



# AI Fairness

Marcin DETYNIECKI, Boris RUF

15 December 2022



L'allégorie de la justice - musée de l'Hospice Comtesse, à Lille.



# 1

A Key Principle of Responsible AI: Fairness

# Responsible AI Principles

« Responsible AI is a **standard for ensuring that AI is safe, trustworthy and unbiased.**

Responsible AI ensures that AI and machine learning (ML) models are **Robust, Explainable, Ethical and Efficient.** » FICO


Several organizations have published AI principles based on values / ethics ...


## OECD'S AI PRINCIPLES

 Inclusive growth, sustainable development and well-being >

 Human-centred values and fairness >

 Transparency and explainability >

 Robustness, security and safety >

 Accountability >

## EU GUIDELINES FOR TRUSTWORTHY AI

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental well-being
- Accountability

## FEAT PRINCIPLES for RESPONSIBLE AI (Monetary Authority of Singapore)

- Fairness (Justifiability, Accuracy and Bias)
- Ethics
- Accountability (Internal and External)
- Transparency

... driving AXA's definition

## AXA'S RESPONSIBLE AI PRINCIPLES

Responsible AI Principle
Technical robustness and safety
Diversity, non-discrimination and fairness
Transparency / interpretability
Privacy and data governance
Human agency and oversight



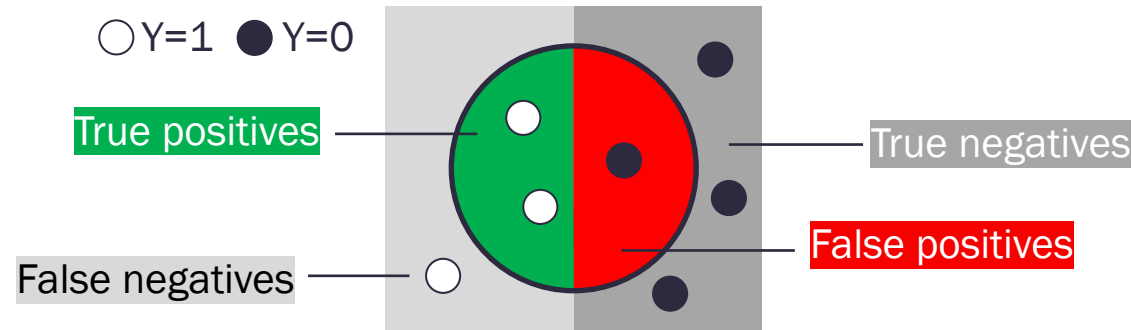
# 2

Insights from R&D

# Unwanted Bias in AI

- Bias: Algorithm performs differently for sensitive sub groups
- Sources are different and less obvious compared to conventional algorithms
  - Data collection (historical bias)
  - Sampling (representation bias)
  - Measurement (measurement bias)
  - Model learning (learning bias)
  - Benchmarking (evaluation bias)
  - Human interpretation (deployment bias)
  - ...
- If not controlled for, bias can get reproduced at scale without being noticed
- Research community has proposed plenty of fairness metrics and bias mitigation methods

# Example: Two conflicting fairness metrics



## Equalized Odds

Metrics: false positive and false negative rates

$$FPR = \frac{FP}{N} \quad FNR = \frac{FN}{P}$$

Rationale: Based on the **true outcome**, the proportion of correct decisions should be equal across all groups.

## Conditional Use Accuracy Equality

Metrics: false discovery and false omission rates

$$FDR = \frac{FP}{TP+FP} \quad FOR = \frac{FN}{TN+FN}$$

Rationale: Based on the **predictions**, the proportion of correct decisions should be equal across all groups.

