# Trustworthy AI by design: Insights from the AI, Big Data and Democracy Taskforce
## The AI4Gov project

George Manias

**University of Piraeus Research Centre (UPRC)**

Technical Coordinator of AI4Gov

AI4Gov
Trusted AI for Transparent Public Governance
fostering Democratic Values

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
**UNIVERSITY OF PIRAEUS**

Webinar, October 1st

# Outline

## Trusted AI in the context of AI4Gov Project

- The AI4Gov Project

- Main pillars

- Data and AI

- Challenges of AI

- Bias in AI cycle

- Ethical aspects of AI

- Trusted AI: The AI4Gov Project

- Integrated solutions

# AI4Gov: Trusted AI for Transparent Public Governance fostering Democratic Values

- Grant Agreement ID: 101094905
  - Funded by the Horizon Europe
  - Call: HORIZON-CL2-2022-DEMOCRACY-01

- Start – End dates: January 1, 2023 –December 31, 2025 (36 months)

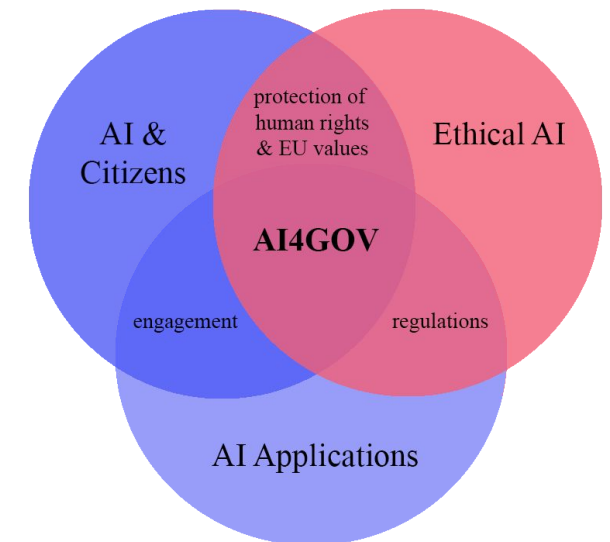## Consortium

# The main pillars

Safeguard and foster the fundamental elements of human rights and democratic values by the utilization of AI and Big Data.

Build upon applicable laws, protocols, and regulations (i.e., GDPR, AI ACT), and in-line with ethical recommendations for AI, e.g., the recommendations of the HLEG

List, explore and remove bias in AI and Big Data by design

Provide enhanced policies that are easy-to-use, transparent, trustworthy, discrimination-free and unbiased.

Raise the awareness and trust of stakeholders on AI by reassuring them that is leveraged responsibly and ethically.

AI & Citizens

protection of human rights & EU values

Ethical AI

AI4GOV

engagement

regulations

AI Applications

# Challenges & Problems

From the use of AI

The use of AI in decision-making requires **transparency, trustworthiness and fainess** to avoid **bias, inequality** and **discrimination**.

Germany's Ruling Party Boss Warns Against Bias in AI Data Sets

Saskia Esken *Photographer: Tobias Schwarz/AFP/Getty Images*

By Aggi Cantrill
10 Μαΐου 2023 στις 12:33 π.μ. CEST

Germany's Social Democrat Leader Saskia Esken called for a closer examination of the discrimination inherent within data sets used to train artificial intelligence models.

