1st Training Workshop: Bias in AI

Ljubljana, Slovenia – 16/05

AI4Gov

Trusted AI for Transparent Public Governance
fostering Democratic Values

AI and bias (bias in algorithms, bias in data)

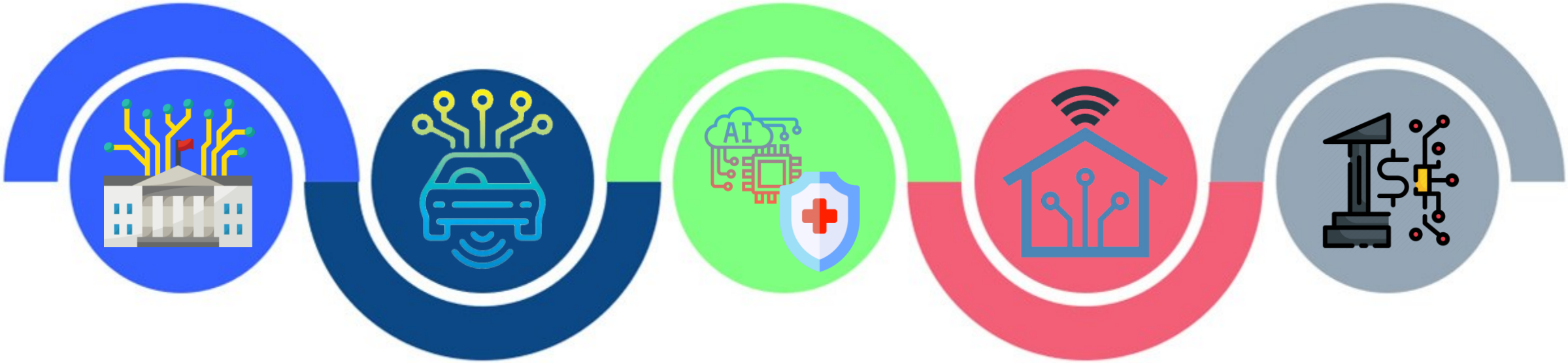
Presenter: George Manias (gmanias@unipi.gr)
Department of Digital Systems, University of Piraeus, Piraeus, Greece

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

OUTLINE

- AI in modern societies
- What is AI bias?
- Examples of AI bias in real world
- What are the types of bias in AI?
- Impact of bias in AI
- Common forms of bias in AI
- Widely used tools to mitigate bias
- How to better tackle bias in AI?
- How AI4Gov identifies and removes bias in AI?
- How AI4Gov creates Responsible and Trustworthy AI?

AI IN MODERN SOCIETIES




AI
~~Data~~ makes the world go around...
...BUT

THREATS OF AI

Germany's Ruling Party Boss Warn 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 oximeters. Photograph: Grace Cary/Getty Images

BROOKINGS

COMMENTARY - TECHTANK

The politics of AI: ChatGPT and political bias

Jeremy Baum and John Villasenor · Monday, May 8, 2023

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

EURACTIV

The Capitals | The Brief | Ukraine | Intelligence

Agrifood | Economy | Energy & Environment | Global Europe | Health | Politics | Technology | Transport

AI Act moves ahead in EU Parliament with key committee vote

By Luca Bertuzzi | EURACTIV.com · Est. 6min · 9:33 (updated: 11:13)

MEP Brando Benifei during the parliamentary committee vote on the AI Act. [European Parliament]

EURACTIV is part of the Trust Project >>>

Print

The European Parliament's leading parliamentary committees have green-lighted the AI Act in a vote on Thursday (11 May), paving the way for plenary adoption in mid-June.

Supporter

AI4TRUST

AI and bias (

WHAT IS AI BIAS?