**Calibration**

**Conditional statistical parity**

**Balance for negative class**

**Predictive equality**

**Equalized opportunities**

**Equal selection parity**



**Conditional use accuracy equality**

**Equalized odds**

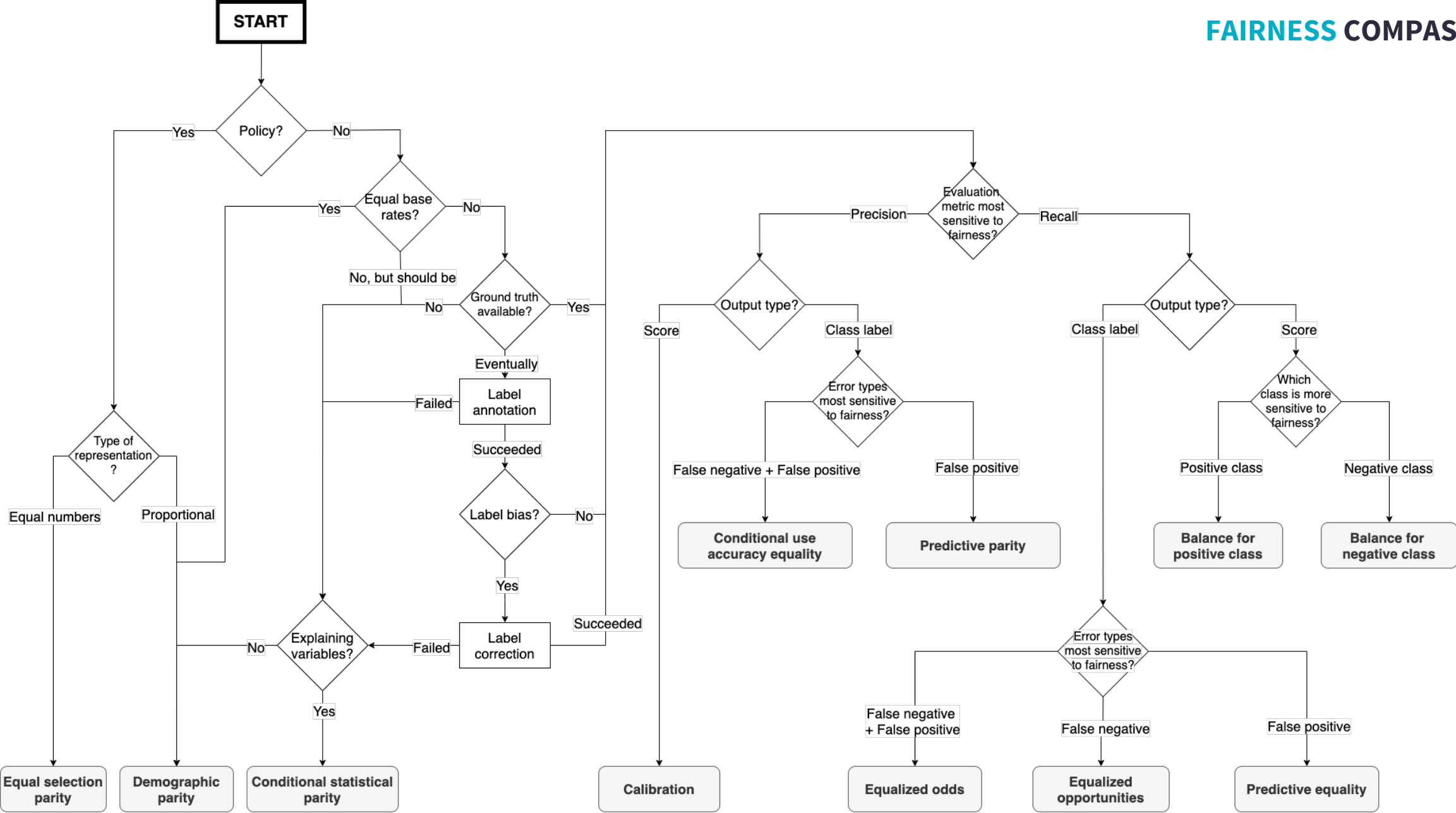
**Predictive parity**

**Balance for positive class**

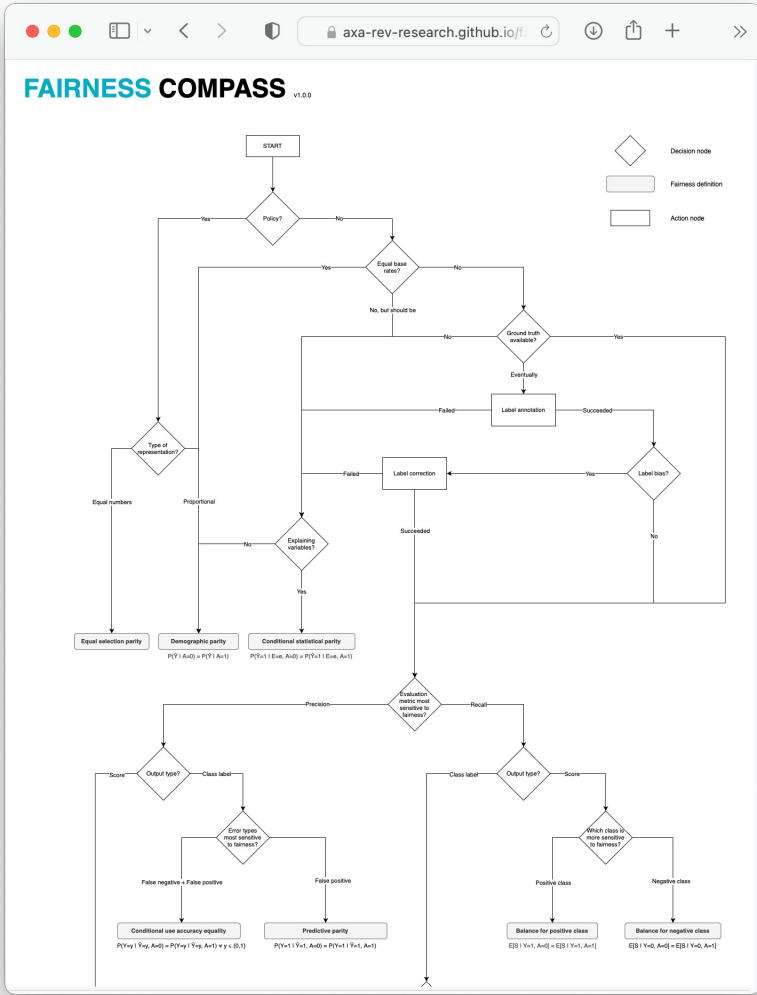
**Demographic parity**

...

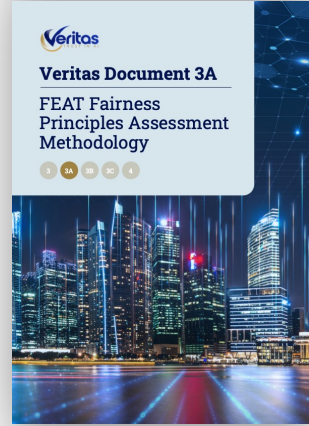
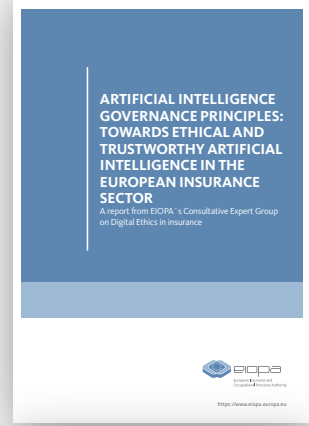




# Web Application and Booklet



Available on GitHub: <https://github.com/axa-rev-research/fairness-compass>



# Future Work for R&D

- Sensitive attributes are missing in practice
  - General Data Protection Regulation (GDPR) prohibits the collection and the processing of sensitive personal attributes in many cases
  - “Fairness by unawareness” insufficient due to many correlations in large datasets
- Existing limitations of research proposals
  - Continuous sensitive attributes (e.g., age)
  - Regression problems (e.g., insurance pricing)
- Other challenges
  - Intersectional group fairness
  - Group fairness vs. individual fairness