Bloomberg

## Artificial intelligence (AI)

# AI poses existential threat and risk to health of millions, experts warn

BMJ Global Health article calls for halt to 'development of self-improving artificial general intelligence' until regulation in place

**Andrew Gregory** and **Alex Hern**

Wed 10 May 2023 00.01 BST

The Guardian

One of the examples of health harm cited by experts involved the use of AI-driven pulse

## Manhattan Institute

Policy Areas    Explore Content    Get Involved    About

View all Articles

**CULTURE, ECONOMICS**
Artificial Intelligence, Technology

March 14th, 2023
**Issue Brief** by David Rozado

13 Minute Read
Share

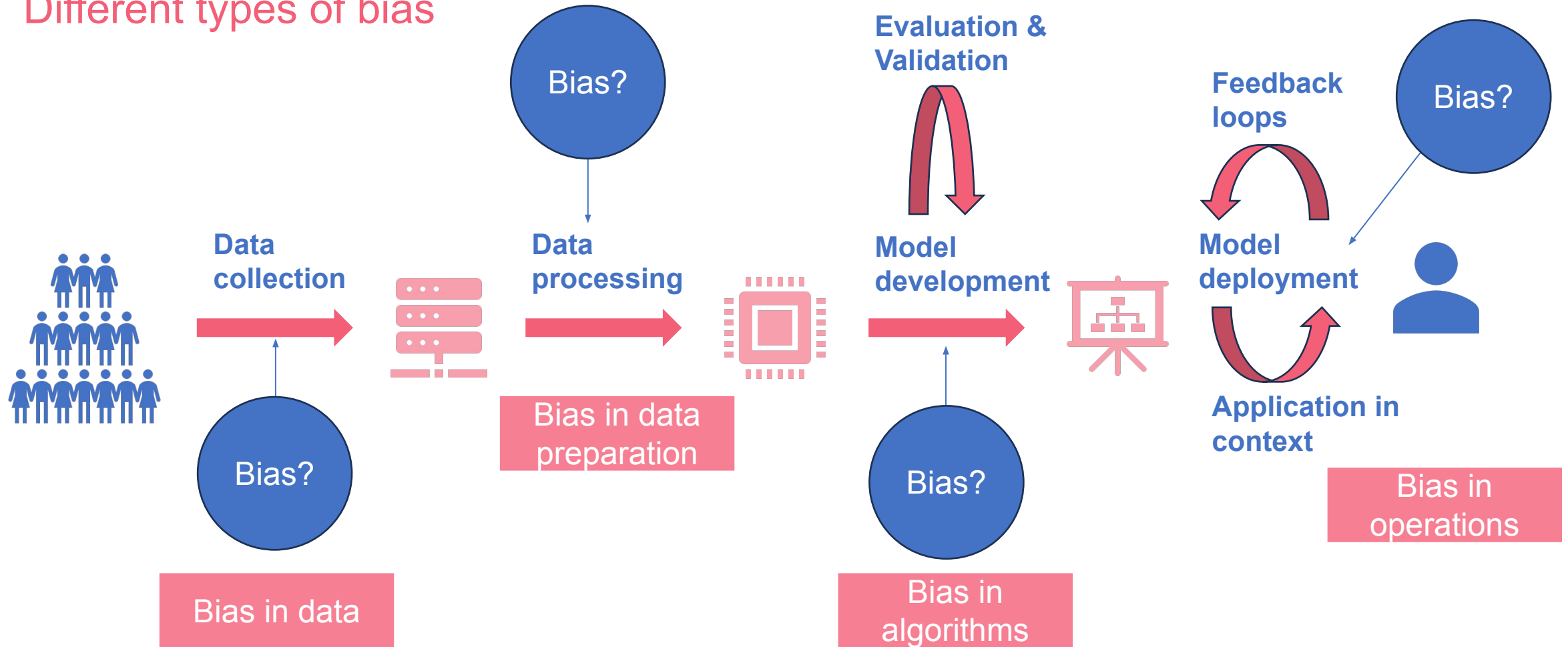# Danger in the Machine: The Perils of Political and Demographic Biases Embedded in AI Systems

## BROOKINGS

**COMMENTARY · TECHTANK**
**The politics of AI: ChatGPT and political bias**
Jeremy Baum and John Villasenor · Monday, May 8, 2023