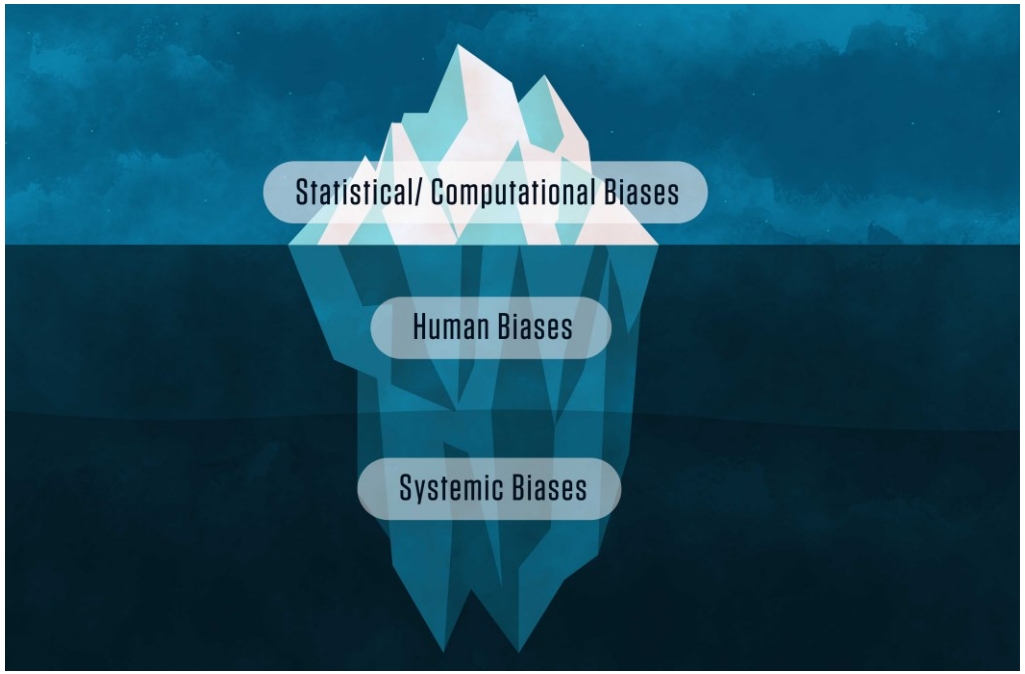
AI bias is an anomaly in the output of AI algorithms. This could be due to the prejudiced assumptions made during the algorithm development process or prejudices in the training data.

EXAMPLES OF AI BIAS IN REAL WORLD

- Racism embedded in US healthcare (due to features/variables used)
- Amazon's Recruiting Engine gender bias (due to training data)
- COMPAS algorithm race bias with reoffending rates (due to training data and algorithm process)
- Google Photos Algorithm race bias (due to features/variables used)
- PredPol Algorithm biased against minorities (due to training data used and human bias)

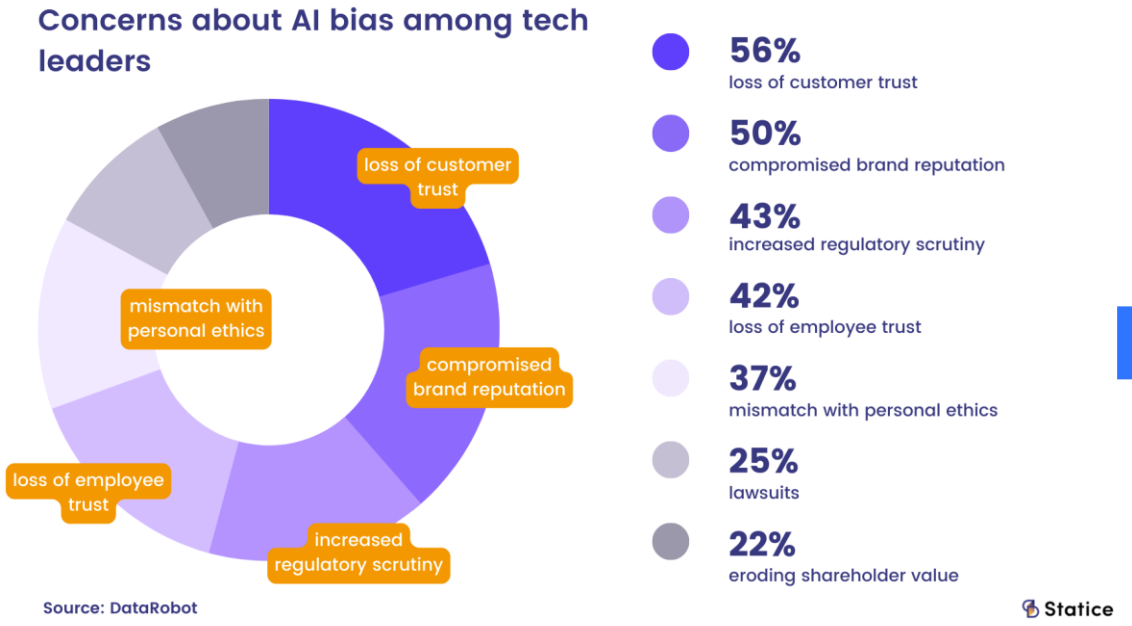
WHAT ARE THE TYPES OF BIAS IN AI?