# From R&D to Practice: AI Governance

# AXA's Responsible AI Circle (RAIC)

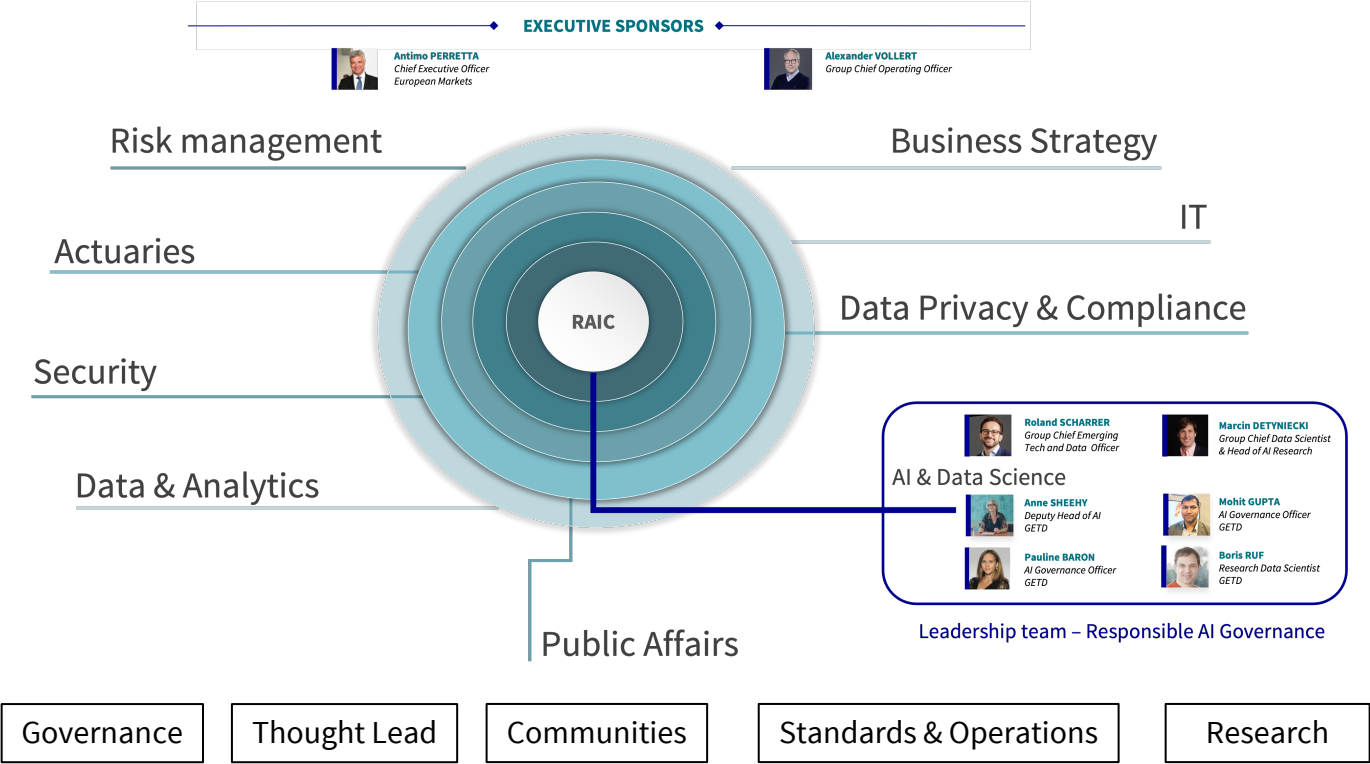
## Purpose

In Jan. 2021, AXA launched the Responsible AI Circle, a **light and agile multi-stakeholder** (Group & entities) **governance body**.

The Circle is in charge of **overseeing the Responsible adoption of AI** within the Group.

**AXA's Responsible AI Circle gathers 25+ permanent members as of June 2022**

**RAIC mission statement**



- 1 Give directions towards trustworthy, performant & delivering value AI
- 2 Accelerate the Responsible adoption of AI for business, in core activities
- 3 Apprehend & contribute to the evolution of AI regulation
- 4 Strengthen AXA's position as a responsible insurer
- 5 Break siloes and simplifying our AI governance





# 4

## Concluding Remarks

# Key Takeaways

- There is no one-fits-all solution for AI fairness, the best solution depends on the context of use case
- Assessing and mitigating unwanted biases without the sensitive attribute is hard
- For now, a process-driven approach with human oversight (AI governance) is the best practice available
- We need to continue to invest in research to ensure a robust/sustainable implementation of Trustworthy AI – an opportunity for a better world



# AI Fairness

[Marcin.DETYNIECKI@axa.com](mailto:Marcin.DETYNIECKI@axa.com)

[Boris.RUF@axa.com](mailto:Boris.RUF@axa.com)

15 December 2022