# Bias in Data and AI lifecycle

## Different types of bias



**Bias?**

**Data collection**

**Data processing**

**Evaluation & Validation**

**Model development**

**Feedback loops**

**Bias?**

**Model deployment**

**Application in context**

Bias in data preparation

**Bias?**

Bias in operations

**Bias?**

Bias in data

Bias in algorithms

# Ethical aspects of AI

## And how to mitigate them

**Regulatory Frameworks**
Develop regulatory frameworks that will protect citizens' rights and ensure transparency.

**01**

**02**

**Personal Data Protection**
Need for strict measures to protect individuals' personal data

**TRUSTED AND TRANSPARENT AI**

**Human in the loop**
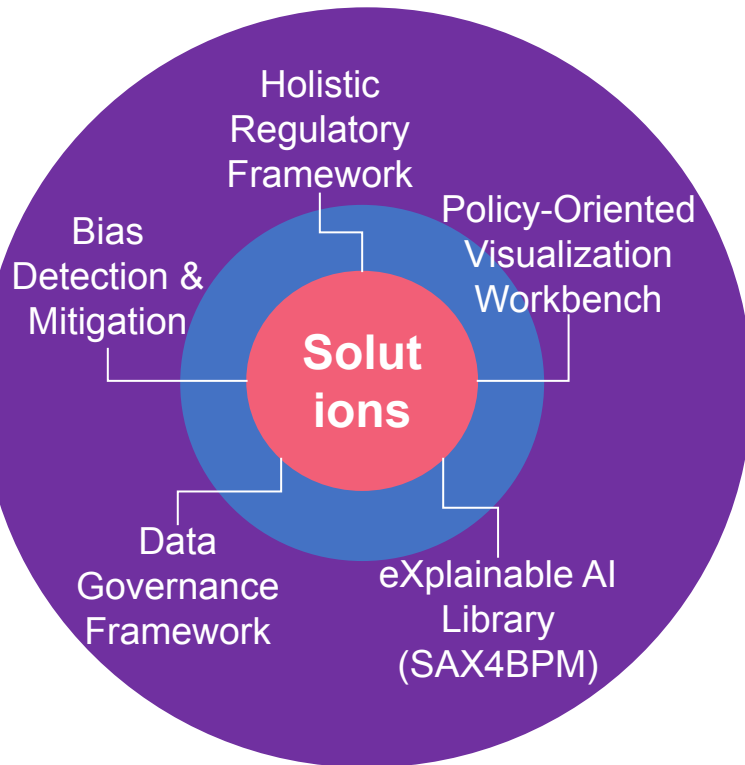Fostering integration between AI and human for optimal results.

**04**

**03**

**Transparency in Algorithms**
Provide transparent, explainable and unbiased AI-based solutions in the decision-making.

Webinar, October 1st

8

# The AI4Gov Project

## And the introduction of Trusted AI in the decision-making



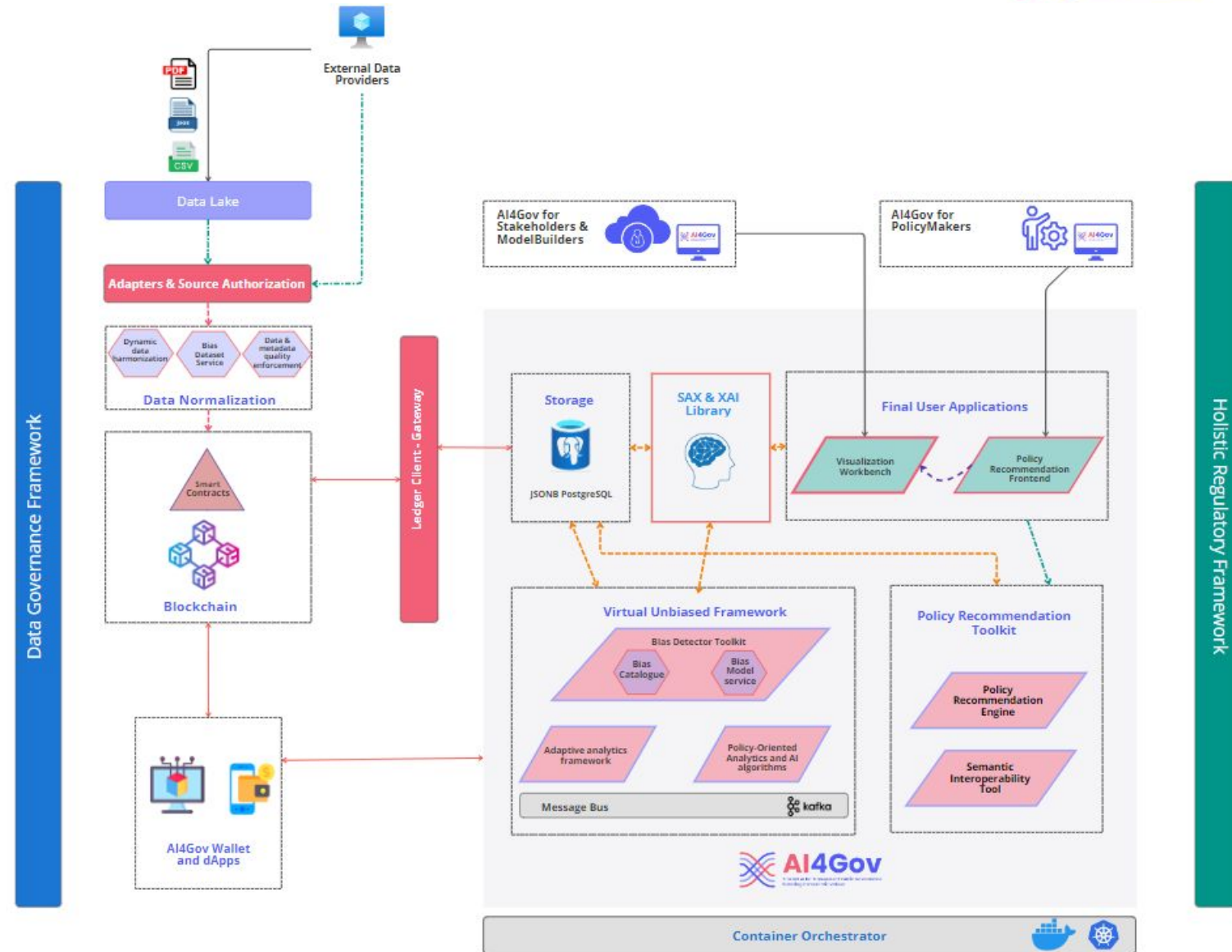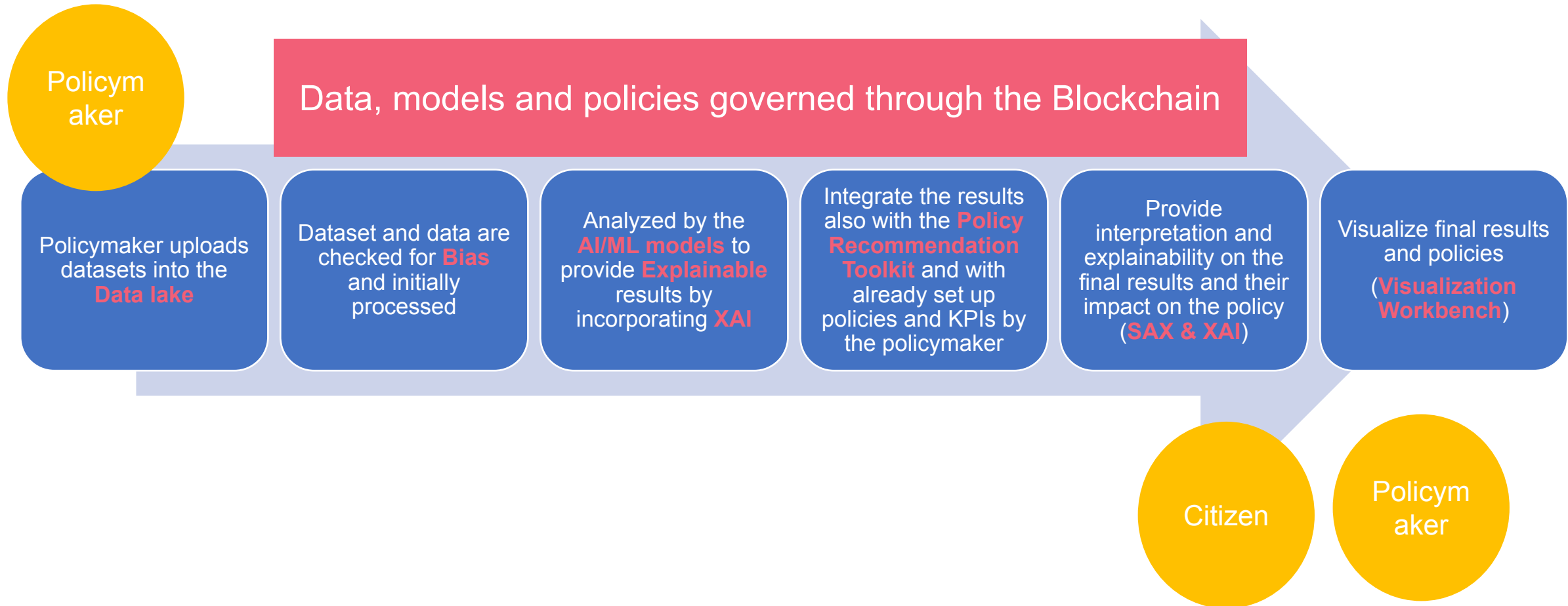| Research on AI and the rule of law | **Research on the effectiveness of monitoring and control protocols** of existing legislation and non-regulatory frameworks related to AI (e.g. ALTAI, EU AI ACT) |
|---|---|
| Fair & Trusted AI towards Addressing Bias | AI4Gov seeks to introduce a **set of tools and solutions covering the issue of bias** in IT (Bias Detector Toolkit, Holistic Regulatory Framework (HRF) and Virtual Unbiased Framework (VUF)) |
| Reliable, transparent and interpretable AI-based solutions | Improving **reliability, fairness and explanations by allowing people to reflect on the results of AI-based models** (eXplainable AI (XAI) & Situation-Aware Explainability (SAX)) |
| Identifying and Modeling Bias and Discrimination | Methods and techniques for **detecting and mitigating bias** (VUF, Bias Detector). Building an **XAI Library** to provide improved explainability. |
| AI-based solutions to enhance citizen representation and participation | Developing **hybrid citizen-centered education approaches** and using AI techniques to facilitate interaction between citizens and organizations |

# AI4Gov Architecture

Integrated platform

# Integrated Utilization

**Policymaker**

**Data, models and policies governed through the Blockchain**

Policymaker uploads datasets into the **Data lake**

Dataset and data are checked for **Bias** and initially processed

Analyzed by the **AI/ML models** to provide **Explainable** results by incorporating **XAI**

Integrate the results also with the **Policy Recommendation Toolkit** and with already set up policies and KPIs by the policymaker

Provide interpretation and explainability on the final results and their impact on the policy (**SAX & XAI**)

Visualize final results and policies (**Visualization Workbench**)

**Citizen**

**Policymaker**

Webinar, October 1st

# Citizens' Benefits from AI4Gov in Public services

Transparency

Increased trust and engagement

Access and inclusive participation

Training and Professional Development of public servants

Service and Process Improvement in public institutions

Enhanced Accessibility and Inclusivity

Webinar, October 1st

If not now, then when?

# Let's catch up!!!

# THANK YOU FOR YOUR ATTENTION!

Webinar, October 1st