

Artificial Intelligence and Fundamental Rights

Dr. David Reichel

Luxembourg, 9 July 2024

Overview

- Policy developments
- Fundamental rights implications when using AI
- Biased algorithms and discrimination
- Safeguards, AI Act and work ahead

European policy developments on AI

- EU institutions active on AI regulation since 2017
 - High Level Expert Group on AI (2018-2020)
 - White paper on AI (2020)
 - AI Act proposed April 2021 – approved on 21 May 2024 – OJ 12 July 2024
- Council of Europe
 - Framework Convention on artificial intelligence and human rights, democracy, and the rule of law – adopted on 17 May 2024
- Other international organisations (OECD, UNESCO, ...)

FRA and AI

The need for a rights-based approach

- Extensive policy development on AI
- Ethics and rights-based approaches
- Fundamental and human rights are legal obligations

The need for an evidence-based approach

- Broad field of AI led to generic and abstract discussions
- Scattered knowledge and lack of evidence on fundamental rights
- Focus on concrete use cases to avoid abstract discussions

FRA reports

- Discrimination and algorithms (2018)
- Data quality and AI (2019)
- Facial recognition technology (2019)
- Getting the future right – Artificial intelligence and fundamental rights (2020)
- Bias in algorithms – Artificial intelligence and discrimination (2022)



What is Artificial Intelligence?

- Different views
- Definition AI-System (AI-Act, Art. 3 (1)):
“means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments;”
- Need for clarity

*“It is an umbrella term.”
(Private company, Estonia)*

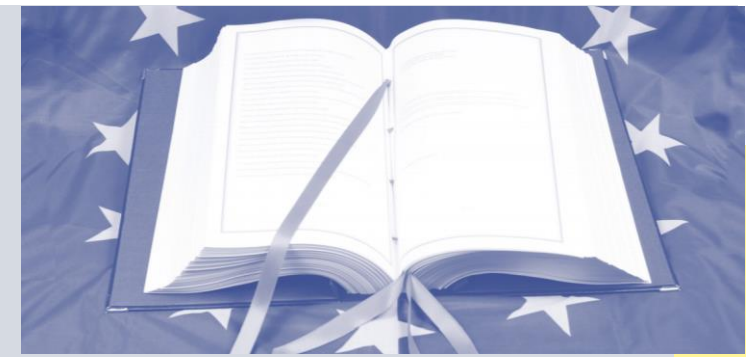
*“There is no lawyer who can tell you the definition of AI and we’ve asked around pretty thoroughly.”
(Public administration, Netherlands)*

Reasons for using AI

“It is mostly used to save time [...] when you have to go through a lot of material.”

“The most important is to deal with cases more efficiently. It’s about making use of your workforce, the people who handle cases, as effectively as possible.”

Fundamental rights framework



- Charter of Fundamental Rights of the EU
- Secondary EU law
 - data protection law
 - non-discrimination law
 - ...
- International treaties (Council of Europe and United Nations)
- National and sector specific laws

Data protection

- High level of awareness about data protection
- Personal data use – when it is used and isn't used
- Automated decision making and human review
- Uncertainty about the applicability of data protection

*“There is the GDPR but it does not give you specific rules. It gives principles but it comes down to ethical issues and interpretation.”
(Private company, Estonia)*

“There is a risk of having too much trust in the machine.” (Public administration, France)

Equality and non-discrimination

- Lack of in-depth analysis of non-discrimination
- Interviewees convinced that their system does not discriminate
- Question of proxies – information linked to protected characteristics
- Rights of special groups almost absent in discussions

“[I]f you do not have access to sensitive personal data, it is impossible to check if you are profiling on that basis.” (Public administration, Netherlands)

Access to effective remedy

- Awareness that people have the right to complain
- Not many complaints received
- Importance of transparency and explainability

*“The topic of transparency is very important nowadays, there are many procedures on how to publish the information, ...”
(Public administration, Spain)*

Other fundamental rights impacted

- Good administration
- Consumer protection
- Social security and assistance
- Freedom of expression
- Right to assembly
- Right to education
- Right to asylum
- Right to collective bargaining and action
- Right to fair and just working conditions
- Right to access preventive care
- ...

