CERTAIN

Learn more about CERTAIN

Towards a Framework for Supporting the Ethical and Regulatory Certification of AI Systems

Co-funded by the European Union

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

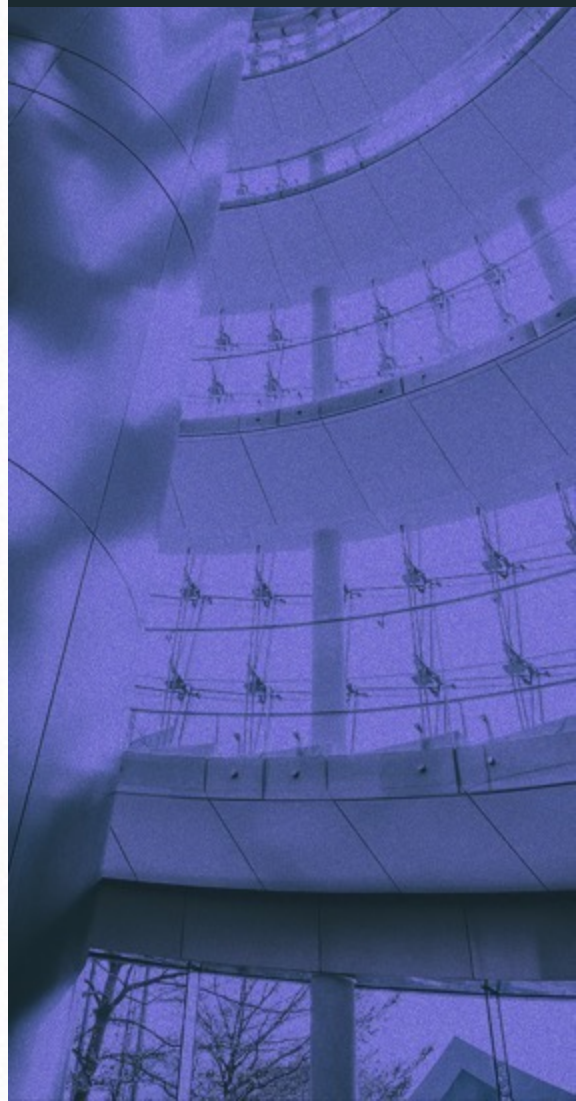
The CERTAIN project received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No 101189650.
This work has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI).

Context and High level presentation

Context

AI, Data, and European Regulation

- The European economy is increasingly driven by **data and AI technologies**
- Multiple stakeholders interact along the **AI value chain**:
  - Data holders
  - Data spaces
  - AI system providers
  - AI system deployers
- The EU has introduced major regulations:
  - AI Act
  - GDPR
  - Data Act
  - Data Governance Act
  - And the "omnibus"...
- Organizations are currently caught between rapid AI adoption and a complex regulatory 'jungle.' CERTAIN closes this gap by turning legal complexity into automated, affordable workflows



CERTAIN: key figures

- **Project Type:** Innovation Action (IA)
- **Programme:** Horizon Europe
- **Topic**

HORIZON-CL4-2024-DATA-01-01 - AI-driven data operations and compliance technologies (AI, data and robotics partnership) (IA)

- **Duration:**
  2025 – 2027 (36 months)
- **Funded under**
  Digital, Industry and Space
- **Consortium:**
  Multi-partner collaboration across Europe – 20 partners from 10 countries
- **Focus Areas:**
  - AI certification
  - Regulatory compliance
  - Data governance
  - Sustainable AI
- **Validation:**
  7 pilots across 6 business sectors

Mission

What is CERTAIN aiming to achieve?

