FONDAZIONE BRUNO KESSLER
MAGAZINE

Artificial intelligence and discrimination

September 4, 2023

At the FBK-ISR international seminar held this summer, researcher Ilaria Valenzi presented the report "Bias in Algorithms - Artificial Intelligence and Discrimination" published by the European Union Agency for Fundamental Rights. We asked her some questions to understand when algorithms generate discrimination and what solutions can be put in place.

– **What is the European Union Agency for Fundamental Rights (FRA) about and what is the specialization of the people who work there?**

FRA aims to contribute to the establishment of a human rights culture in the European Union through the dissemination of the principles contained in the Charter of Fundamental Rights. This is done through advocacy in numerous areas, such as integration of migrants, discrimination, and combating racism and xenophobia. A particular area of action is the protection of personal data and Artificial Intelligence. In these areas, FRA collects and analyzes data, identifies trends, collaborates with EU institutions and member states and equal opportunity bodies, provides advice to policymakers, and promotes policy responses in line with fundamental rights. FRA draws on expertise in law, political and social sciences, and statistical sciences, as well as communication specialists.

– **What was the objective of the study “Bias in Algorithms – Artificial Intelligence and Discrimination”?**

The study examined the use of artificial intelligence in predictive policing and the detection of offensive speech online to demonstrate how biases in algorithms tend to reveal themselves and amplify over time, affecting people’s lives and potentially creating discriminatory situations. Starting from the assumption that the use of artificial intelligence for our societies is central, the study wanted to show how algorithms work in practice, developing models and evaluating their reliability for rights, especially of minorities.

– **Can you give us some examples of algorithms that have generated discrimination?**

There are numerous cases that have occurred in actuality. The report starts with one of the most notorious, the one that led to the resignation of the Dutch government in 2021. This was the case of an algorithmic system for allocating welfare benefits, which falsely accused 26,000 families of tax fraud, families who all had a migration background. Also among the most notorious is the Compas case, a predictive justice algorithm for predicting the risk of recidivism of convicted offenders, with obvious racial-ethnic discrimination against the African American population. Still, there are numerous cases of artificial intelligence that has “learned” to discriminate according to gender or ethnicity for resume selection or facial recognition (in which the machine recognizes, for example, the face of white male people but not that of black female people), or by associating skin color exclusively with certain job tasks. The scientific literature is full of such cases.

– **What emerged from the study?**

The results highlight how certain terms related to protected characteristics of people, such as religion, ethnicity, gender, and sexual orientation, contribute to classifying a text, post, or comment online, as offensive. Certain religious affiliations, especially when declined according to gender (such as the use of the term “Muslim” or “Jew”), are the most reported. The use in the Italian language of terms such as “foreign” increases the prediction of offensiveness. However, the algorithmic model does not always work: comments may be flagged as offensive that actually are not while situations of unrecognized offensiveness escape scrutiny. This risks limiting some fundamental rights or failing to ensure their protection. The lack of linguistic diversity is an aggravating factor.

– **What are the main issues to be resolved and what can be the solutions for the future?**

This involves paying special attention to the use of machine learning algorithms and automated decision making, making sure that periodic evaluations are carried out, improving data quality. From this perspective, algorithms need to be trained by taking into account gender differences and other protected characteristics of people, such as ethnicity, religion, and health status. It will also be increasingly necessary to promote linguistic diversity to mitigate biases in algorithms and restore the variety of our social and cultural contexts.

Ilaria Valenzi holds a doctorate from the University of Rome, Tor Vergata, School of Law. Since 2019 she has been a research fellow at Fondazione Bruno Kessler’s Center for Religious Studies. She is one of the principal investigators of the research project entited “Atlas of religious or belief minorities rights”. She is a member of the scientific council of the “Religions, Law and Economics in the Mediterranean Area” Research Center (REDESM) at the University of Insubria. She collaborates with the Waldensian Faculty of Theology for the teaching of law and religion. Researcher at the Centro Studi e rivista Confronti, where she deals with religious freedom in the post-secularization era. A lawyer, she deals with minorities, anti-discrimination law and data protection.

PERMALINK

https://magazine.fbk.eu/en/news/artificial-intelligence-and-discrimination/

TAGS

- #algorithms
- #artificialintelligence
- #bias
- #religiousstudies

AUTHORS

- Viviana Lupi