# Can AI truly be free of bias, or is it a reflection of the world we create?

Clara Rus

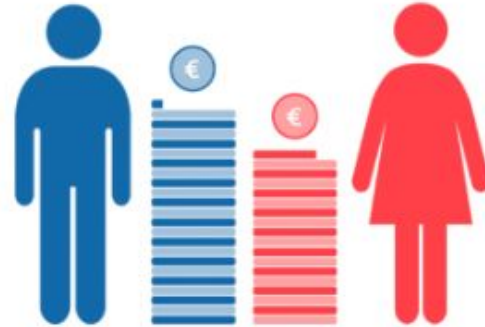**Annie Romein-Verschoorlezing 2025**

# Why is AI biased?



Under-representation

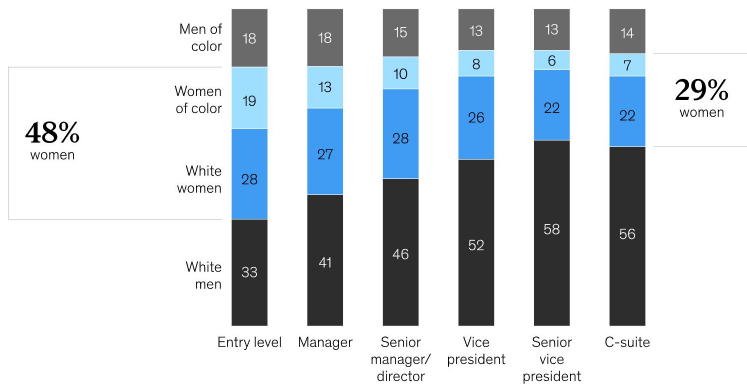Women being underrepresented in the dataset

Historical Bias

Women being associated with low income in the dataset

# Is it a reflection of our society?

## Women remain underrepresented at every stage of the corporate pipeline, regardless of race and ethnicity.

**Representation in corporate role, by gender and race,** % of employees



**48%** women

**29%** women

| | Entry level | Manager | Senior manager/ director | Vice president | Senior vice president | C-suite |
|---|---|---|---|---|---|---|
| Men of color | 18 | 18 | 15 | 13 | 13 | 14 |
| Women of color | 19 | 13 | 10 | 8 | 6 | 7 |
| White women | 28 | 27 | 28 | 26 | 22 | 22 |
| White men | 33 | 41 | 46 | 52 | 58 | 56 |

| | Entry level | Manager | Senior manager/ director | Vice president | Senior vice president | C-suite |
|---|---|---|---|---|---|---|
| Total women, 2024 | 48 | 39 | 37 | 34 | 29 | 29 |
| Total women, 2020 | 47 | 38 | 33 | 29 | 28 | 21 |
| Total women, 2015 | 45 | 37 | 32 | 27 | 23 | 17 |

Note: Figures may not sum to 100%, because of rounding. Total percent of women per level may not sum to overall corporate pipeline totals, because overall figures do not include employees with unreported race data.
Source: *Women in the Workplace 2024*, McKinsey & Company and LeanIn.Org

McKinsey & Company

---

**The unadjusted gender pay gap, 2022**
(difference between average gross hourly earnings of male and female employees as % of male gross earnings)
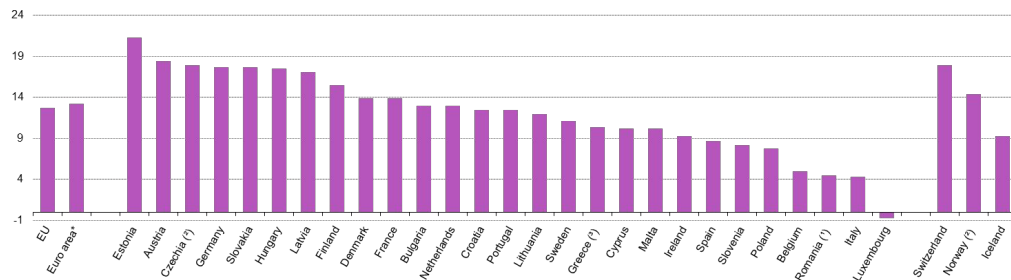


Note: For all the countries except Czechia and Iceland: data for enterprises employing 10 or more employees, NACE Rev. 2 B to S (-O); Czechia: data for enterprises employing 1 or more employees, NACE Rev. 2 B to S; Iceland: NACE Rev. 2 sections C to H, J, K, P, Q. Gender pay gap data for 2022 are provisional until benchmark figures, taken from the Structure of Earnings survey, become available in December 2024.
\* Euro area (2015-2022)
(¹) Estimated data.
(²) Definition differs (see metadata).
(³) 2018 data.
*Source*: Eurostat (online data code: sdg_05_20)