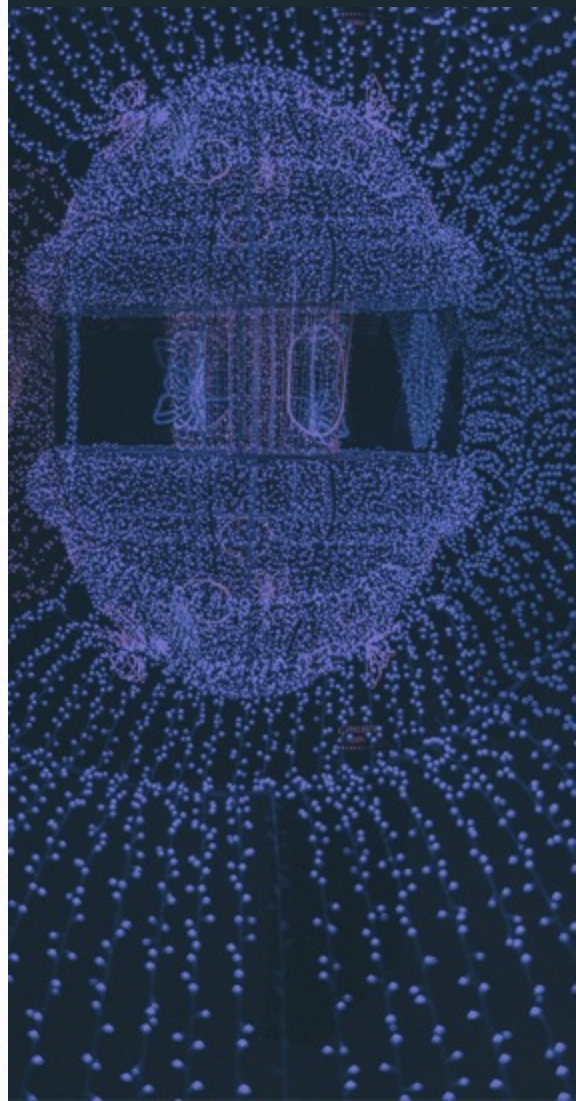
- Help organisations to navigate complex regulatory landscapes
- Provide them with tools and guidelines to ensure compliance with the regulation
- Encourage data holders to share their data with clear guidelines and explaining the benefit of sharing their data
- Promote responsible AI development and helps with its development

OBJECTIVES

1. **Traceability:**
   Enable traceability of critical information of AI systems
2. **Guidelines:**
   Produce guidelines for legally and ethically compliant AI system assessment regarding EU regulations
3. **Compliance tools:**
   Design tools for dataspace providers and data holders to help them to be compliant with EU regulations related to AI and minimise energy consumption
4. **Assessment:**
   Develop methods to improve and assess the compliance of AI systems with EU regulations related to AI
5. **Trustworthiness Testing:**
   Develop a framework for testing AI systems: protocols and competence of 3rd party labs
6. **Evidence:**
   Empirical evidence of the applicability and adequacy of the proposed framework across multiple sectors
7. **Community:**
   To enable the development of an open, dynamic, multi-disciplinary and sustainable community around the EU AI ecosystems and liaised initiatives and actions, towards an EU regulation compliance frameworks

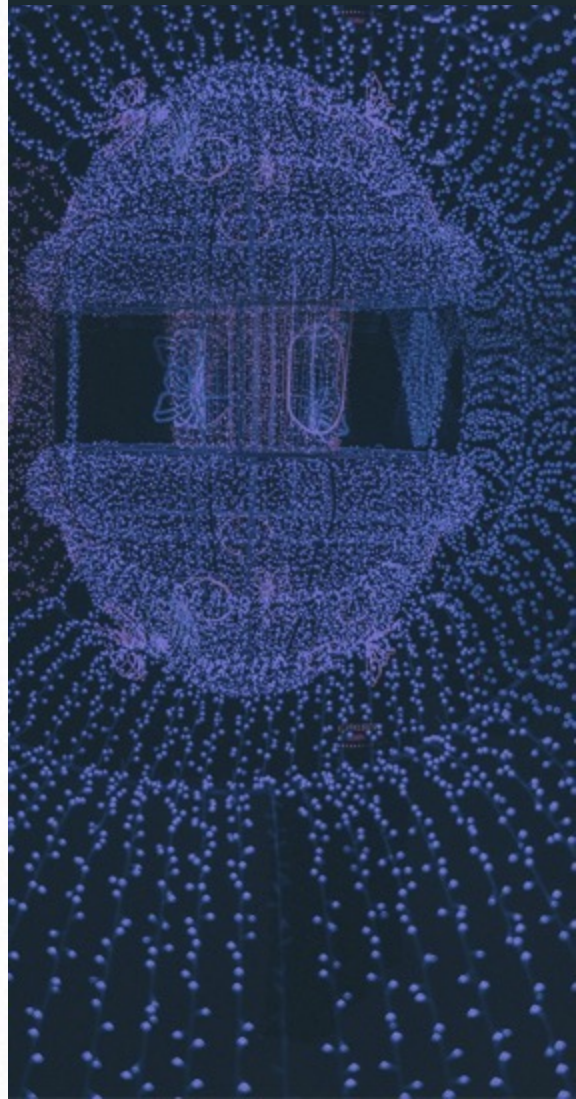
Key outcomes (1/2)