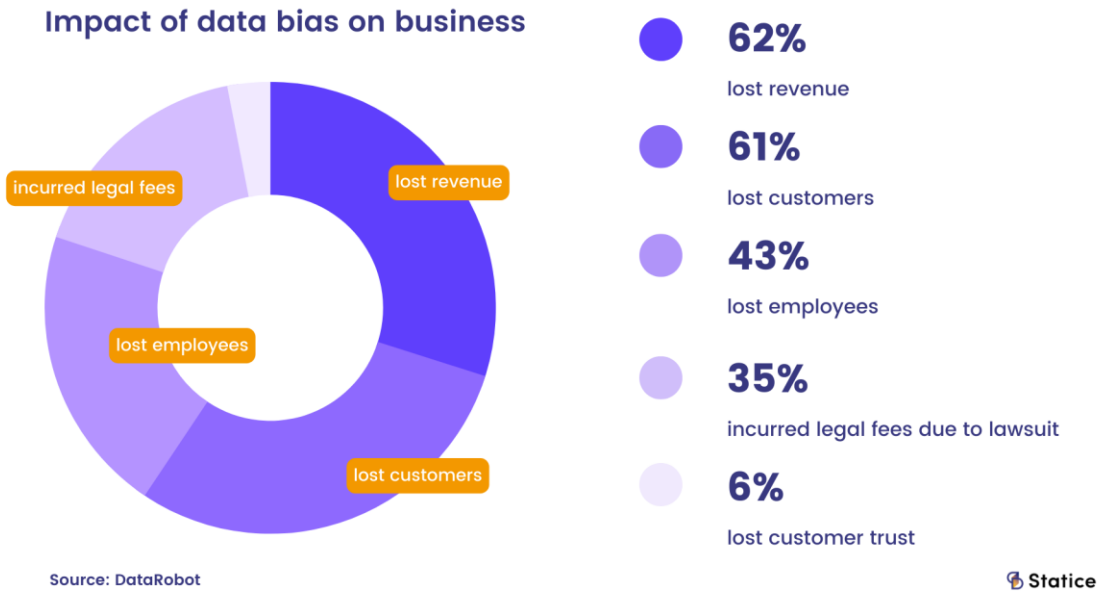
- Statistical/Computational Bias
  - Stem from the datasets and algorithmic processes used (e.g., heterogeneous data, undersampling in data, algorithmic biases such as over- and under-fitting, erroneous treatment of outliers, selection of features and data cleaning)
- Human Bias
  - Present in the institutional, group, and individual decision-making processes across the AI lifecycle and in the use of AI applications once deployed (e.g., social assumptions and norms create blind spots and expectations in thinking)
- Systemic Bias
  - Historical, societal, institutional (e.g., institutional racism and sexism, smart-homes that exclude people with disabilities)



* Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). Towards a standard for identifying and managing bias in artificial intelligence. NIST Special Publication, 1270, 1-77.