eurostat

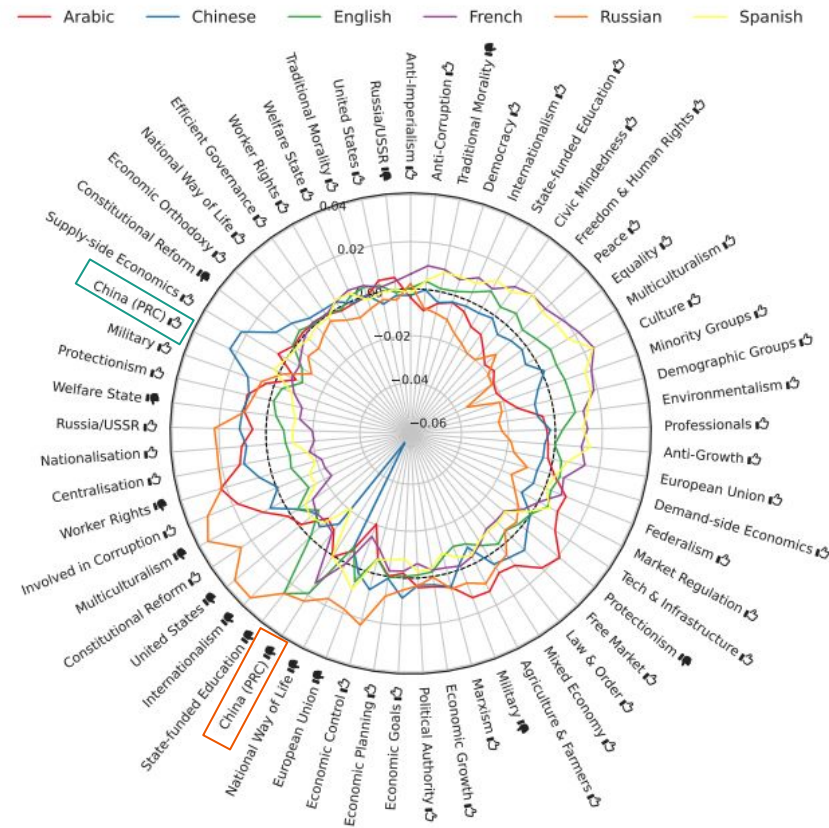## Such imbalances are also present in our society

# Is it a reflection of our society?

- AI is overwhelmingly shaped by men → affects how AI systems are designed, trained, and deployed.

- It reflects the values and perspectives of those who create it and of the world we live in → perpetuating the biases and inequalities of our society.

- Aim at having more diversity in teams and leadership roles developing AI systems → enforce a more diverse perspective in the outputs of AI systems.
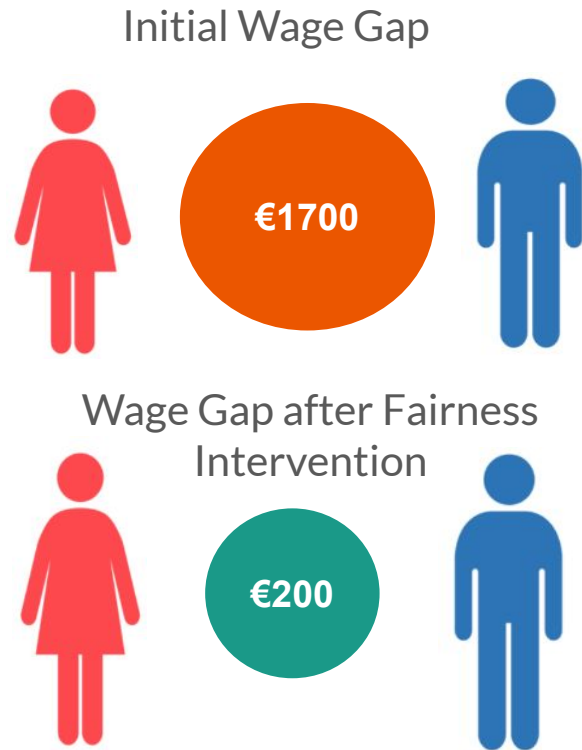
# Is it a reflection of our society?

- Chinese language (blue line)
  → close to China - thumbs up
  → far from China - thumbs down

- A country has a good opinion about itself and this is propagated in the LLMs responses.



Figure 3: Per ideology tag, the zero-centered average score in each UN language. Centering was done by subtracting the overall average score per tag, and the overall average score per language. The dotted line marks the average (zero) across languages.

# **What can we do?**

- Try to combat the problem!
  → fairness interventions
  (e.g. adversarial methods)

- At the job category level there was increased wage gap for some
  → still quite some work to be done on this area

- "Our AI model is free of bias"
  → dangerous statement

Initial Wage Gap

€1700

Wage Gap after Fairness Intervention

€200

Rus C, Luppes J, Oosterhuis H, Schoenmacker GH. Closing the gender wage gap: Adversarial fairness in job recommendation. HR@Recsys 2022 Sep 20.

# Intersectionality



- Individuals may face multiple layers of discrimination, such as gender combined with race, disability, or socioeconomic status.

- **Challenges:** lack of diverse datasets, granularity of groups.

Rus C, de Rijke M, Yates A. Counterfactual Representations for Intersectional Fair Ranking in Recruitment. InHR@ RecSys 2023

# Call for action!

- Developers must challenge the AI system and the potential biases.

- We need more diverse teams working in AI to favor inclusivity from the design and development stage.

- Policymakers should regulate AI ethics and hold corporations accountable.

- Interdisciplinary collaboration between policymakers, developers, researchers and diverse people.

- Advocate for transparency and for AI education.

# Final Remarks



- AI is a mirror of society.

- Change is possible!

- "What kind of AI future do we want to create?"