- Comprehensive legal, ethical and societal analysis
- Guidelines dedicated to data holders to help them entering the data market
- Guidelines for AI system providers and AI system deployers detailing AI regulation compliance essentials
- Training and communication to share CERTAIN results with CERTAIN stakeholders and enable exploitation
  - Data holders
  - Data space providers
  - AI system providers
  - AI system deployers
  - Compliance assessment bodies
  - Independent test laboratories
  - Data intermediation services
- Methods and tools to develop Energy-efficient AI during learning and inference
- Methods and tools to develop Privacy-preserving AI



Key outcomes (2/2, technical)

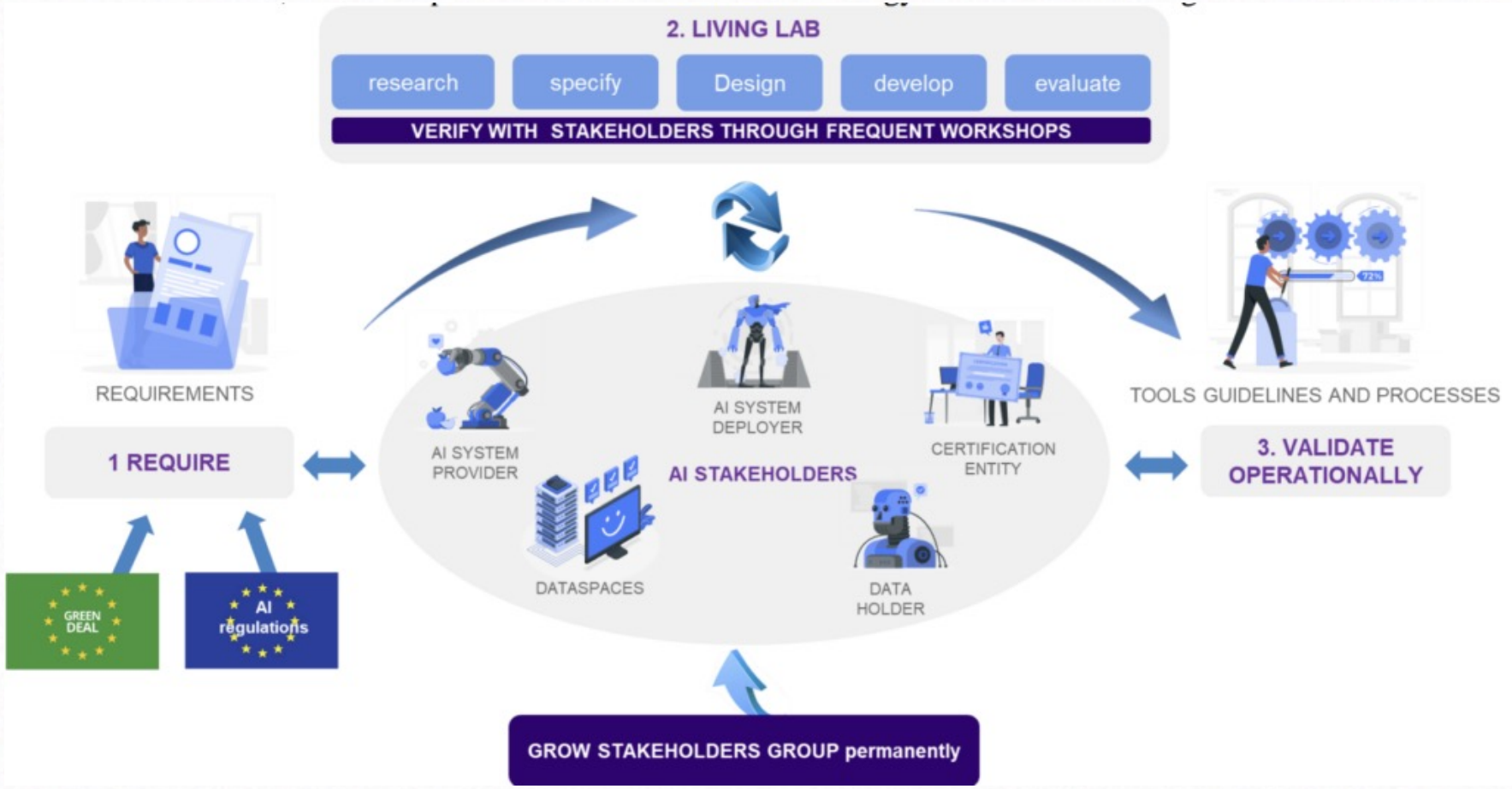
- Dataspace:
  - Identity and access management, and other core services
  - Configurable RegOps workflow
  - Standard compliant connectors to "plug" multiple tools for data holders (e.g. used to analyse data and assess if it complies certain rules)
  - Optimized to minimize energy
- Semantic MLops engine:
  - Add on to MLops pipeline connected to the dataspace to gather data along the lifecycle of an AI system
  - Based on ontologies capturing the whole life cycle of AI systems
  - Extended to document auditability and traceability features required per the AI act
  - Supports cybersecurity risk mitigation through data usage tracking
- Tools for AI system testing
  - Synthetic data generation for testing
  - Testing tools and test result dashboard
  - Framework to enable independent trustworthiness testing by third parties (templates to develop test protocols, qualification of independent testing 3rd parties)