Impact assessments in practice

- Many calls for fundamental rights impact assessments
- Focus on technical aspects, not fundamental rights
- Those using AI are not always sure which bodies are responsible for overseeing AI systems

“When testing the system, we did not really look at the legal aspects, we looked at whether the system is profitable.” (Private company, Estonia)

“Yes, we assess the legality of personal data protection and the conformity with their specific legal acts.” (Public administration, Estonia)

The use of AI should ...

- process data legally
- not lead to unfair treatment
- be contestable
 - make people aware that AI is used
 - inform people where and how to complain
 - make sure that decisions can be explained

“We are proactive not only among ourselves to mitigate risks.” (Private company, Estonia)

Biased algorithms

Biased algorithms: what's the problem?

- Many examples of biased algorithms
- Lack of knowledge and empirical evidence
- Many ongoing policy efforts to prevent and mitigate harms

What's bias?

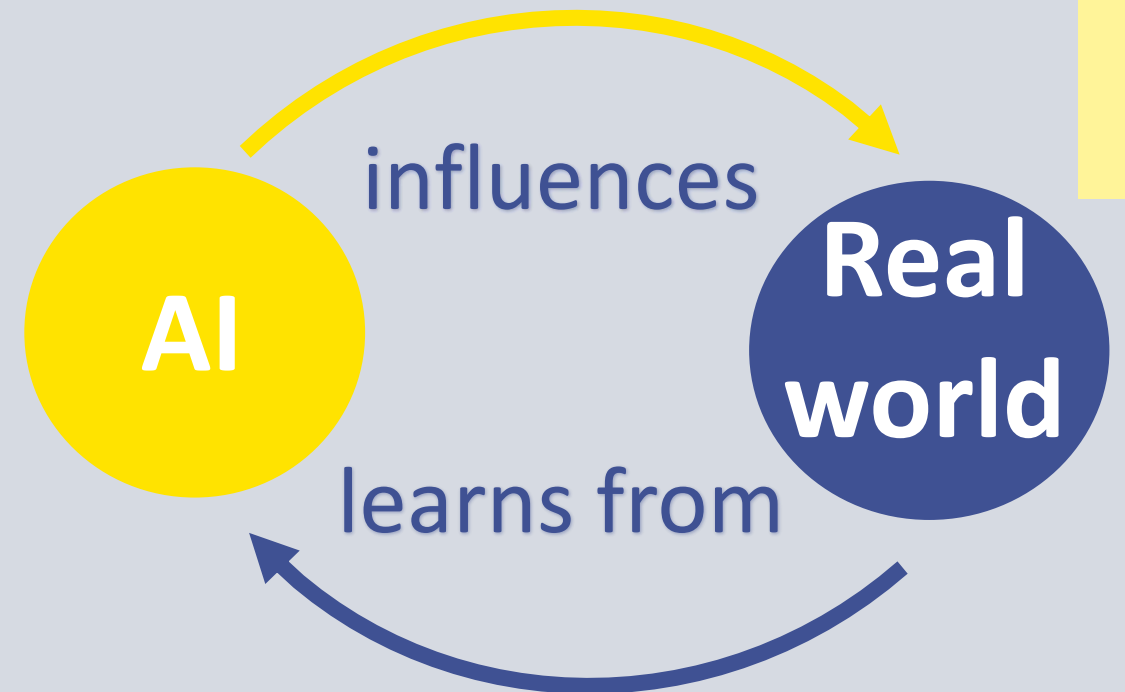
- Bias is defined differently
- Bias relates to differential treatment of people based on protected characteristics
- Bias occurs frequently and needs to be assessed in context

