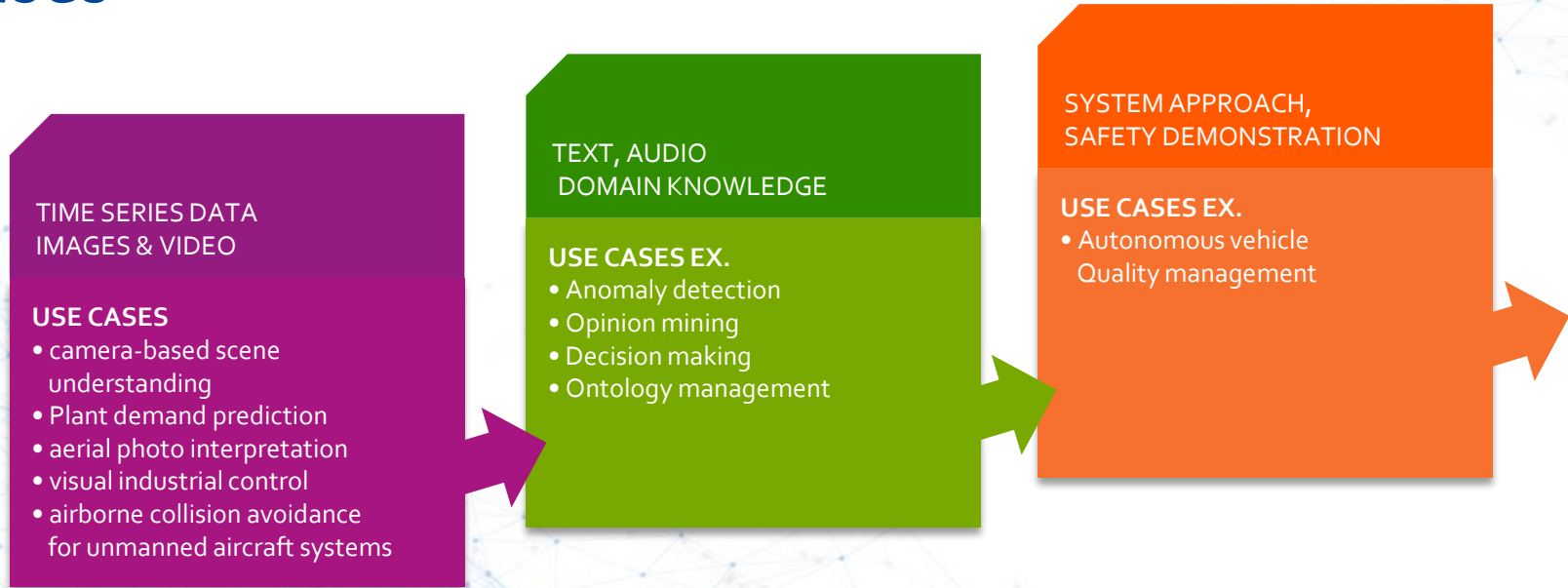
- User interaction - AI-based system
- Designer/certifier interaction - AI-based system

## TRUST AND LEARNING DATA



- Global approach to data/knowledge for learning
- Building data/knowledge for learning
- Qualifying data/knowledge for learning

# An incremental roadmap validated by industrial use-cases



Annual batch delivery of integrated technological and methodological components to be transferred to the consortium' industrial partners

# 43 Partners joining forces

## LARGE GROUPS



## RESEARCH ORGANIZATIONS/UNIVERSITIES



## STARTUP / SME



## STANDARD S<sub>1</sub> NORMES



## ECOSYSTEM



## SUPPORTS







[www.confiance.ai](http://www.confiance.ai)



AIRBUS

Air Liquide

Atos



Inria

NAVAL GROUP

Renault Group

SAFRAN



sopra steria

SystemX

THALES

Valeo



---

**CHALLENGE WHAT'S POSSIBLE**

A portrait of Prof. Dr. Frank Ortmeier, a man with dark hair and glasses, wearing a dark suit jacket over a white shirt. He is smiling slightly and looking directly at the camera. The background is a plain, light-colored wall.

**YOUR SPEAKER**  
Prof. Dr. Frank Ortmeier



**Artificial intelligence for  
humans – and not against  
them**



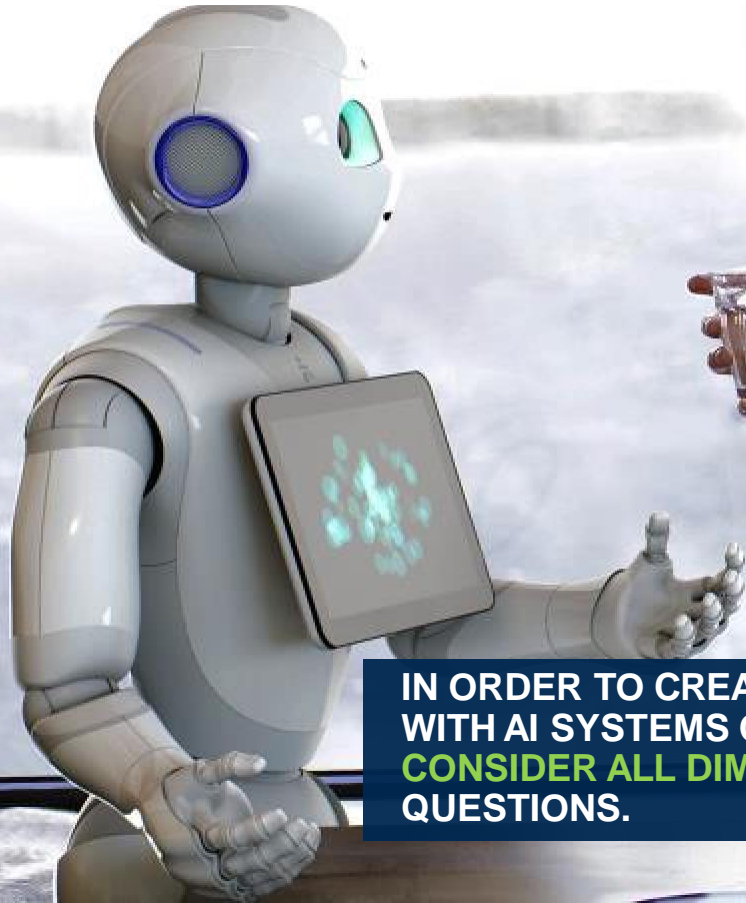
ARTIFICIAL INTELLIGENCE HAS THE POTENTIAL TO ENHANCE  
OUR **DAILY LIFE** AND **WORK**.



AT THE SAME TIME, INTRODUCTION OF ARTIFICIAL INTELLIGENCE BRINGS **MANY** -  
NOT ONLY TECHNICALLY SOLVABLE **CHALLENGES**.



**HUMAN-CENTERED AI-DESIGN HAS TWO FUNDAMENTAL DIMENSIONS:  
DESIGN OF THE AI AND INTEGRATION IN REAL-WORLD PROCESSES**



IN ORDER TO CREATE SOLUTIONS IN WHICH PEOPLE **LIKE TO INTERACT** WITH AI SYSTEMS OR WORK TOGETHER EFFICIENTLY, WE SHOULD **CONSIDER ALL DIMENSIONS** AND CRITERIA WITH THEIR SPECIFIC QUESTIONS.



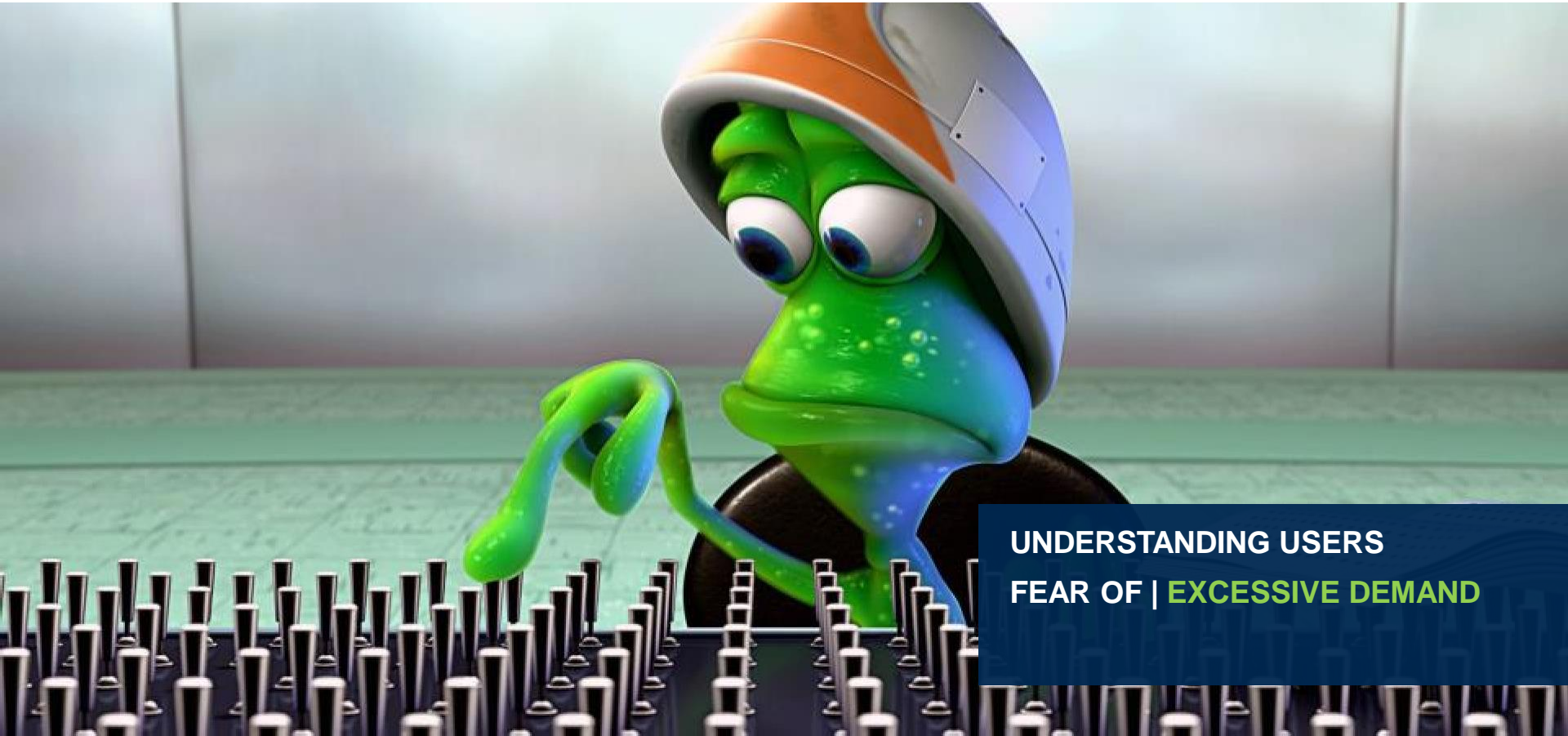
# ASSUMPTION BASED YOU HAVE BECOME. THE DARK SIDE

IT IS THEREFORE ABOUT UNDERSTANDING **ALL USERS** - I.E. THEIR **BACKGROUNDS, FEARS AND WISHES** - AND THUS INVOLVING THEM AT AN **EARLY STAGE IN DEVELOPMENT OR DECISION-MAKING PROCESSES.**

**FEARS**



**UNDERSTANDING USERS**  
**FEAR OF | USELESSNESS**



UNDERSTANDING USERS  
FEAR OF | EXCESSIVE DEMAND



UNDERSTANDING USERS  
FEAR OF | **SURVEILLANCE**

**DESIRES**



**UNDERSTANDING USERS**  
**DESIRE FOR | SELF-DETERMINATION**

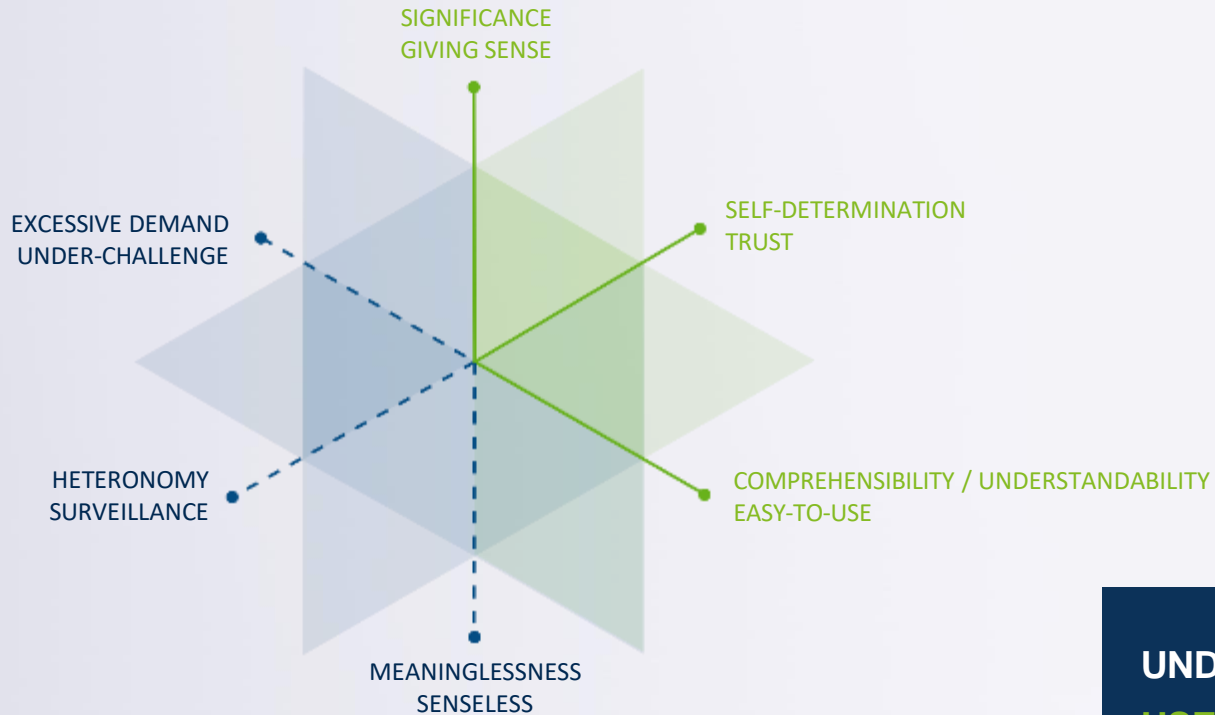


**UNDERSTANDING USERS**  
**DESIRE FOR | UNDERSTANDABILITY**





UNDERSTANDING USERS  
DESIRE FOR | **SIGNIFICANCE**



**UNDERSTANDING USER**  
**USER-ACCEPTANCE-CUBE**



SO WHAT CAN WE DO?

ESTABLISHING ERROR CULTURE FOR  
HUMANS AND AI

BINDING DEFINITION  
OF ROLES

SUPPORTING  
USERS/END USERS

INCREMENTAL  
RESPONSIBILITY-  
TRANSFER

SIGNIFICANCE &  
EFFECT  
VISUALIZE

INTERACTIVE  
INFORMATION-  
EXPLORATION

USAGE

ORGANISATION

TECHNOLOGY

REGULATING AI

EXPLAINING AI

CREATING  
DATA SOVEREIGNTY

CREATING AI  
TOGETHER

INCREMENTAL, ITERATIVE  
INTRODUCTION OF AI

REALTIME  
SCENARIO-  
EXPLORATION

**BRIDGEFIELD'S**  
**HCAI TOOLBOX**



**EXAMPLE: CAST ASSIST**  
**INTERACTIVE INFORMATION EXPLORATION**



# A CONCLUSION

- ✓ SIGNIFICANCE & MEANINGFUL
- ✓ SELF-DETERMINATION & TRUST
- ✓ UNDERSTANDABILITY & TRACEABILITY



**TOP THREE ACCEPTANCE  
CRITERIA FOR AI**



**THANK YOU VERY MUCH!**



# Questions & Answers



# THANKS

Follow the project updates

<https://www.dih4ai.eu/>