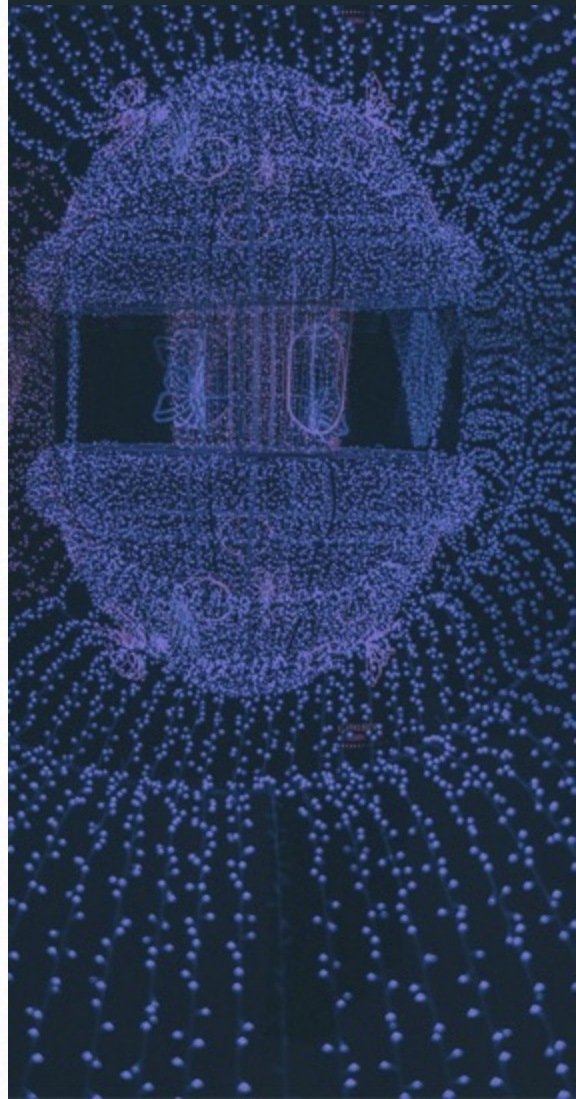


A methodology based on pilots

Methodology

A continuous feedback loop where researchers, providers, and regulators co-create and validate tools in real-time