Feedback loops

Simulating predictive policing

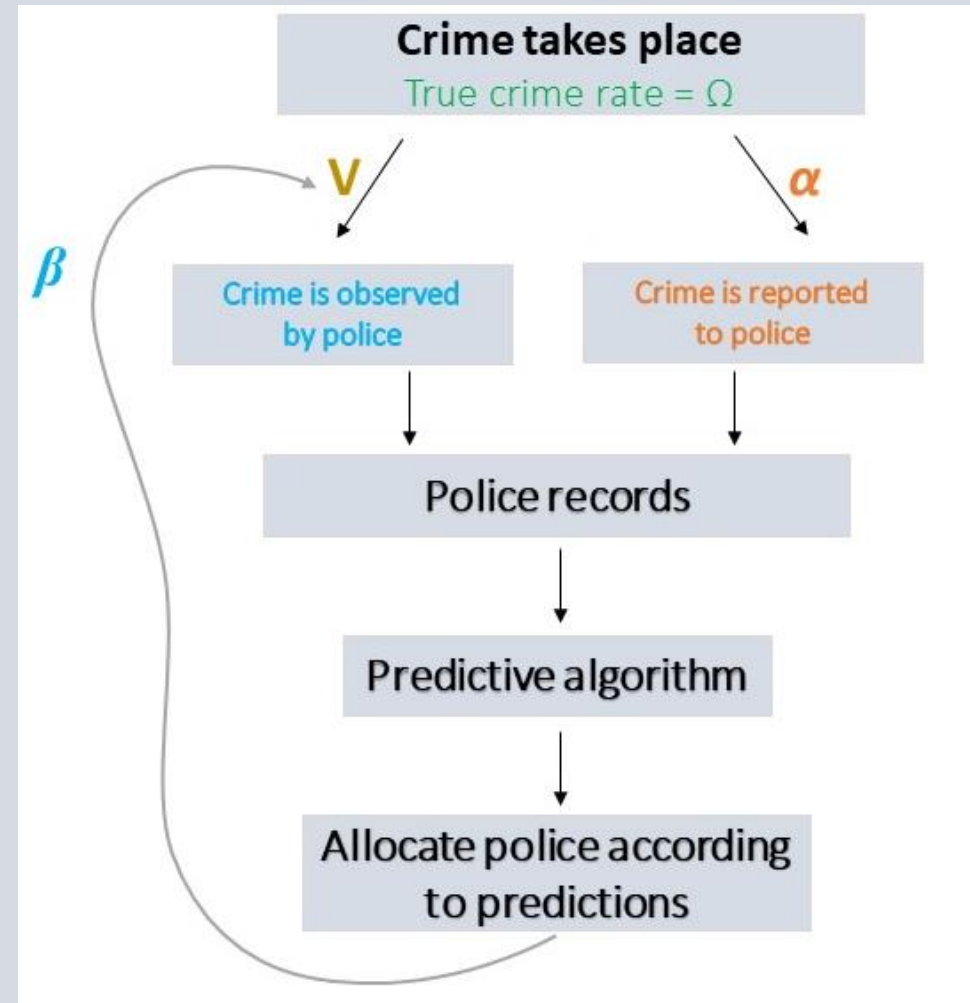
Feedback loops

- Algorithms influence algorithms, because they influence the reality on the ground
- Feedback loops can 'run away'



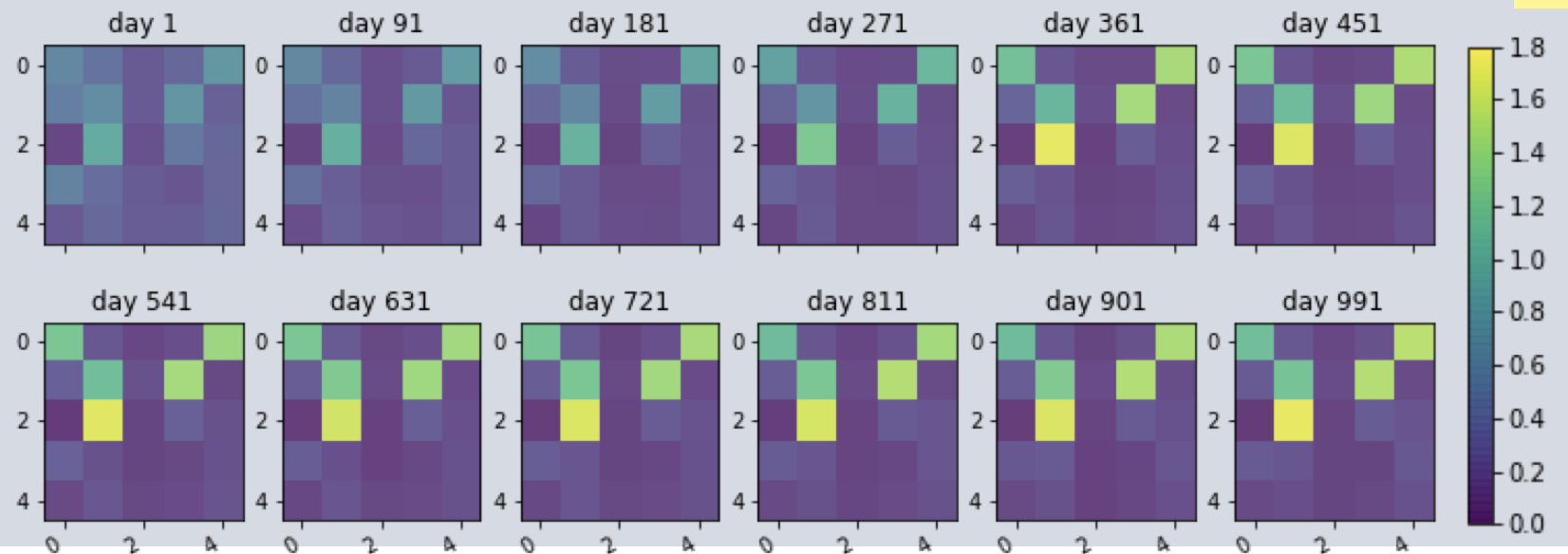
Simulating predictive policing

- Crime reporting rates (α)
- Observability of crime (V)
- True crime distribution (Ω)
- Police distribution (β)



Several sources of bias

- Machine learning model can introduce bias
- Crime reporting rates
- Crime observability



Ethnic and gender bias in offensive speech detection

Ethnic bias in speech detection

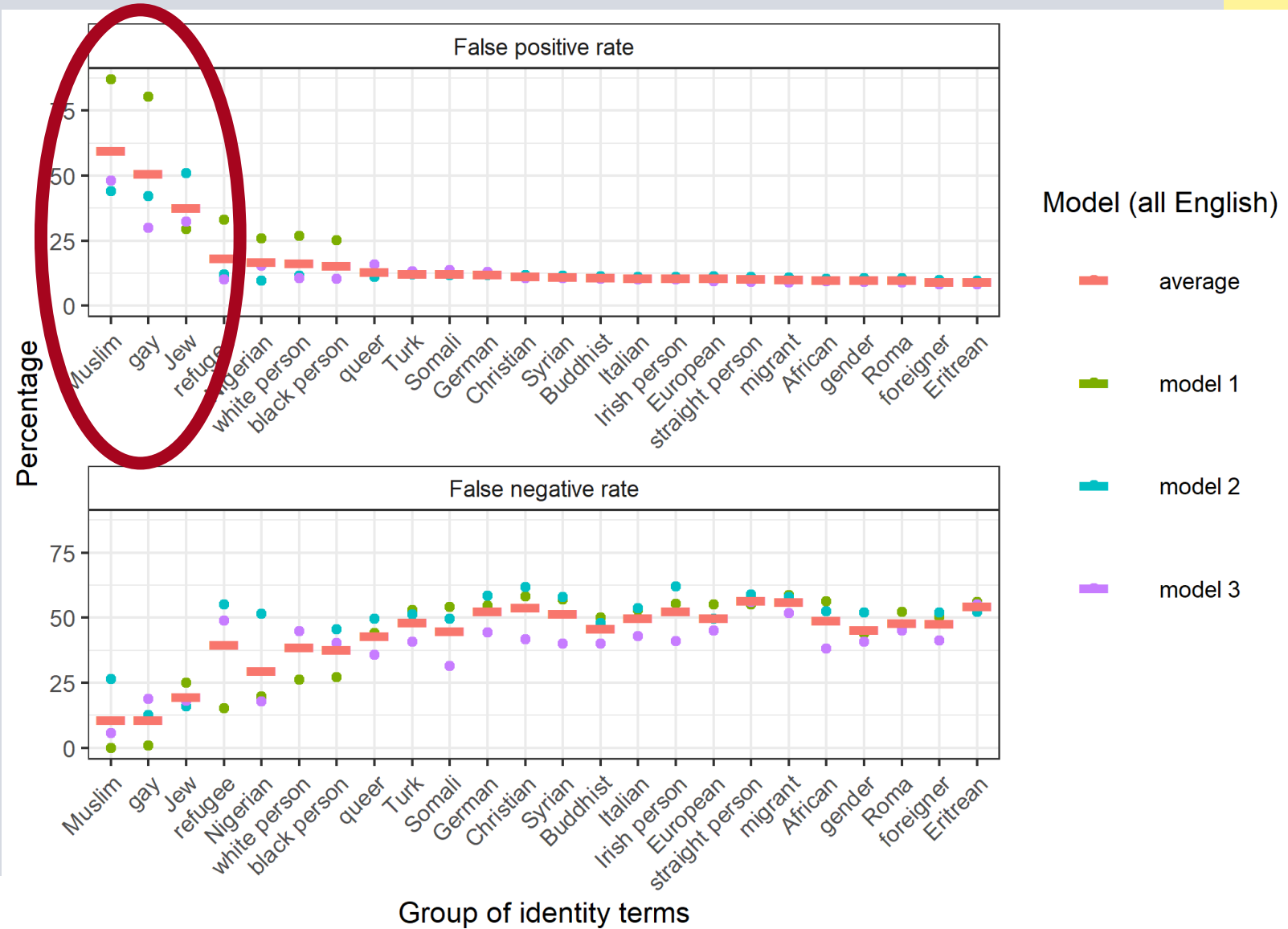
- Building algorithms based on social media data
- Three languages – EN, DE and IT
- Three models – basic, more advanced, highly advanced
- Using invented text to test the models for detecting offensive speech

Ethnic bias in speech detection

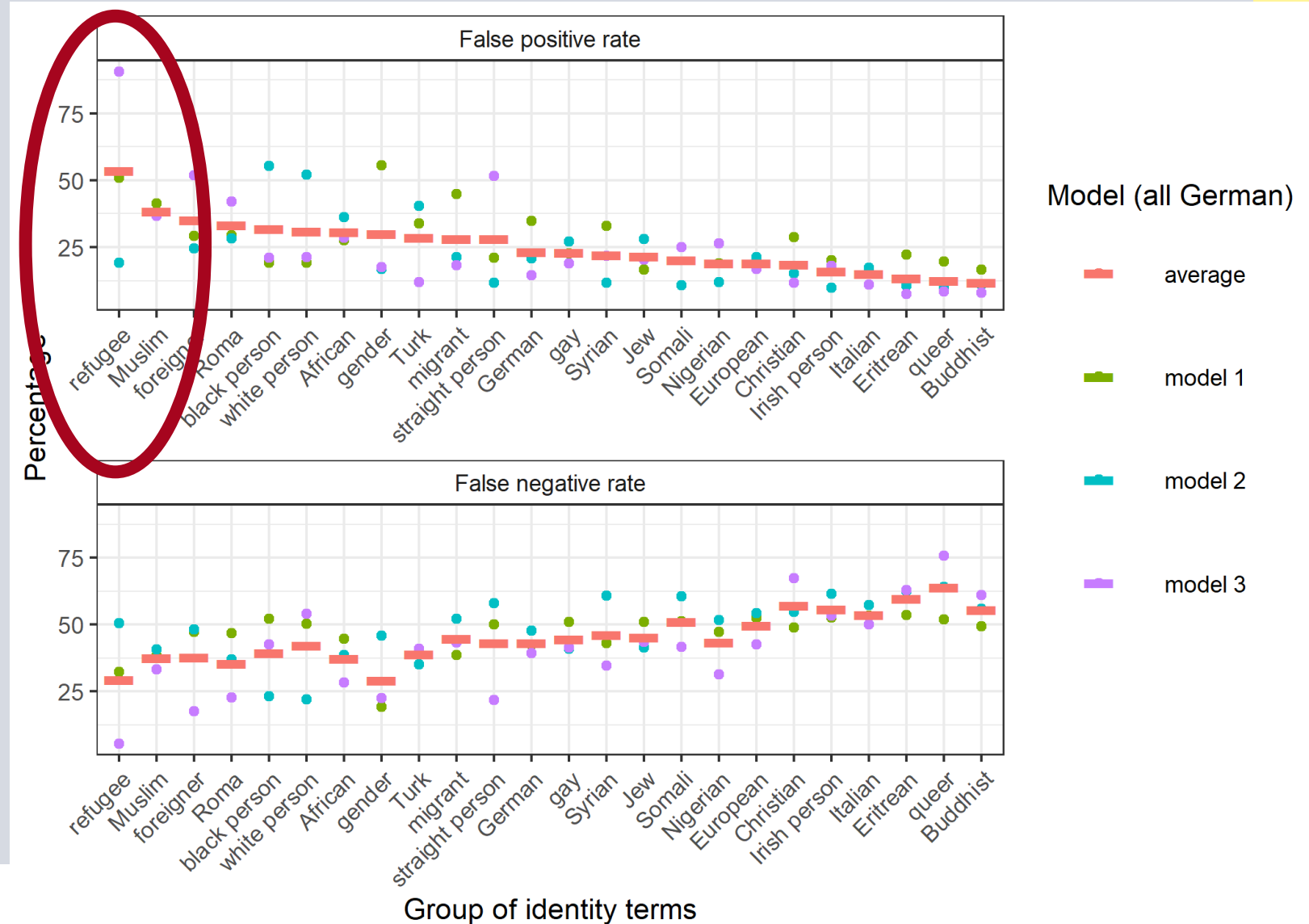
- Certain terms lead to strong predictions of offensiveness – even if the text is not offensive at all
- Strong bias against muslims and jews
- Varying results across languages, models and ethnic groups

Offensiveness predictions			
text	mean	min	max
i am buddhist	11	2	23
i am christian	5	2	9
i am jewish	33	5	86
i am muslim	72	28	94

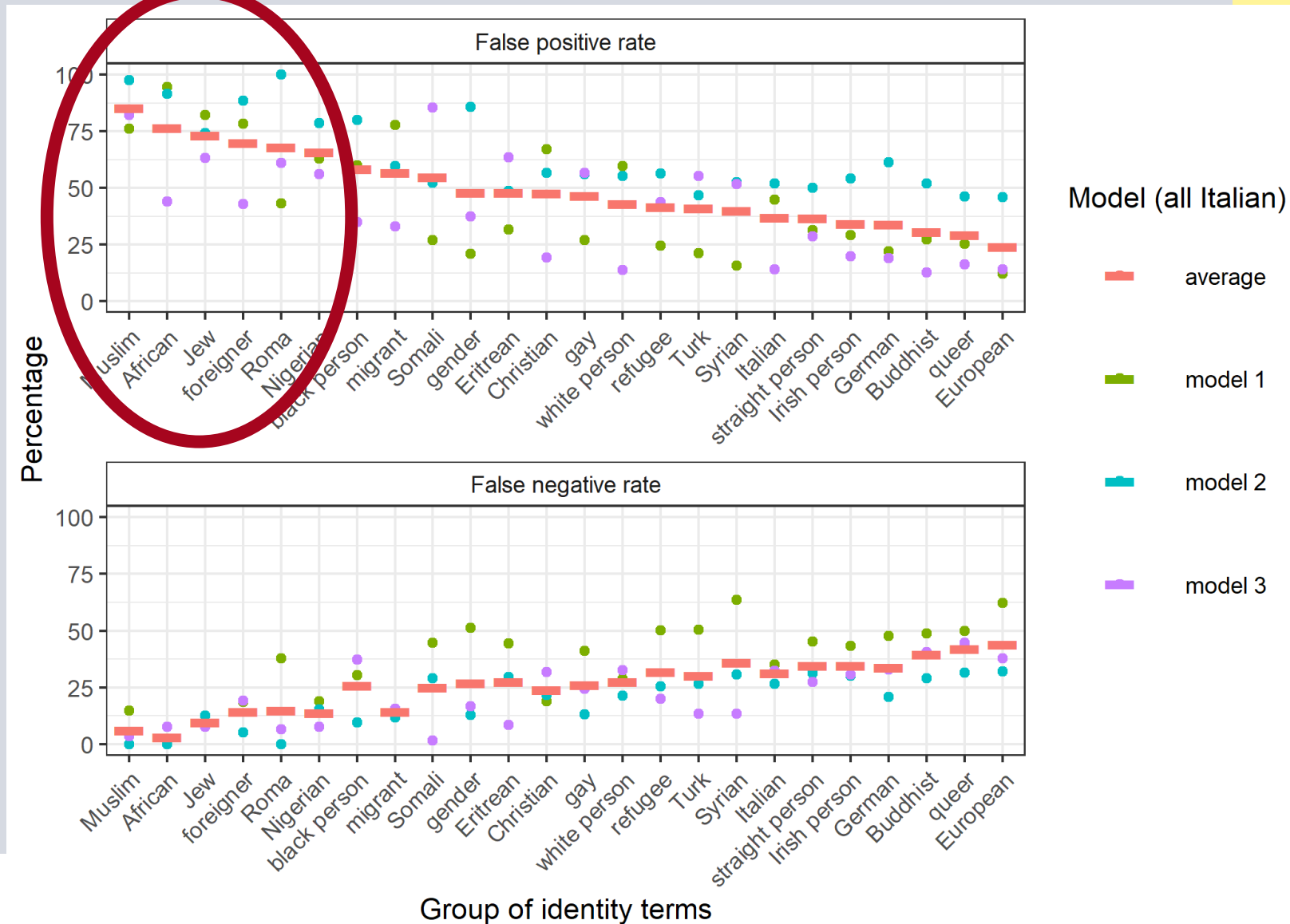
Varying results across models



Varying results across models



Varying results across models



Gender bias

- Masculine nouns slightly lower rated
- Intersectionality
- Only one additional dimension of gender inequality

Text	Gender	mean	min	max
sono buddhista*	M/F	30	1	46
sono cristiana	F	43	0	93
sono cristiano	M	34	0	59
sono ebrea	F	47	5	95
sono ebreo	M	67	17	95
sono musulmana	F	90	70	100
sono musulmano	M	78	46	94

Ways forward

AI Act – FR safeguards (selection)

- Article 5: Prohibitions
- Article 6 (3): ‚the filter‘ – for AI use that does not influence decisions
- Article 9: Risk management
- Article 10: Data governance
- Article 11: Technical documentation (Annex IV)
- Article 14: Human oversight
- Article 77: Access to documentation for national bodies protecting FR

- Article 27: Fundamental rights impact assessment

Relevant information needed to assess AI

- Purpose and context of systems, and legal basis
- Possible harm (false positives and false negatives)
- Technology
- Data description
 - Where do the data come from? What is included? Who is covered?
 - Missing information? Time frame and geographical coverage?
- Accuracy and tests (across groups)
- Most influential predictors

Ongoing projects – future outputs

- Assessing High-Risk AI (2023-2025)
 - Interviews with providers, users, experts and rights holders
- Remote Biometric Identification for law enforcement purposes (2023-2025)
 - Interviews, non-participants observation, small surveys, etc.
 - Cooperation with Europol
- Digitalisation of Justice (2024-2025)
 - Interviews with practitioners in the justice field
- Other projects (ETIAS, DSA, ...)

Thank you!

FRA – EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS

Schwarzenbergplatz 11 – 1040 Vienna – Austria

T +43 158030-0 – F +43 158030-699

 facebook.com/fundamentalrights

 twitter.com/EURightsAgency

 linkedin.com/company/eu-fundamental-rights-agency

fra.europa.eu