

FRA Press Release
Vienna, 8 December 2022

Test algorithms for bias to avoid discrimination

Artificial intelligence is everywhere and affects everyone. A new report from the EU Agency for Fundamental Rights (FRA) looks at the use of artificial intelligence in predictive policing and offensive speech detection. It demonstrates how bias in algorithms appears and how it can affect people's lives. This is the first time FRA provides hands-on evidence on how biases develop. The Agency calls on policymakers to ensure AI is tested for biases that could lead to discrimination.

"Well developed and tested algorithms can bring a lot of improvements. But without appropriate checks, developers and users run a high risk of negatively impacting people's lives," says FRA Director [Michael O'Flaherty](#). *"There is no quick fix. But we need a system for assessing and mitigating bias before and while using algorithms to protect people from discrimination."*

For its new report '[Bias in algorithms – Artificial intelligence and discrimination](#)', FRA developed two case studies to test for potential bias in algorithms:

- 1) **Predictive policing** shows how bias can amplify over time, potentially leading to discriminatory policing. If the police only go to one area based on predictions influenced by biased crime records, the police will mainly detect crime in that area. This creates a so-called feedback loop. In this case, algorithms influence algorithms, reinforcing or creating discriminatory practices that may disproportionately target ethnic minorities.
- 2) **Offensive speech detection** analyses ethnic and gender bias in offensive speech detection systems. It shows that tools used to detect online hate speech can lead to biased results. Algorithms may even flag harmless phrases such as 'I am Muslim' or 'I am Jewish' as offensive. There is also a gender bias in gendered languages, such as German or Italian. This can lead to unequal access to online services on potentially discriminatory grounds.

These results call for a comprehensive assessment of algorithms. FRA thus calls on the EU institutions and EU countries to:

- **Test for bias** – algorithms can be biased or develop bias over time, potentially leading to discrimination. Testing for bias before and during use, especially in automated decision making, reduces this risk.
- **Provide guidance on sensitive data** – to assess potential discrimination, data on protected characteristics (e.g. ethnicity, gender) may be needed. This requires guidance on when such data collection is allowed. It has to be justified, necessary and with effective safeguards.
- **Assess ethnic and gender biases** – ethnic and gender biases in speech detection and prediction models are strong. They need to be assessed case by case. Such assessments need to be evidence-based and made available to oversight bodies and the public.
- **Consider all grounds of discrimination** - biases are wide-ranging. So all prohibited grounds of discrimination, such as sex, religion or ethnic origin, need to be assessed. Various existing and proposed EU laws are needed to tackle discrimination by algorithms, including the proposed Equal Treatment Directive.

- **Strive for more language diversity** – speech detection models tend to focus on English. There is a need to promote and fund research on other languages to promote the use of properly tested, documented, and maintained language tools for all official EU languages.
- **Increase access for evidence-based oversight** – what lies behind AI systems can be largely unknown. Effective oversight requires improved access to the data and data infrastructures for identifying and combating the risk of bias in algorithms.

These findings aim to contribute to ongoing regulatory developments by informing policymakers, human rights practitioners, tech industry, and the public about the risk of bias in AI.

They are part of [FRA's work on artificial intelligence and big data](#). Previous research [identified pitfalls in the use of AI](#) and called on the EU and Member States to ensure that AI protects all fundamental rights.

For more, please contact: media@fra.europa.eu / Tel.: +43 1 580 30 653