Methodology

1. REQUIREMENTS: Setting the foundation

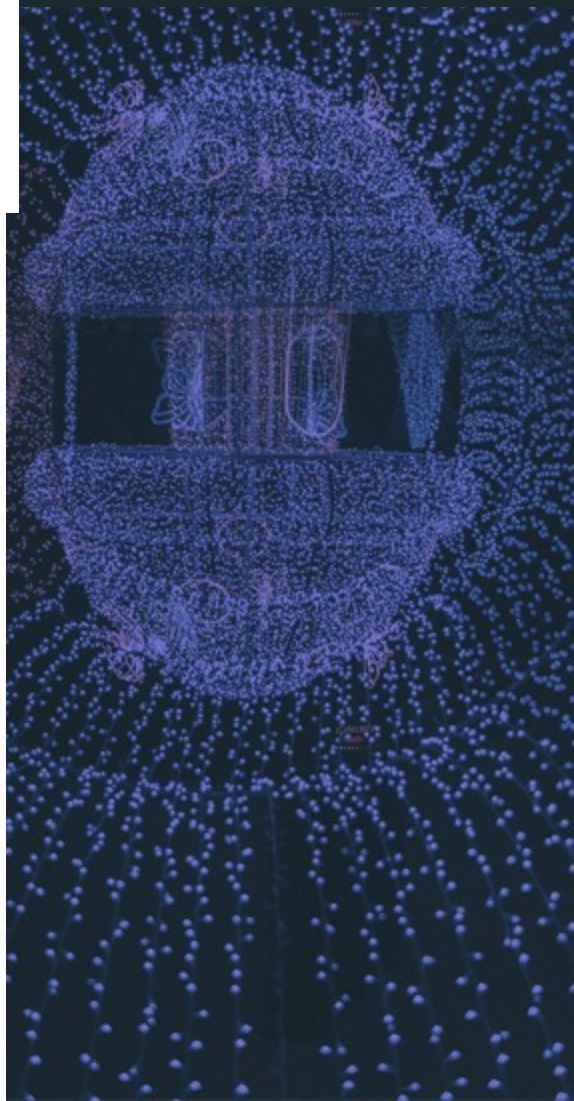
- The process begins by ingestive external drivers, specifically the **European Green Deal** and **AI Regulations**. CERTAIN translates these complex legal frameworks into a clear set of **requirements**, ensuring that the project's starting point is firmly rooted in EU compliance and sustainability goals.

2. THE LIVING LAB: Co-creation & continuous verification

  - At the heart of the methodology is a **Living Lab** that acts as an iterative engine for innovation. This phase follows a structured cycle:
  - **Research & Specify:** Defining the technical boundaries of the problem.
  - **Design & Develop:** Creating the actual tools and frameworks.
  - **Evaluate:** Testing the outputs against the initial requirements.
  - **Stakeholder Verification:** Crucially, this cycle is constantly verified through **frequent workshops** with a diverse group of **AI Stakeholders**, including data holders, AI system providers, deployers, and certification entities.

3. VALIDATE OPERATIONALLY: Real-world impact

- The final stage of the loop converts the Living Lab's outputs into tangible **tools, guidelines, and processes**. These are then **validated operationally** within the project's pilots to ensure they are effective, scalable, and ready for industry adoption.
- The entire methodology is underpinned by a commitment to **grow the stakeholder group permanently**, ensuring that the ecosystem remains open, transparent, and aligned with the evolving European data economy.



PILOTS (7)

Pilots are at the heart of CERTAIN methodology.
The project's requirements are taking into account the individual pilots requirements defined with suitable stakeholders.
The project's results will **be tested in operational conditions**, to achieve high maturity, equal or higher to the TRL expected in the call for technical results.
It will also prepare the project results for adoption by the AI stakeholders.

Biometrics

Data holders

Health

Bank, finance

Energy

(small) IT companies

Human Resources

Biometric Systems for Border Control

PILOT 1

- **Sector**

Biometrics / Security

- **Core Values**

Fairness, Privacy, and Trust.

- Pilot Purpose

To test a “black box” setup for fairness, robustness against identity fraud (morphing/inference attacks), and the trade-off between energy efficiency and performance

- **Business Focus**

Developing AI-driven biometric recognition (facial data) for high-risk border management (1:1 passport checks) and airline boarding (1:N passenger checks).

- Performance Targets

Bias Reduction: ≥50% reduction in demographic bias.
Energy Efficiency: Tenfold reduction in energy consumption compared to Q1 2024 solutions.
Security: Maintain error rates for presentation and morphing attacks below 10%.

AI Support for Childhood Psychiatry

PILOT 2

- **Sector**

Healthcare

- **Core Values**

Ethics, Clinical Trust, Autonomy, and Transparency.

- Pilot Purpose

To validate clinical impact through Living Labs, ensuring the system provides transparent “risk scores” for shared decision-making while preventing demographic bias.

- **Business Focus**

Utilizing explainable AI and multimodal data (video, audio, sensors) to enhance early detection and monitoring of psychiatric risks in youth and creating privacy preserving Data Sets for secondary use.

- Performance Targets

Bias & Accuracy: Reduce diagnostic bias by 30% and achieve ≥70% sensitivity and specificity.
Acceptance: ≥85% clinician trust and ≥70% trust from parents/children.
Efficiency: 30% reduction in time to diagnosis [Pilot 2 Measurable Improvements].

AI-Driven Energy Planning

PILOT 3

- **Sector**

Energy / Sustainability

- **Core Values**

Accountability, Inclusion, Trust and Reliability.