IMPACT OF BIAS IN TECH LANDSCAPE

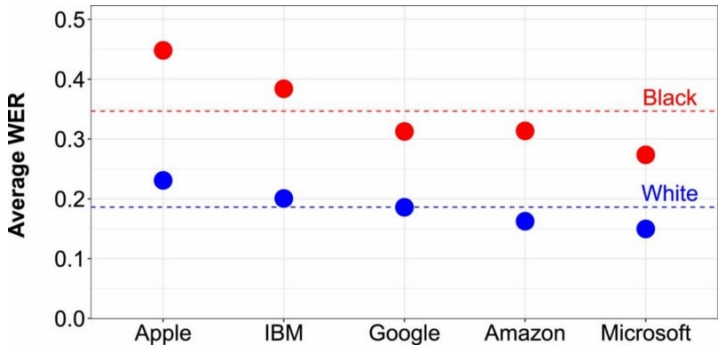


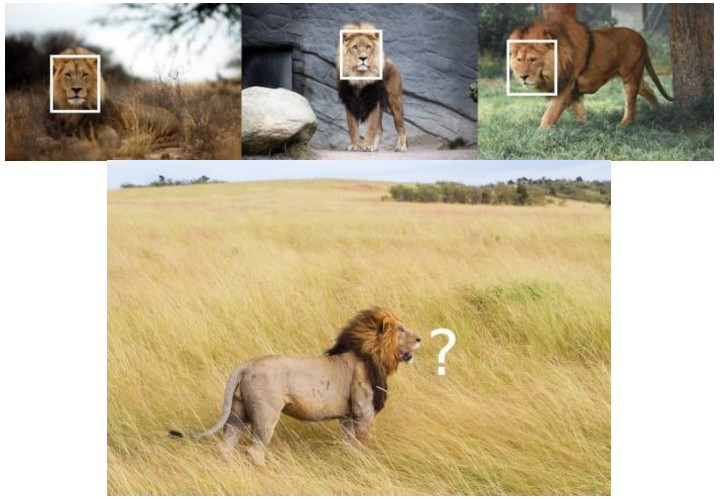


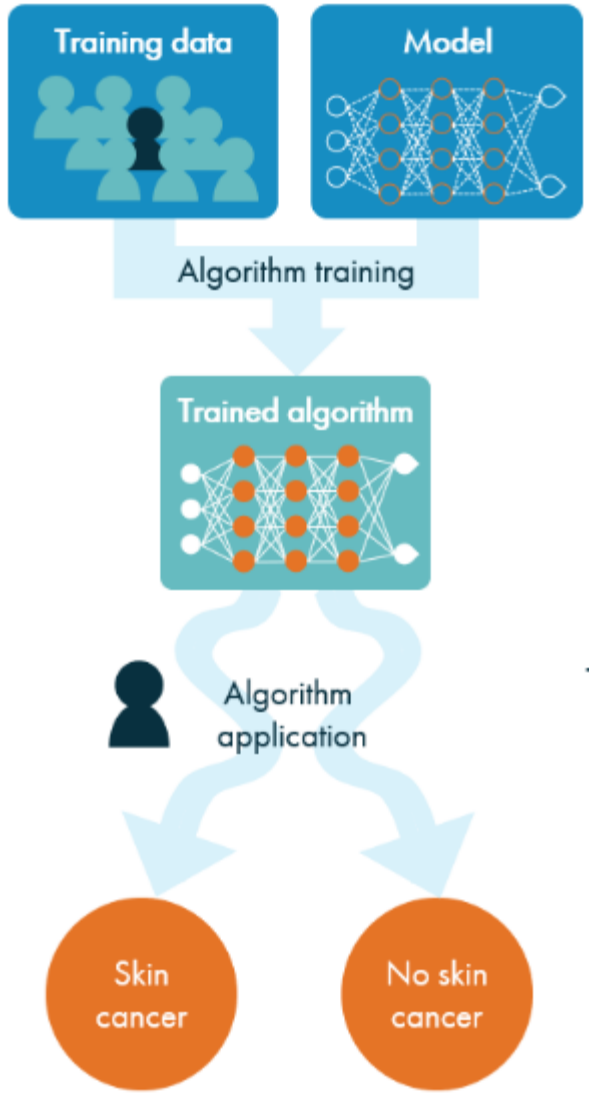
* "State of AI Bias" report, DataRobot, 2022

COMMON FORMS OF BIAS IN AI

- Historical Bias
- Sample/Selection Bias
- Feature/Label Bias
- Aggregation Bias
- Confirmation Bias
- Interaction bias
- Evaluation Bias







WIDELY USED TOOLS TO MITIGATE BIAS IN AI

- Google's "What-If Tool"
- IBM's AI Fairness 360
- Academic fairness-in-AI projects are under progress funded by the National Science Foundation (NSF) in collaboration with Amazon.
- Facebook AI is using a new "radioactive data" technique to detect if a dataset that was used to train a ML model underlines bias.
- audit-AI Toolkit for bias testing in ML applications
- FairLens
- PwC's Responsible AI
- …

HOW TO BETTER TACKLE BIAS IN AI?

The challenge of bias in AI is complex and multi-faceted. While there are many approaches and tools for mitigating this challenge there is no quick fix.

Employ human-in-the-loop technology for constant improvement

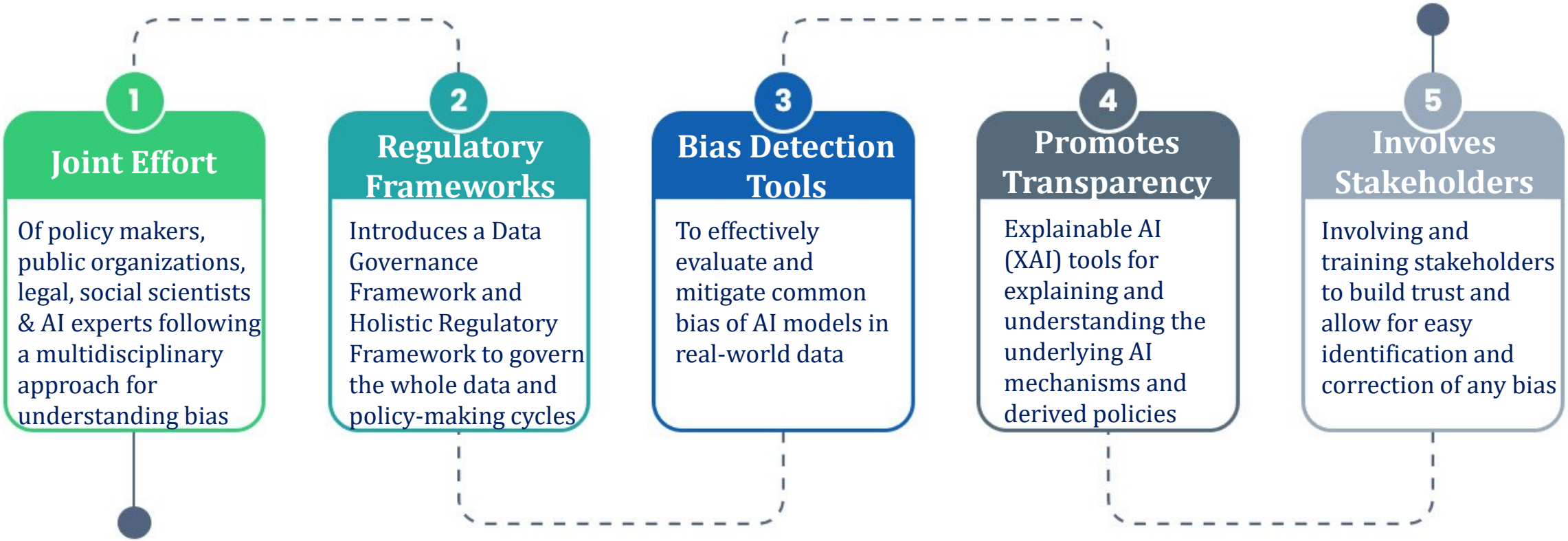
Attention on data and algorithms for constant feedback

- Underlying data
- User-generated data
- Deploying suitable tools to identify bias and inaccuracies

Diverse and inclusive team of experts and extensive research in bias

Algorithm test in a real-world setting

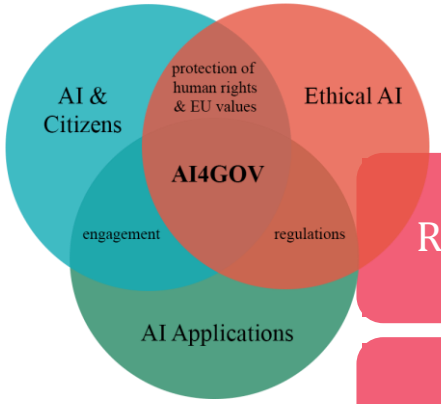
HOW AI4GOV MITIGATES BIAS?



In-line with applicable laws, protocols, and regulations (i.e., the GDPR), but also with ethical recommendations for AI, e.g., the recommendations of the HLEG and the EU AI Act.

Raising the awareness of stakeholders will reassure them that AI is being leveraged responsibly and ethically

HOW AI4GOV CREATES RESPONSIBLE AND TRUSTWORTHY AI?

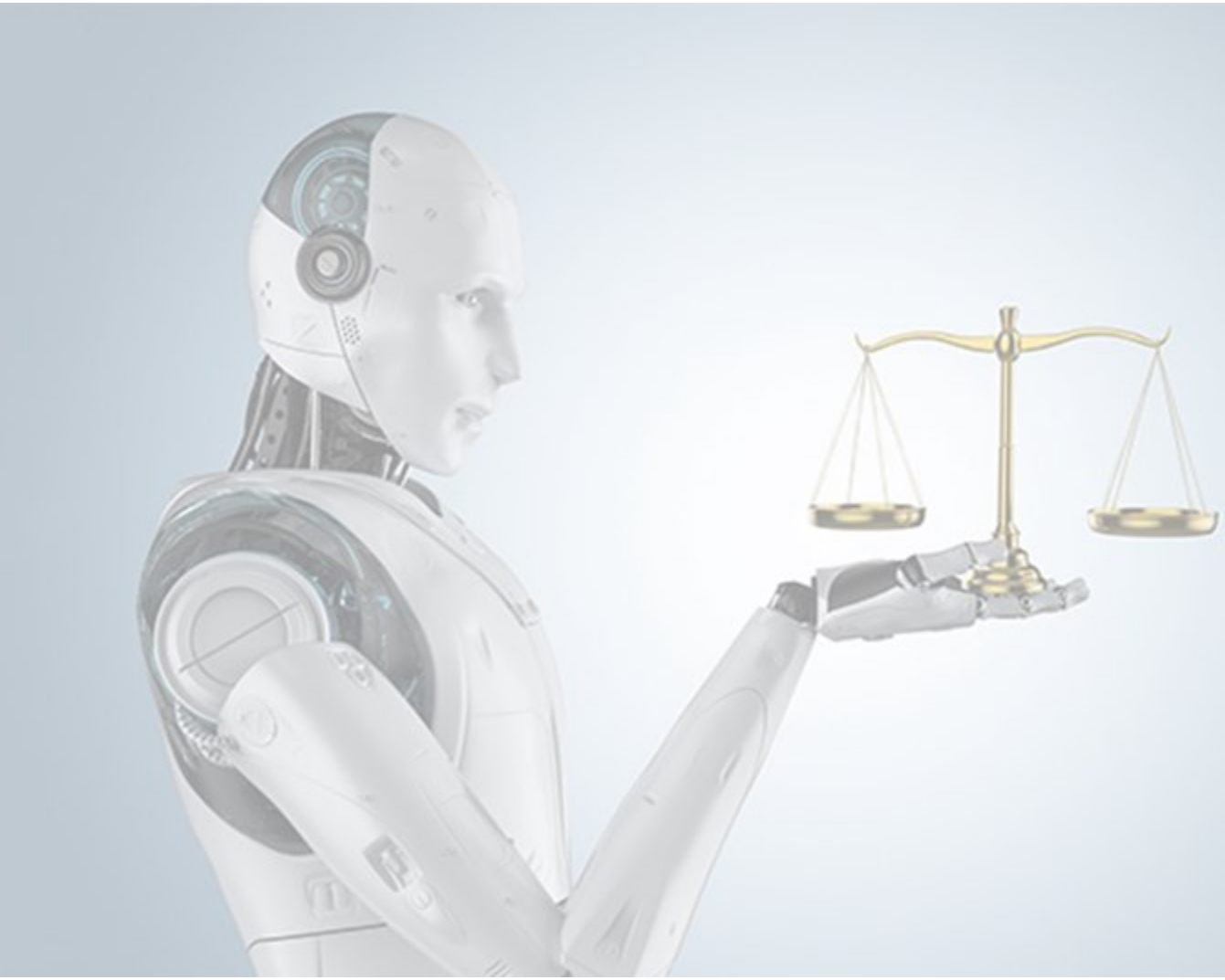


Research on AI and Rule of Law	Examine the effectiveness of monitoring and control protocols of established legislation and non-regulatory measures over AI and Big Data (e.g., **ALTAI, EU AI ACT**)
AI Fairness and Bias Mitigation	AI4Gov seeks to introduce a set of tools and solutions covering the issue of bias in AI across the complete AI chain. (**Bias Detector Toolkit, HRF and Virtual Unbiased Framework**)
Trusted, Transparent, and Interpretable AI solutions	Enhance trustworthiness, fairness and explainability, by enabling humans to reason about the outcomes of AI-based models (**eXplainable AI (XAI)** & **Situation-Aware Explainability (SAX)**)
Identification and Modelling of Bias and Discrimination and their Sources	Methods and techniques to identify and mitigate bias (**VUF, Bias Detector**). **XAI Toolkit** will be combined with VUF to provide bias removal recommendations.
AI applications in the context of Citizen Engagement and Participation	Establish hybrid **citizen-centric training approaches** and utilize AI techniques to facilitate the interaction between citizens and organizations

AI and bias (bias in algorithms, bias in data)

AI4Gov
Trusted AI for Transparent Public Governance fostering Democratic Values

If not now, then when?



THANK YOU FOR YOUR ATTENTION