XR5.0

Human-Centric AI-Enabled
Extended Reality Applications
for the Industry 5.0 Era

27 January 2025

# Ethical Management and AI Governance Framework in XR/AI projects

Marcelo Corrales Compagnucci

Funded by
the European Union

# XR5.0 Project Overview

**Goal:** Develop adv. XR apps (VR, AR and MR) designed specifically for industrial workers as part of the industry 5.0.

**Why this matters?** Current XR tools improve safety and production but lack *personalization.*

XR5.0 focuses on *customized, worker-centric* XR experiences that adapt to individual skills, traits, and work environments, promoting efficiency and safety.

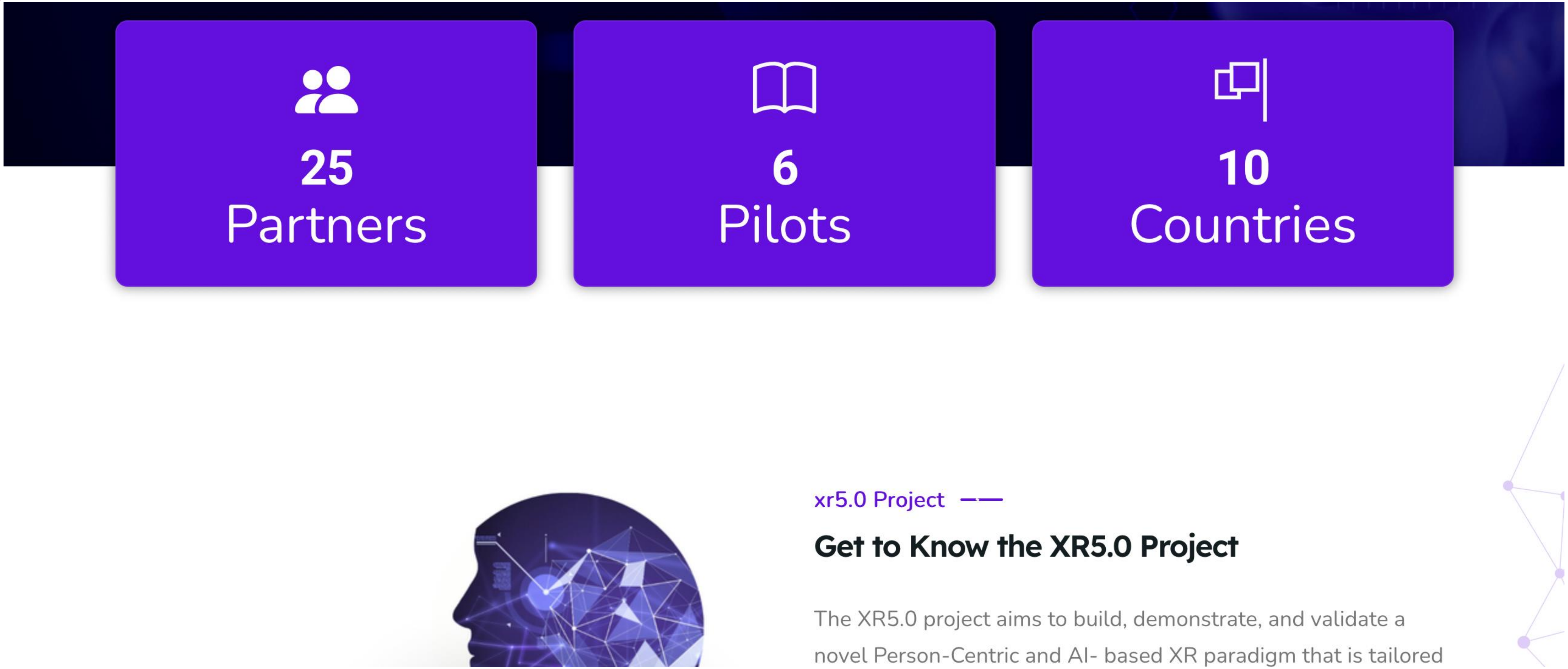
# Key Highlights

**AI integration:** Advanced AI will enhance XR by creating dynamic and *intelligent visualizations* for industrial tasks.

**Human-centered DT:** Virtual replicas of physical processes will enable better monitoring, training, and personalization.

**European values & compliance:** Technology will reflect European standards for *safety*, *privacy* and *ethical* AI use.

**Training and upskilling:** Industrial workers will train using *ergonomic* and *customized* XR environments.



25
Partners

6
Pilots

10
Countries

xr5.0 Project

### Get to Know the XR5.0 Project

The XR5.0 project aims to build, demonstrate, and validate a novel Person-Centric and AI- based XR paradigm that is tailored

# Pilots 1,2 & 3

## XR5.0 Pilots

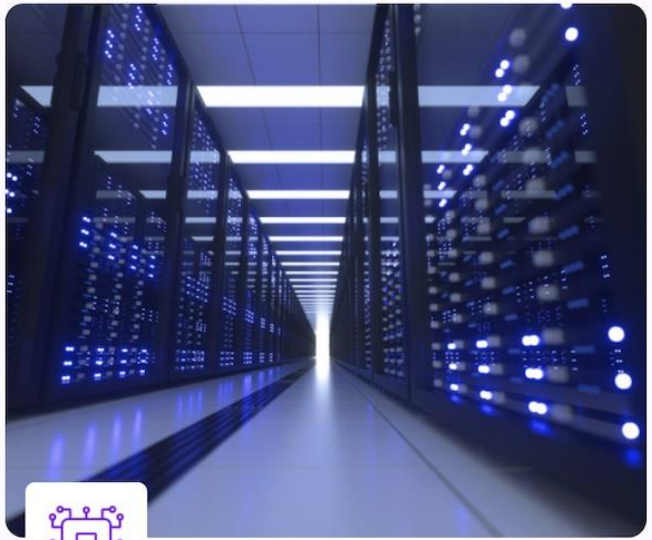
**Rapid Human Centric AI-Enabled Product Design**

**Human Centred Remote Maintenance and Asset Management**

**Operator 5.0 Training for Smart Water Pipes based on XR Streaming**

# Pilots 4, 5 & 6

**Worker Centric Aircraft Maintenance Training**



**Increased Effectiveness and Safety of Product Assembly and Repair Processes**

**Human Centric Guidance and Troubleshooting for Customer Service**

# Ethical and Legal Framework

### Mandatory legal provisions

General Data Protection Regulation (GDPR)

EU AI Act

Liability reform for AI: Amendment to the Product Liability Directive & AI Liability Directive

Cybersecurity regulations: NIS 2, Cybersecurity Act, Cyber Resilience Act, etc.

### Ethical Principles & Guidelines

The High-Level Expert Group Guidelines on Trustworthy AI

OECD Council Recommendation on AI

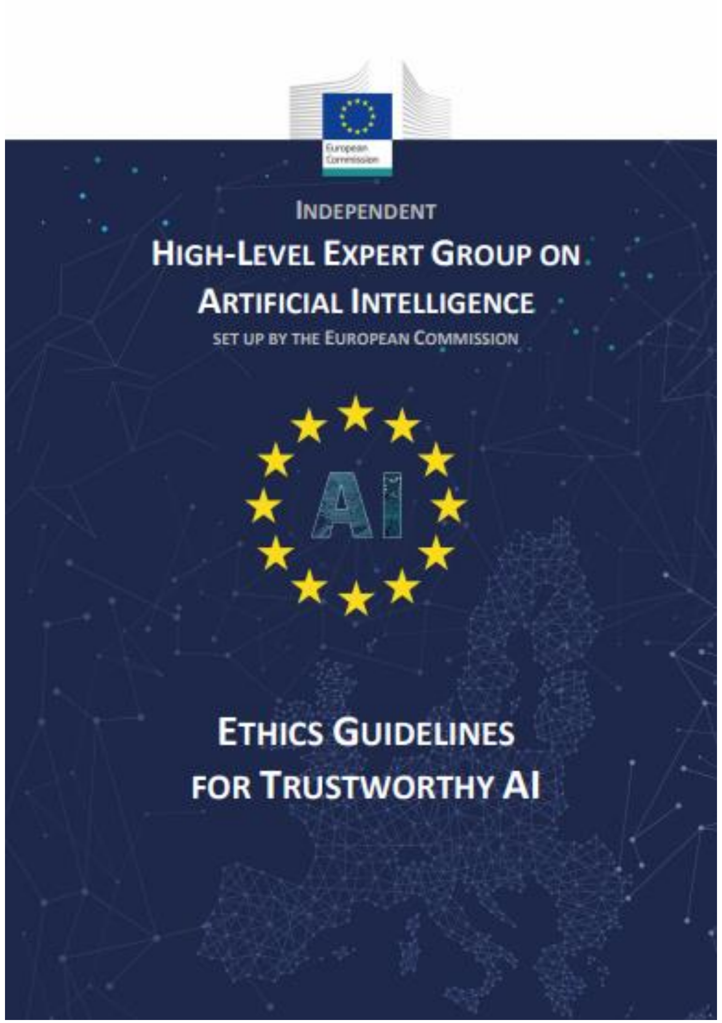
### Standards & Frameworks

ISO Standards: ISO 27001, ISO 27701, ISO 23053, ISO 23894, ISO 42001, ISO 31000, ISO Guide 51

NIST AI Risk Management Framework (NIST AI RMF)

HUDERIA (Human Rights, Democracy, and Rule of Law Impact Assessment)

# HLEG Guidelines & ALTAI



European Commission

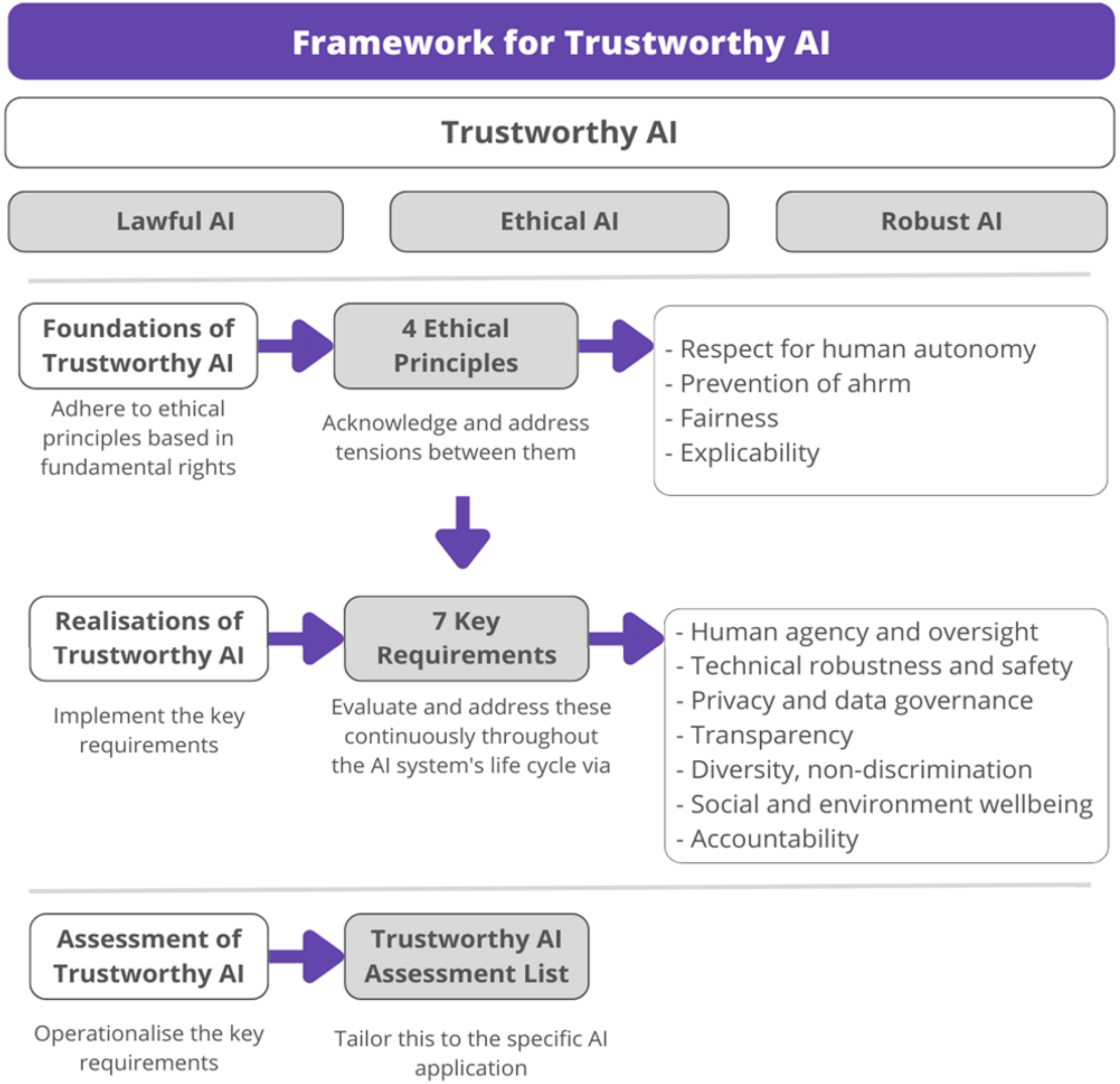
INDEPENDENT

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE

SET UP BY THE EUROPEAN COMMISSION

AI

ETHICS GUIDELINES FOR TRUSTWORTHY AI



# Framework for Trustworthy AI

## Trustworthy AI

Lawful AI

Ethical AI

Robust AI

**Foundations of Trustworthy AI** → **4 Ethical Principles** →

- Respect for human autonomy
- Prevention of ahrm
- Fairness
- Explicability

Adhere to ethical principles based in fundamental rights

Acknowledge and address tensions between them

**Realisations of Trustworthy AI** → **7 Key Requirements** →

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination
- Social and environment wellbeing
- Accountability

Implement the key requirements

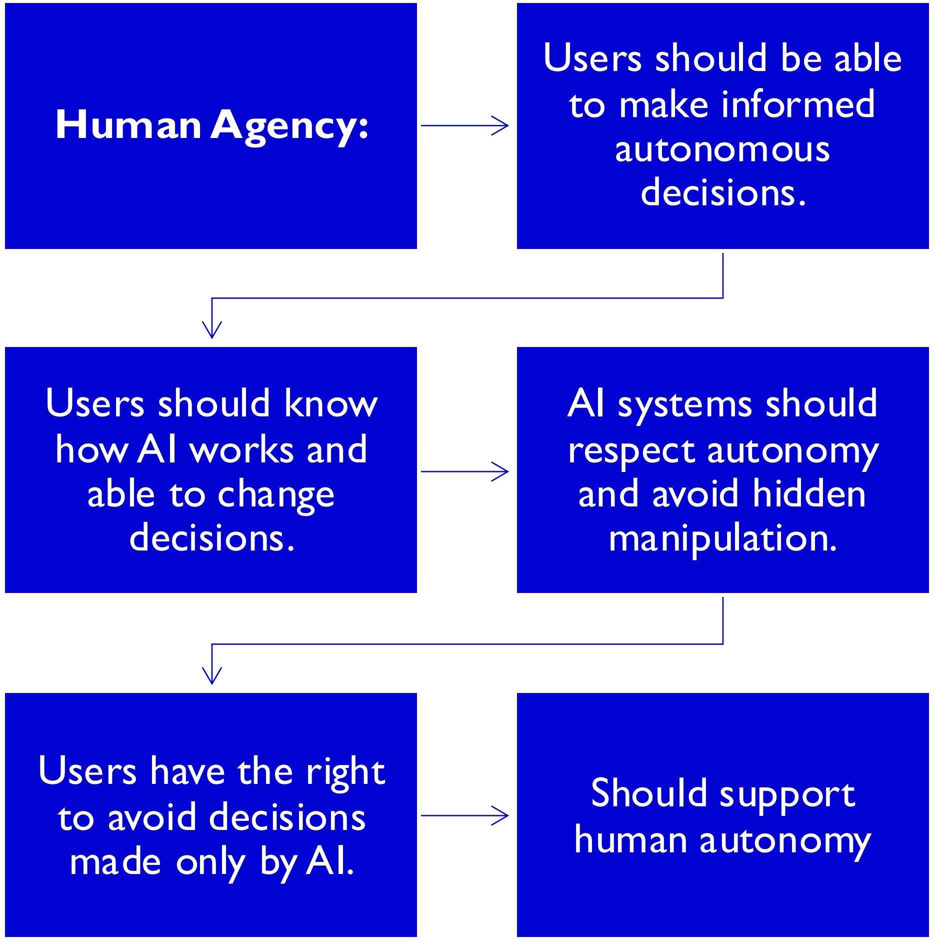
Evaluate and address these continuously throughout the AI system's life cycle via

**Assessment of Trustworthy AI** → **Trustworthy AI Assessment List**

Operationalise the key requirements

Tailor this to the specific AI application

# **Requirement 1-** Human Agency and Oversight



**Human Agency:**

Users should be able to make informed autonomous decisions.

Users should know how AI works and able to change decisions.

AI systems should respect autonomy and avoid hidden manipulation.

Users have the right to avoid decisions made only by AI.

Should support human autonomy

Source Chat GPT: AI generated image depicting a worker using XR

# Human Oversight

- **Human in the Loop (HITL):** A person can step in at each stage of the AI's decision cycle.

- **Human on the Loop (HOTL):** A person monitors the AI during its decision and operation but does not interfere everytime. Human intervention is during the design and monitoring.

- **Human in Command (HIC):** A person oversees the entire system and can choose when to use or stop the AI. Considers the broader impact (econonomic, societal, legal & ethical).

- Makes sure the AI systems does not harm people or take away their control.
- Key action points: determine whether the AI system is *self-learning* vs. *fixed-rule AI.*
- Self-learning needs closer monitoring or the ability to override.
- Fixed-rule AI is more predicatable, less oversight needed.

# Requirement 2 - Technical Robustness & Safety

### RESILIENCE TO ATTACKS

- ENSURE AI IS CERTIFIED

- ASSESS ADVERSARIAL ATTACKS

- ATTACKS: DATA, MODEL, INFRASTRUCTURE

IMPL. SECURITY

### FALLBACK PLAN AND GENERAL SAFETY

- ENSURE A FALLBACK PLAN

- IDENTIFY & ASSESS THE RISKS WITH UC

- INFORM USERS OF POTENTIAL RISKS

### ACCURACY

- HOW WELL AN AI SYSTEM MAKES CORRECT DECISIONS/PREDICTIONS

- ENSURE TRAINING DATA IS ACCURATE, HIGH QUALITY, REPRESENTATIVE

- IMPLEMENT MEASURES TO MONITOR METRICS

### RELIABILITY AND REPRODUC.

- RELIABLE :THE SYSTEM WORKS PROPERLY WITH A RANGE OF INPUTS AND SITUATIONS.

- REPRODUCIBLE: GIVES CONSISTENT AND CORRECT RESULTS. EXHIBITS THE SAME BEHAVIOUR WHEN REPEATED UNDER THE SAME CONDITIONS.

# **Requirement 3** – Privacy & Data Governance

| | |
|---|---|
| Assess | If AI system processes personal data (including sensitive data). |
| Implement | Necessary GDPR measures (DPIA, DPO, oversight mechanism, etc.). |
| Implement | Privacy-by-Design measures (anon. and pseudonym.); Access protocols. |
| Integrate | Data subject rights into the development of AI systems. |
| Consider | Data protection of non-personal data. |
| Align | The AI system with relevant data management and governance standards (e.g., ISO 27701 PIMS and IEEE). |
| Ensure | Data quality and integrity (avoid biases, inaccuracies, errors, mistakes) – Needs to be addressed before the training of the dataset. |

# Requirement 4 – Transparency

**Traceability:** How the AI system works, what it does, and how affects people (like the data it uses, how decisions are made, and algorithms involved). Should all be documented. It facilitates auditability.

**Explainability:** People should be able to understand how the AI system works and why it makes specific decisions.

**Communication:** People have the right to know when they are interacting with an AI system. The system's strenghts, weaknesses and accuracy should be clearly communicated in a manner appropriate to the UC.

# **Requirement 5** – Diversity, Non-Discrimination, and Fairness

### Avoidance of unfair bias

AI systems use data (training and operation) to learn and make decisions. If this data contains historical biases, imcompleteness or errors, the AI might treat people unfairly or discriminate.

### Accessibility and universal design

AI systems should be *user-centric* and design so that *everyone* can use them. Follow universal design principles and use accessibility standards for people with disabilities.

### Stakeholder participation

All stakeholders (workers, customers, communities) should be involved in its development and deployment. Get feedback during and after development, like consultations and worker committees.

# **Requirement 6** – Societal and Environmental Well-Beign

### Sustainable and environmentally friendly AI

AI systems should help solve big problems while minimizing the environmental impact. AI systems can consume a lot of energy and resources – should use energy-efficiency methods and ensure it is env. responsible.

### Social impact

AI systems affect how people interact (relationships, mental and physical well-being). It can improve social skills but also harm relationships – should monitor how AI changes behavior, especially in areas like education and work.

### Society and democracy

AI systems can influence democracy and societal institutions (e.g., elections). Improper use in political context can harm society – carefully evaluate AI systems, ensure it strenghthens democratic principles and institutions.

# Requirement 7 – Accountability

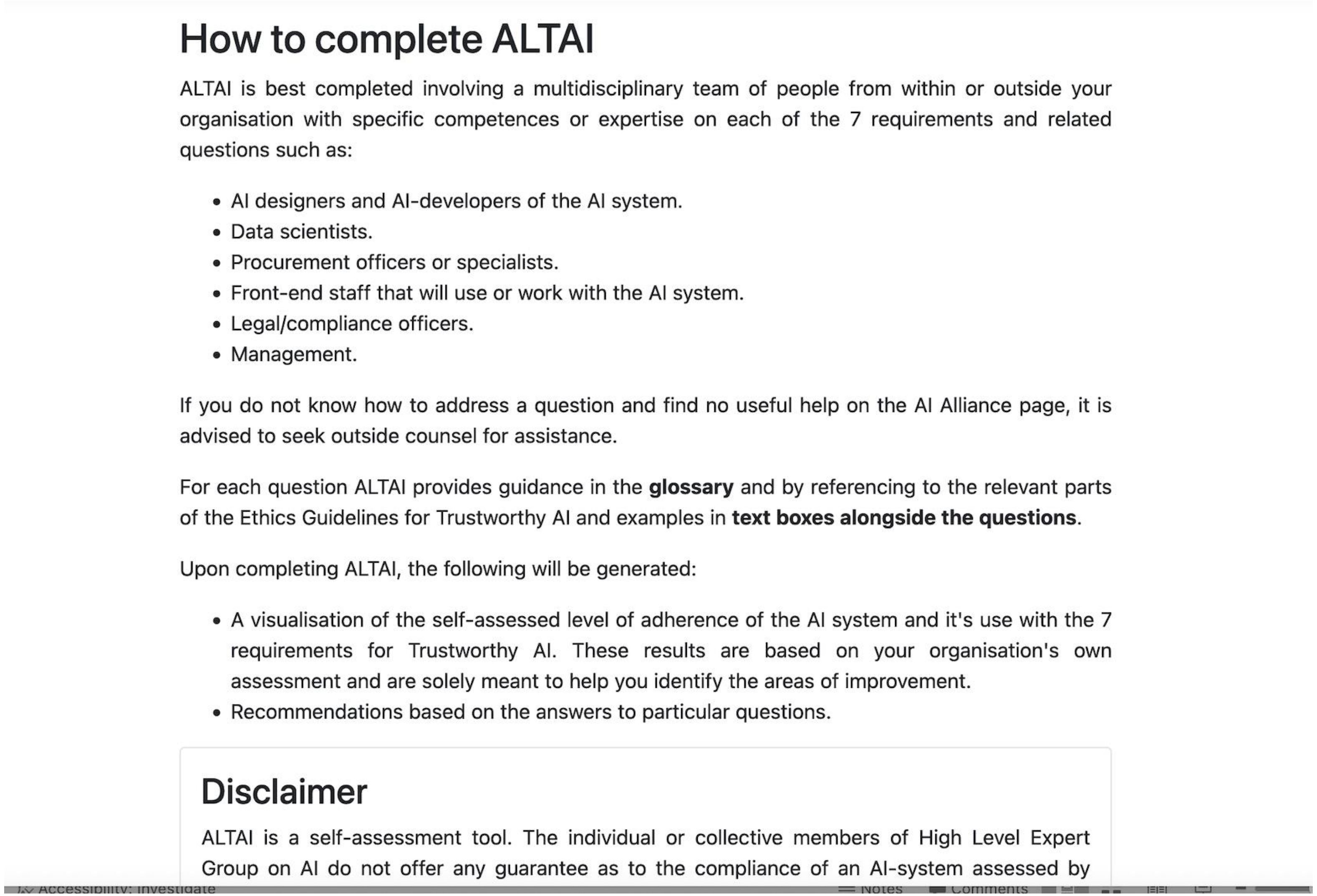
**Auditability:** AI systems need to be checked to make sure they work as intended and don't cause harm (incl. algorithms, data and design). Independent audits are essential for systems that impact fundamental rights like safety-critical tools.

**Minimization and reporting of negative impacts:** AI systems should be design to reduce harm and address problems – should use tools like impact assessments (e.g., simulations and tests) to reduce risks. Ensure people can report issues.

**Trade-offs:** sometimes ethical principles or systems goals might conflict – should balance these conflicts.

**Redress:** if someone is unfairly impacted by an AI system, there should be ways to correct the harm (esp. vulnerable groups).

# How to complete the ALTAI?



## How to complete ALTAI

ALTAI is best completed involving a multidisciplinary team of people from within or outside your organisation with specific competences or expertise on each of the 7 requirements and related questions such as:

- AI designers and AI-developers of the AI system.
- Data scientists.
- Procurement officers or specialists.
- Front-end staff that will use or work with the AI system.
- Legal/compliance officers.
- Management.

If you do not know how to address a question and find no useful help on the AI Alliance page, it is advised to seek outside counsel for assistance.

For each question ALTAI provides guidance in the **glossary** and by referencing to the relevant parts of the Ethics Guidelines for Trustworthy AI and examples in **text boxes alongside the questions**.

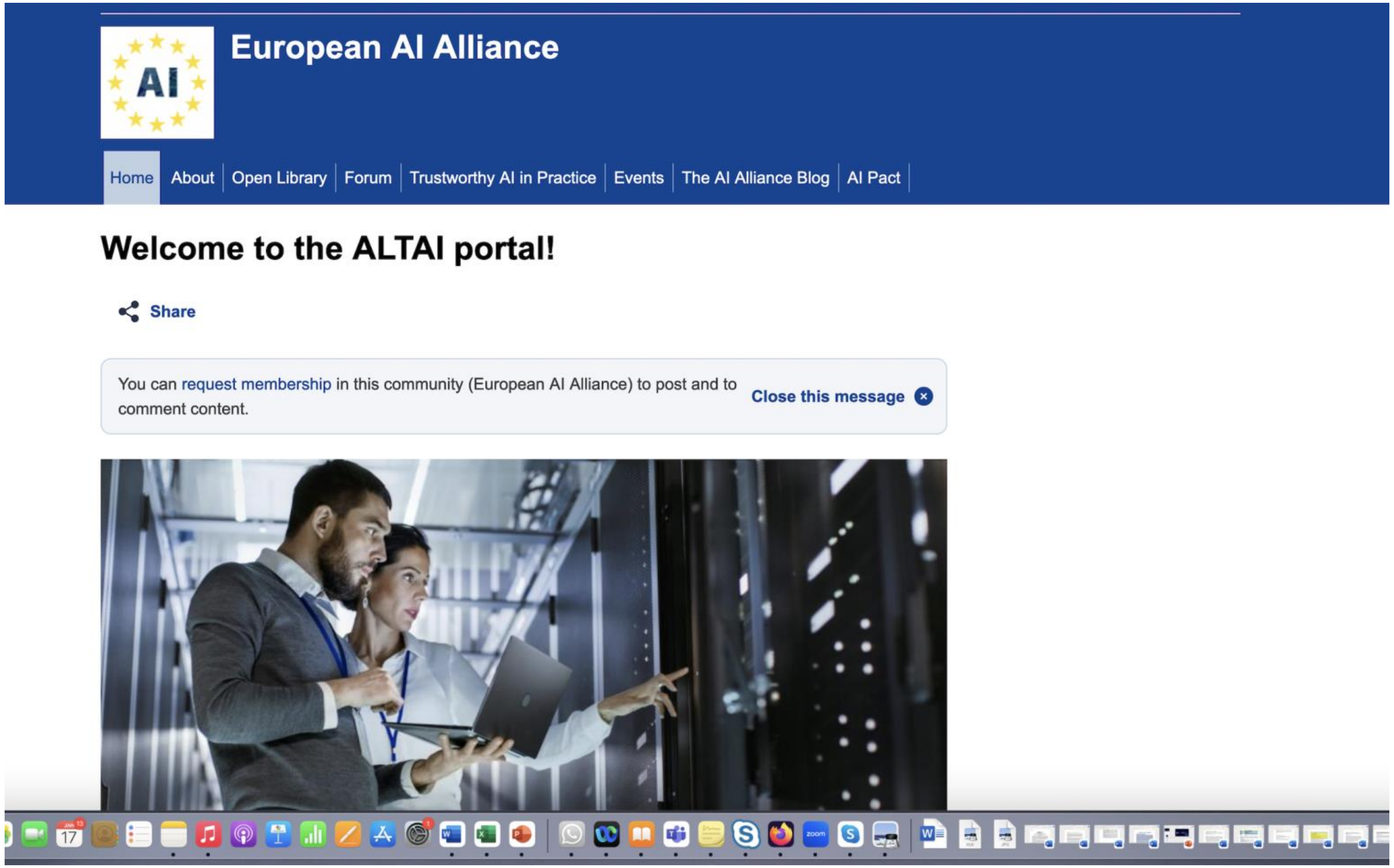
Upon completing ALTAI, the following will be generated:

- A visualisation of the self-assessed level of adherence of the AI system and it's use with the 7 requirements for Trustworthy AI. These results are based on your organisation's own assessment and are solely meant to help you identify the areas of improvement.
- Recommendations based on the answers to particular questions.

## Disclaimer

ALTAI is a self-assessment tool. The individual or collective members of High Level Expert Group on AI do not offer any guarantee as to the compliance of an AI-system assessed by

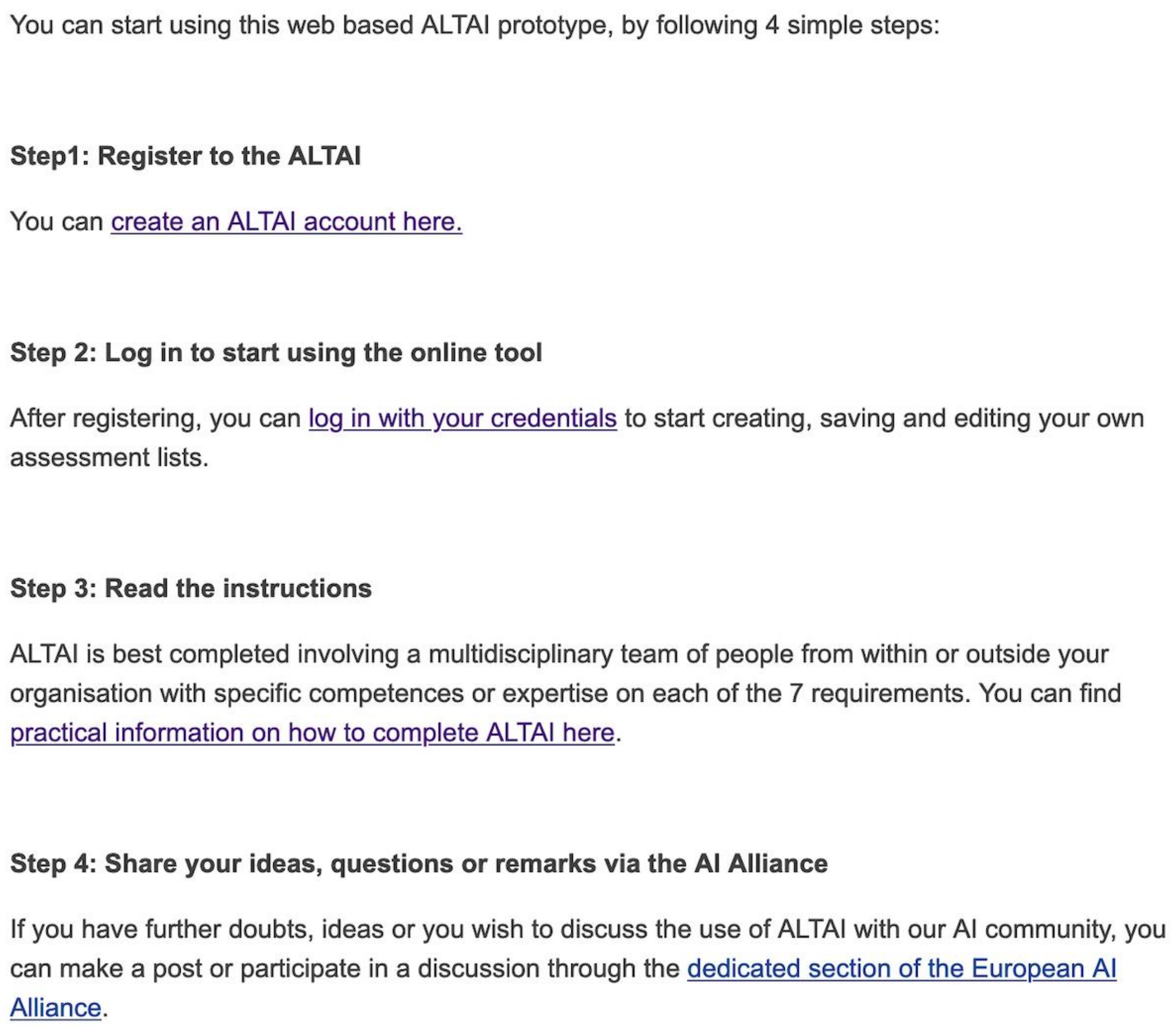
# Go to the ALTAI portal: [https://futurium.ec.europa.eu/en/european-ai-alliance/pages/welcome-altai-portal](https://futurium.ec.europa.eu/en/european-ai-alliance/pages/welcome-altai-portal)



Welcome to the ALTAI portal!

You can request membership in this community (European AI Alliance) to post and to comment content.

# Follow these 4 simple steps



You can start using this web based ALTAI prototype, by following 4 simple steps:

**Step1: Register to the ALTAI**

You can create an ALTAI account here.

**Step 2: Log in to start using the online tool**

After registering, you can log in with your credentials to start creating, saving and editing your own assessment lists.

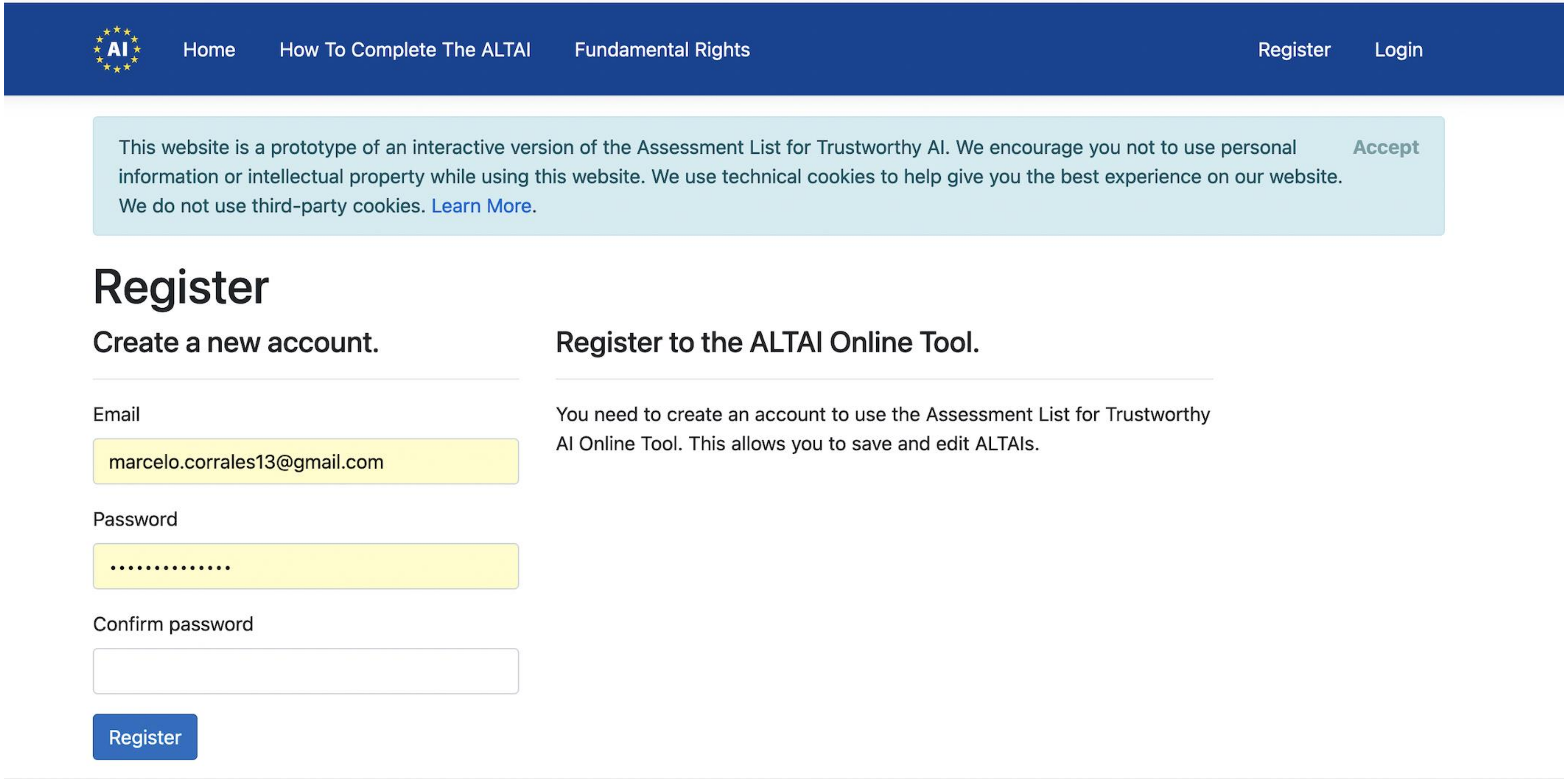
**Step 3: Read the instructions**

ALTAI is best completed involving a multidisciplinary team of people from within or outside your organisation with specific competences or expertise on each of the 7 requirements. You can find practical information on how to complete ALTAI here.

**Step 4: Share your ideas, questions or remarks via the AI Alliance**

If you have further doubts, ideas or you wish to discuss the use of ALTAI with our AI community, you can make a post or participate in a discussion through the dedicated section of the European AI Alliance.

# Registration page – Create a new account



This website is a prototype of an interactive version of the Assessment List for Trustworthy AI. We encourage you not to use personal information or intellectual property while using this website. We use technical cookies to help give you the best experience on our website. We do not use third-party cookies. Learn More.

Register

Create a new account.

Email: marcelo.corrales13@gmail.com

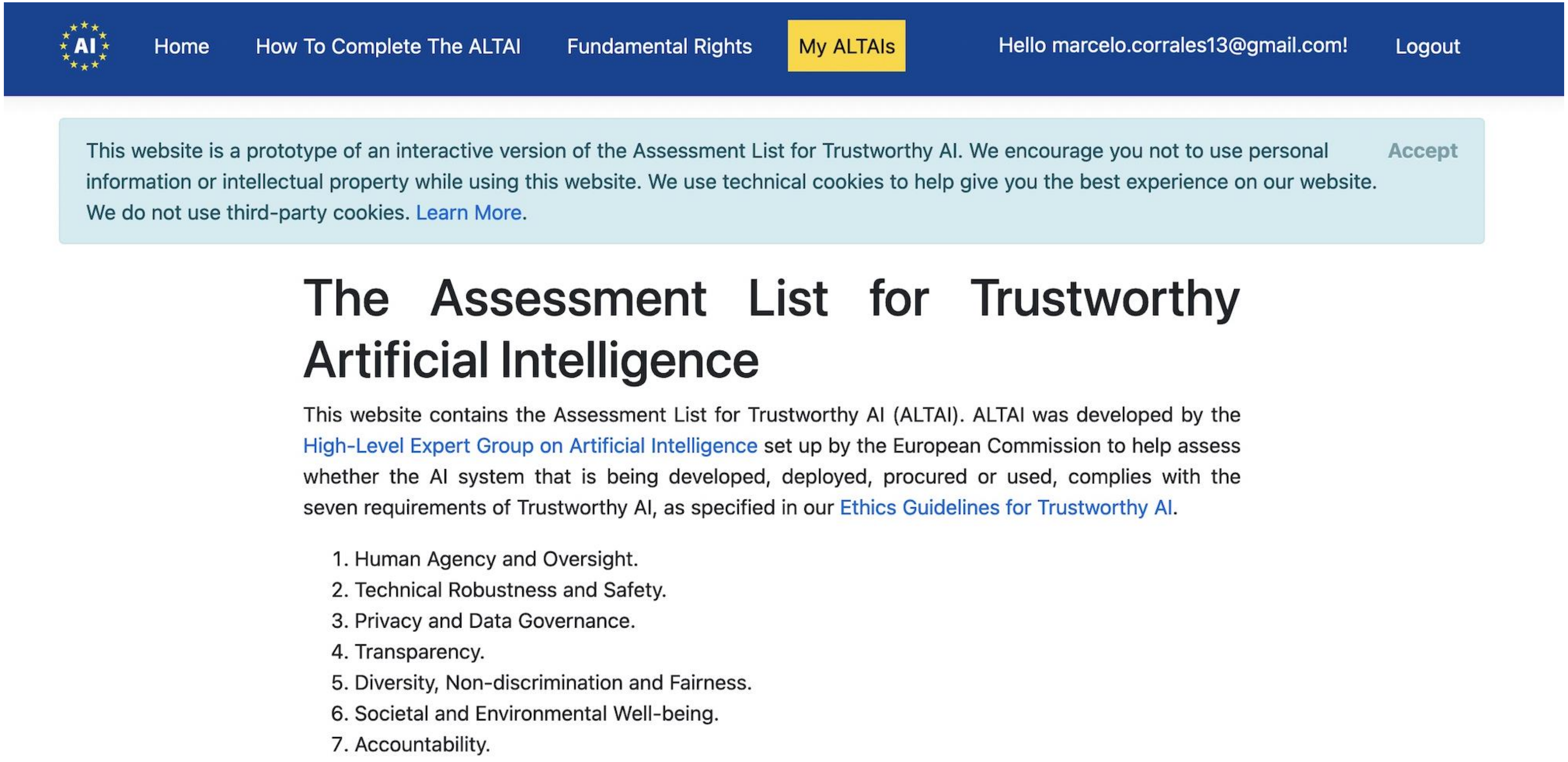
Password: ••••••••••••••

Confirm password:

Register to the ALTAI Online Tool.

You need to create an account to use the Assessment List for Trustworthy AI Online Tool. This allows you to save and edit ALTAIs.

# Login page – Go to “My ALTAIs”



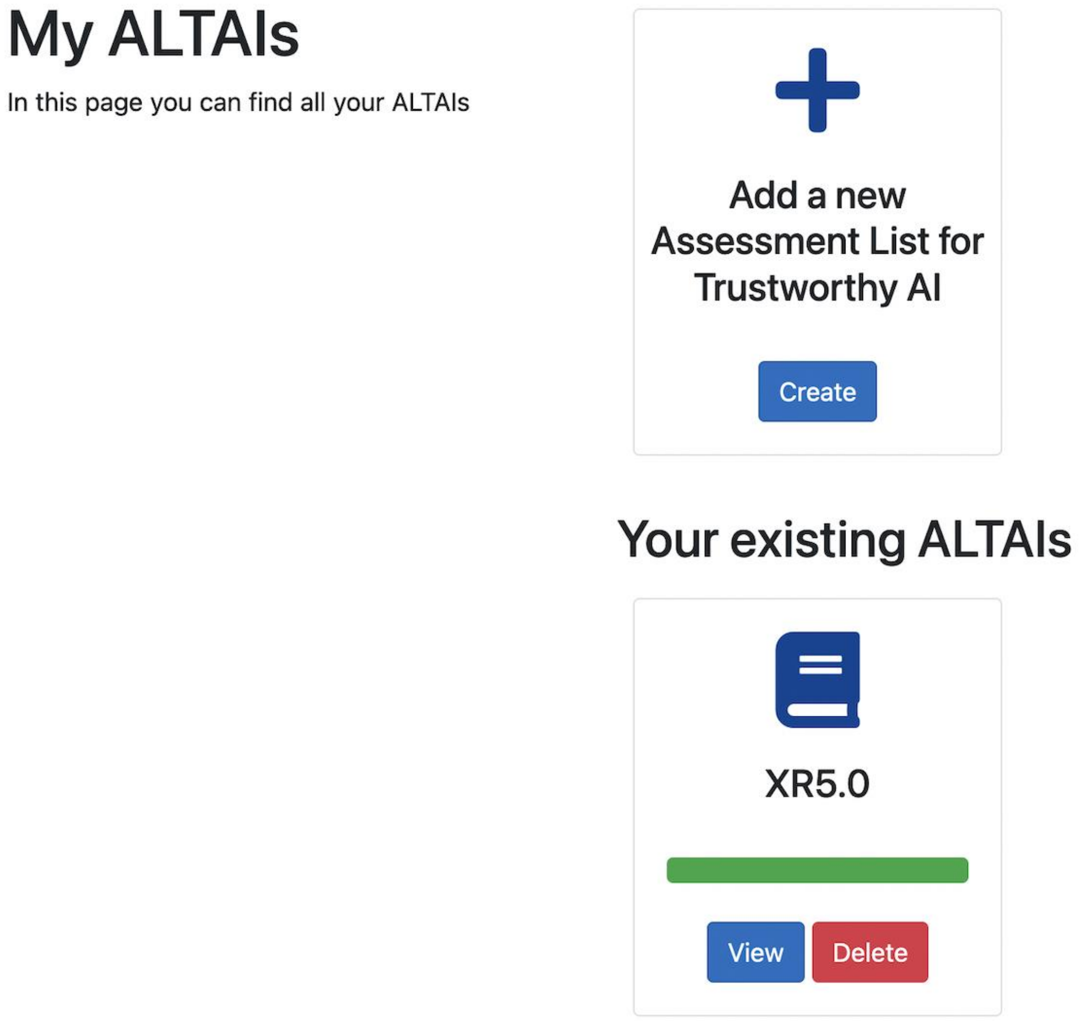
AI
Home
How To Complete The ALTAI
Fundamental Rights
My ALTAIs
Hello marcelo.corrales13@gmail.com!
Logout
Accept

This website is a prototype of an interactive version of the Assessment List for Trustworthy AI. We encourage you not to use personal information or intellectual property while using this website. We use technical cookies to help give you the best experience on our website. We do not use third-party cookies. Learn More.

## The Assessment List for Trustworthy Artificial Intelligence

This website contains the Assessment List for Trustworthy AI (ALTAI). ALTAI was developed by the High-Level Expert Group on Artificial Intelligence set up by the European Commission to help assess whether the AI system that is being developed, deployed, procured or used, complies with the seven requirements of Trustworthy AI, as specified in our Ethics Guidelines for Trustworthy AI.

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.



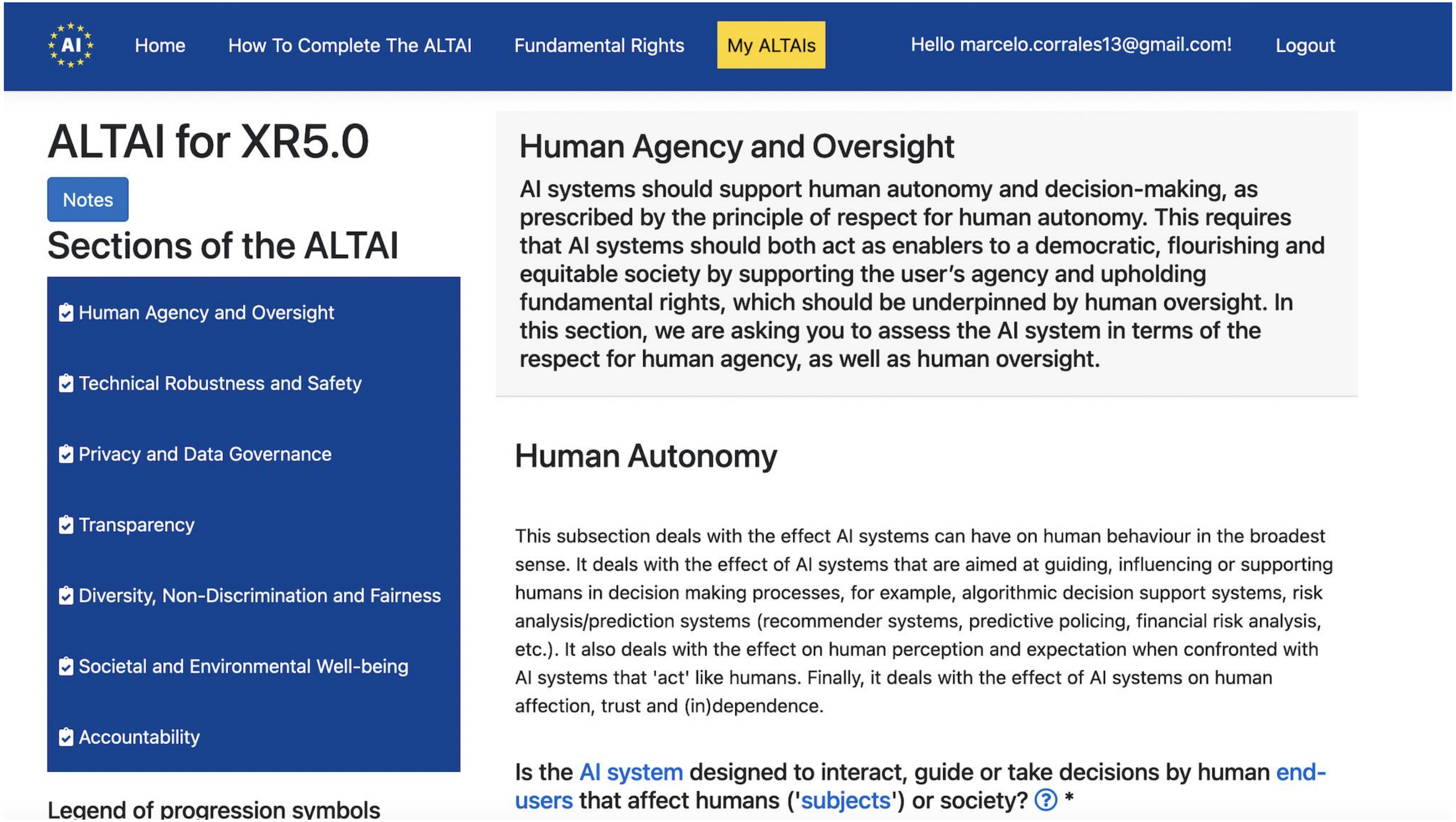
# My ALTAIs

In this page you can find all your ALTAIs

Add a new Assessment List for Trustworthy AI

## Your existing ALTAIs

XR5.0


AI
Home
How To Complete The ALTAI
Fundamental Rights
My ALTAIs
Hello marcelo.corrales13@gmail.com!
Logout
Notes

# ALTAI for XR5.0

## Sections of the ALTAI

- ✓ Human Agency and Oversight
- ✓ Technical Robustness and Safety
- ✓ Privacy and Data Governance
- ✓ Transparency
- ✓ Diversity, Non-Discrimination and Fairness
- ✓ Societal and Environmental Well-being
- ✓ Accountability

Legend of progression symbols

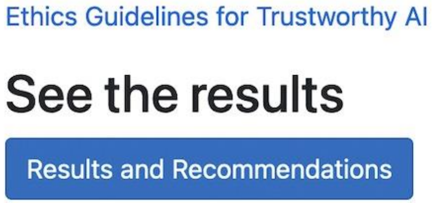
## Human Agency and Oversight

AI systems should support human autonomy and decision-making, as prescribed by the principle of respect for human autonomy. This requires that AI systems should both act as enablers to a democratic, flourishing and equitable society by supporting the user’s agency and upholding fundamental rights, which should be underpinned by human oversight. In this section, we are asking you to assess the AI system in terms of the respect for human agency, as well as human oversight.

## Human Autonomy

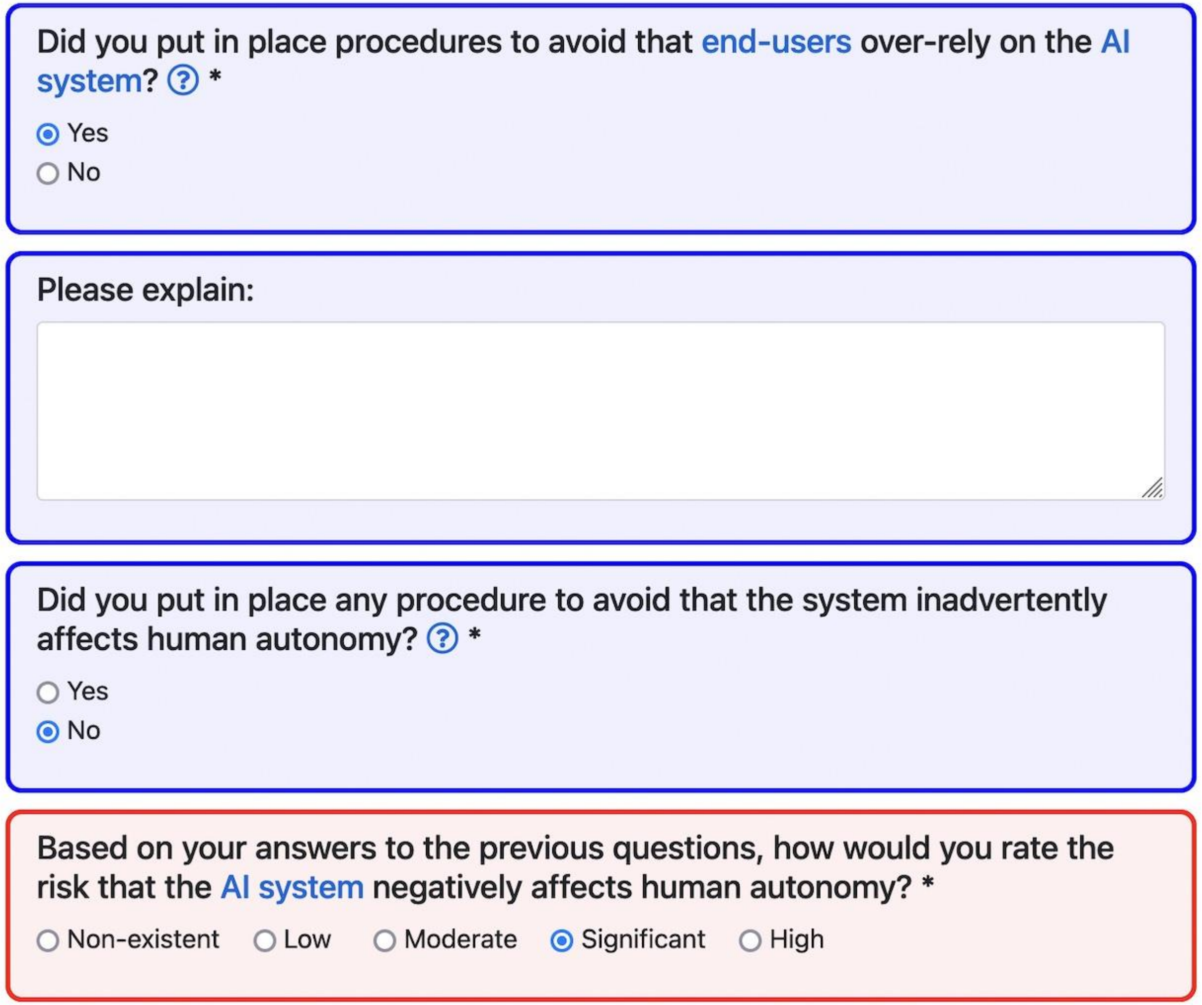
This subsection deals with the effect AI systems can have on human behaviour in the broadest sense. It deals with the effect of AI systems that are aimed at guiding, influencing or supporting humans in decision making processes, for example, algorithmic decision support systems, risk analysis/prediction systems (recommender systems, predictive policing, financial risk analysis, etc.). It also deals with the effect on human perception and expectation when confronted with AI systems that 'act' like humans. Finally, it deals with the effect of AI systems on human affection, trust and (in)dependence.

Is the AI system designed to interact, guide or take decisions by human end-users that affect humans ('subjects') or society? ? *



Ethics Guidelines for Trustworthy AI

## See the results



Did you put in place procedures to avoid that end-users over-rely on the AI system? ? *

● Yes
○ No

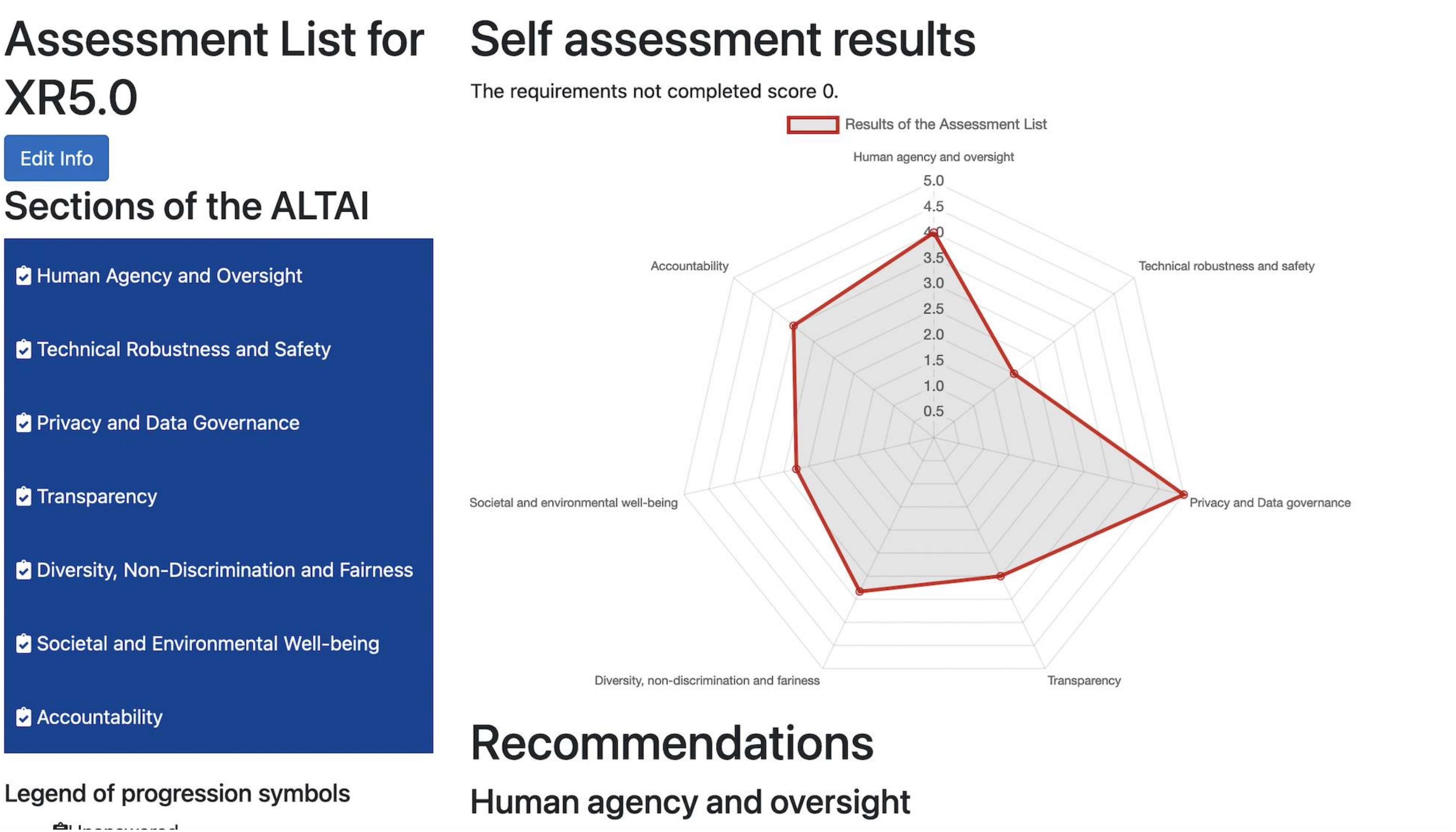
Please explain:

Did you put in place any procedure to avoid that the system inadvertently affects human autonomy? ? *

○ Yes
● No

Based on your answers to the previous questions, how would you rate the risk that the AI system negatively affects human autonomy? *

○ Non-existent ○ Low ○ Moderate ● Significant ○ High



# Assessment List for XR5.0

Edit Info

## Sections of the ALTAI

- ☑ Human Agency and Oversight
- ☑ Technical Robustness and Safety
- ☑ Privacy and Data Governance
- ☑ Transparency
- ☑ Diversity, Non-Discrimination and Fairness
- ☑ Societal and Environmental Well-being
- ☑ Accountability

### Legend of progression symbols

Unanswered

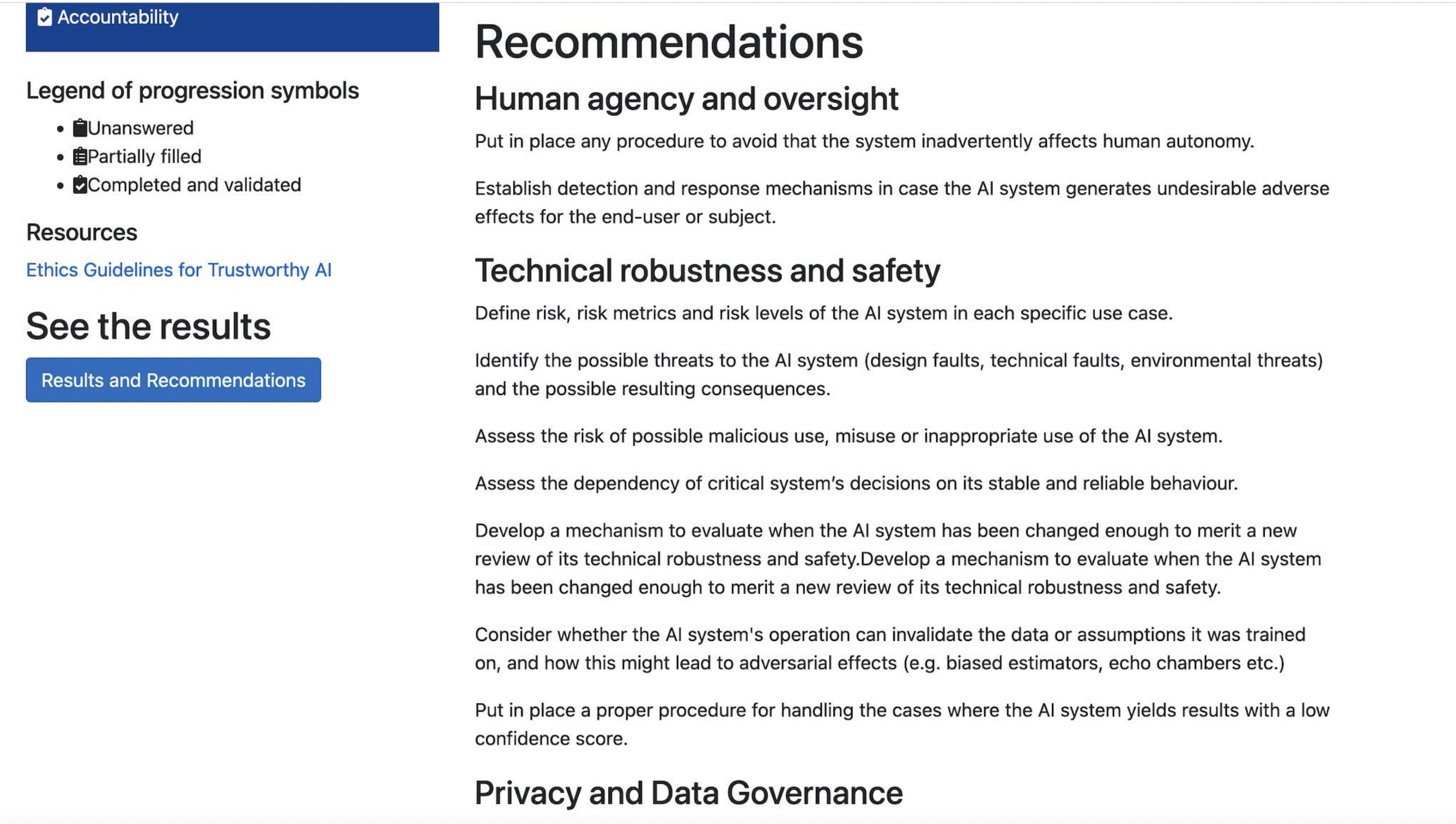
# Self assessment results

The requirements not completed score 0.

Results of the Assessment List
Human agency and oversight
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
Technical robustness and safety
Privacy and Data governance
Transparency
Diversity, non-discrimination and fariness
Societal and environmental well-being
Accountability

# Recommendations

## Human agency and oversight



☑ Accountability

### Legend of progression symbols

- ☐Unanswered
- ▤Partially filled
- ☑Completed and validated

### Resources

Ethics Guidelines for Trustworthy AI

## See the results

Results and Recommendations

# Recommendations

## Human agency and oversight

Put in place any procedure to avoid that the system inadvertently affects human autonomy.

Establish detection and response mechanisms in case the AI system generates undesirable adverse effects for the end-user or subject.

## Technical robustness and safety

Define risk, risk metrics and risk levels of the AI system in each specific use case.

Identify the possible threats to the AI system (design faults, technical faults, environmental threats) and the possible resulting consequences.

Assess the risk of possible malicious use, misuse or inappropriate use of the AI system.

Assess the dependency of critical system's decisions on its stable and reliable behaviour.

Develop a mechanism to evaluate when the AI system has been changed enough to merit a new review of its technical robustness and safety.Develop a mechanism to evaluate when the AI system has been changed enough to merit a new review of its technical robustness and safety.

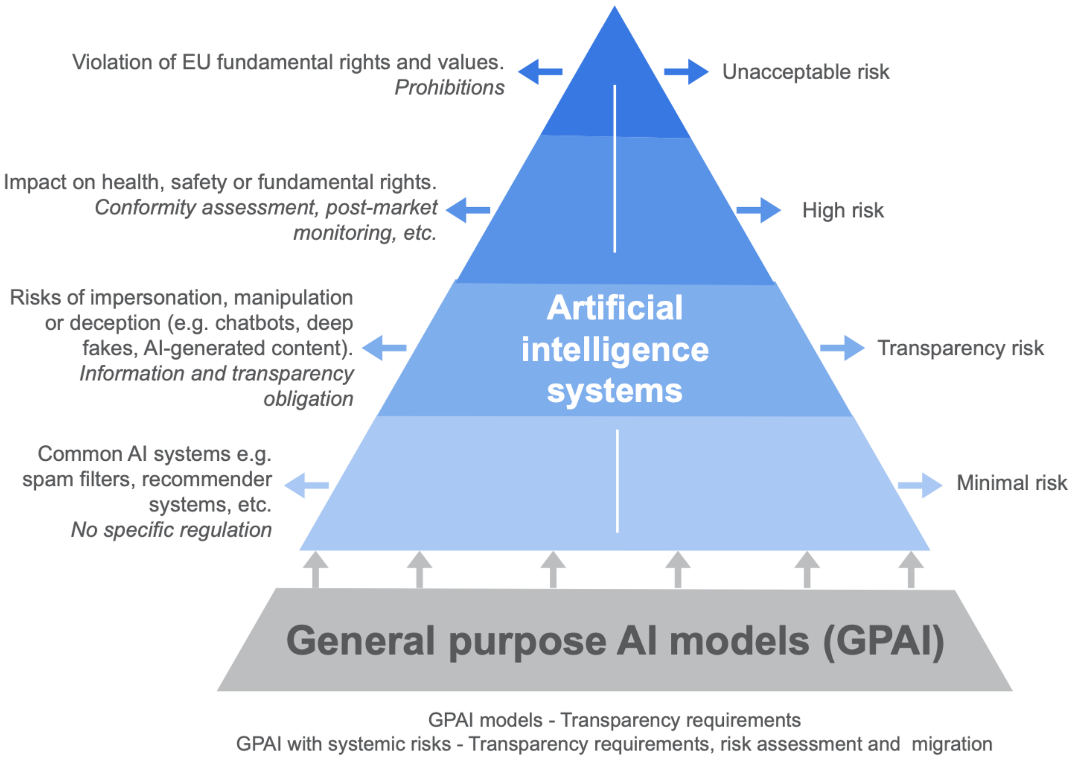
Consider whether the AI system's operation can invalidate the data or assumptions it was trained on, and how this might lead to adversarial effects (e.g. biased estimators, echo chambers etc.)

Put in place a proper procedure for handling the cases where the AI system yields results with a low confidence score.

## Privacy and Data Governance

# The AI Act takes a risk-based approach.

## The project needs to identify under which category each AI systems fall



Violation of EU fundamental rights and values.
*Prohibitions*

Unacceptable risk

Impact on health, safety or fundamental rights.
*Conformity assessment, post-market monitoring, etc.*

High risk

Artificial intelligence systems

Risks of impersonation, manipulation or deception (e.g. chatbots, deep fakes, AI-generated content).
*Information and transparency obligation*

Transparency risk

Common AI systems e.g. spam filters, recommender systems, etc.
*No specific regulation*

Minimal risk

General purpose AI models (GPAI)

GPAI models - Transparency requirements
GPAI with systemic risks - Transparency requirements, risk assessment and  migration

**Source: European Commission**

## The higher the risk to cause harm to society, the stricter the rules

- The majority of the AI Act requirements apply to high-risk AI systems.
- Limited-risk AI systems entail fewer requirements compared to high-risk systems, albeit with specific transparency obligations.
- Distinct requirements apply to providers and deployers of limited-risk AI systems.
- Deployers are required to: Inform and obtain consent from individuals exposed to authorized emotion recognition or biometric categorization systems.

# What does the AI Act do?

Creates harmonized rules for the placing on the market, putting into services & use of AI systems.

Prohibits certain AI practices.

Establishes requirements fo high-risk AI systems.

Creates transparency rules for certain AI systems.

Creates rules fo the placing on the market of general-purpose AI models.

Creates rules on market monitoring & enforcement.

Establishes measures to support innovation.

# High-level Requirements & Recommendations to Project Partners According to the AI Act

## High Level Requirements

- Robust data and data governance.
- Record keeping and logging.
- Transparency and provision of information to users.
- Human oversight.
- Accuracy, robustness and cybersecurity.
- Quality management system.
- Conformity assessment.

## General Recommendations

- Identify High-Risk AI Systems.
- Determine Provider or Deployer (User) Status.
- Conduct Gap Analysis.
- Stay Informed on Technical Standards.
- Integrate AI ethics alignment throughout the lifecycle.

XR5.0

Human-Centric AI-Enabled
Extended Reality Applications
for the Industry 5.0 Era

# Thank You!

www.xr50.eu

Funded by
the European Union