- **Business Focus**

AI systems for Renewable Energy Communities (RECs) to manage demand forecasting and consumption optimization for households and small businesses.

- Pilot Purpose

Evaluate GDPR compliance regarding personal energy data and assess AI fairness to ensure optimization models benefit small producers, not just large ones.

- Performance Targets

Efficiency: 80% reduction of energy surplus [Pilot 3 Expected Improvements].
Adoption: 30% increase in trust and adoption of automation solutions [Pilot 3 Expected Improvements].

EU Human Capital Market Demand Analysis

PILOT 4

- **Sector**

Human Resources (HR)

- **Core Values**

Data Sovereignty and Explainability; Interoperability & Trustworthiness; Security & Privacy; Transparency and Usability.

- Pilot Purpose

To test the readiness, usability and compliance of the CERTAIN data space for data holders/end-users, focusing on automated regulatory compliance and data sovereignty.

- **Business Focus**

Enabling the trusted and secure publishing of sensitive HR data (GDPR, EU Legal Requirements).
Showcasing the added value of Large Language Models (LLMs) to automate candidate matching, improve recruitment efficiency and labor market trends forecasting.

- Performance Targets

Matching: 30–40% increasing the matching of professionals with job opportunities
Skills: 20-30% Enhancing skills and awareness in high-demand sectors.
Forecasting: 20–30% improving the ability to forecast Human Capital trends [Pilot 4 Expected Improvements].

Compliance Guidance for Data Holders

PILOT 5

- **Sector**

Data Economy / SMEs

- **Core Values**

Accessibility, Demonstrating Compliance, and Cost-Effectiveness

- **Business Focus**

Supporting SMEs in navigating legal and ethical complexities when entering AI data markets through tailored compliance “wizards”.

- Pilot Purpose

To evaluate the usefulness and accessibility of CERTAIN’s simplified guidelines for non-expert users to reduce the burden of EU regulatory compliance.

- Performance Targets

Usability: System Usability Scale (SUS) score > 68.
Efficiency: Significant reduction in time and cost of compliance checks compared to manual audits.

Personalized Investment Recommendations

PILOT 6

- **Sector**

Finance / Retail Banking

- **Core Values**

Suitability, Transparency and Explainability.

- Pilot Purpose

To develop and test counterfactual explanation methods (identifying how different inputs change decisions) to reveal algorithmic biases and improve trustworthiness [Pilot 6 Evaluation Purpose].

- **Business Focus**

Generating personalized (1to1) investment recommendations aligned with customers' risk profile.

- Performance Targets

Reliability: Achieve at least 80% accuracy and suitable recommendations.
Speed: Recommendations generated within 20 seconds.

Automating MLOps for regulatory compliance in digital solutions deployment

PILOT 7

- **Sector**

Information Technology (IT) / RegOps , MLOps

- **Core Values**

Compliance, Efficiency, Ethics, Risk Reduction, and Market Advantage.

- Pilot Purpose

To evaluate the effectiveness of streamlining the compliance process in addressing legal, ethical, and operational challenges in a product deployment environment.

- **Business Focus**

Delivering AI-boosted and compliant-by-design technical solutions and products that adhere to EU regulations.

- Performance Targets

Compliance: ≥90% of digital solutions successfully pass compliance checks.
Cost & Speed: ≥15% reduction in compliance costs and ≥60% reduction in time for security assessments.

CONSORTIUM

IDEMIA PUBLIC SECURITY

UNIVERSITÉ DU LUXEMBOURG

fh st.pölten

Netcompany

Univerza v Mariboru
Faculty of Electrical Engineering and Computer Science

UKC MARIBOR Univerzitetni klinični center Maribor

NATIONAL BANK OF GREECE

eit Digital

ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
HAROKOPIO UNIVERSITY

UNIVERSITY OF TARTU
1632

nvision

EMPOWER
BÜRGERENERGIEGEMEINSCHAFT

RED ALERT LABS
IoT Security

DEXAI
ARTIFICIAL ETHICS

INCOM Ltd.
INTERNATIONAL CONSULTING & MANAGEMENT

novelcore
RESEARCH & INNOVATION

ANADELTA

idiap
RESEARCH INSTITUTE

DIGITAL FOR PLANET

CERTAIN

Thank you!

www.certain-project.eu

info@certain-project.eu

Follow our advancements
on www. certain-project.eu

@certain-project @certainproject @certain-project

The CERTAIN project received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No 101189650.
This work has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI).

Co-funded by the European Union

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI