

Big data, algorithms and discrimination

IN BRIEF

With enormous volumes of data generated every day, more and more decisions are based on data analysis and algorithms. This can bring welcome benefits, such as consistency and objectivity, but algorithms also entail great risks. A FRA focus paper looks at how the use of automation in decision making can result in, or exacerbate, discrimination.



How it works

Put simply, algorithms are sequences of commands that allow a computer to take inputs and produce outputs. Using them can speed up processes and produce more consistent results. But risks abound.

Training algorithms

Algorithms are created in different ways. So-called ‘training data’ are used to find out which calculations predict a certain outcome most accurately. Take a classic example – spam filters. To create algorithms that take over the task of separating spam from legitimate emails, a set of several thousands of emails is first identified as either ‘spam’ or ‘not spam’. This is then used to identify characteristics that define the differences between the two types of email – for example, certain words or combination of words. Conducting this analysis permits setting rules for spam and then applying them.

Such ‘training’ methods can work well, but are not flawless – as we all know from both receiving emails trying to sell us questionable services and occasionally having to search the spam box for erroneously filtered messages.

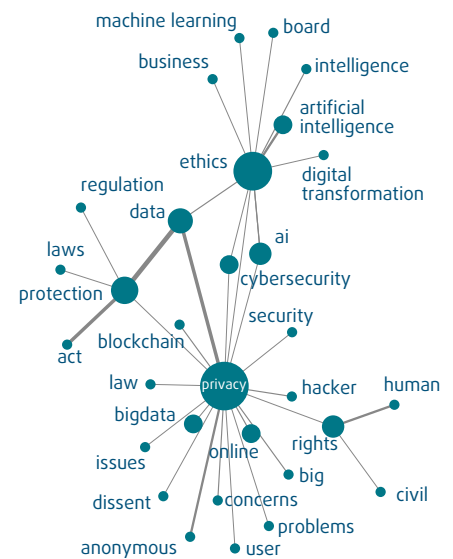
Nonetheless, algorithms are permeating all areas of life. Globally, they are also being applied in extremely delicate contexts – for example, to decide whether or not to jail someone pending their trial, or to determine which families to investigate based on child welfare concerns.

‘Garbage in, garbage out’

Algorithms are only as good as the data they are fed. If the data are outdated, incorrect, incomplete or poorly selected, results too will be questionable. With endless volumes of data being so quickly produced on the internet, without quality control concerning how these data are produced and then used, this is a serious concern.

The risk of discrimination also looms large. This can take different forms. For example, hiring decisions made by humans may sometimes be based on discriminatory behaviour or stereotypical thinking. If an algorithm is then used to make further hiring decisions, and it is trained with the data resulting from the human discriminatory behaviour, the algorithm itself will perpetuate the discrimination.

Or algorithms might be trained on data that are not necessarily biased but that are unrepresentative, meaning they do not allow for generalisation to other groups. For example, an algorithm may be trained using job applicants from a field that is predominantly male. Its predictions might be problematic when applied in another occupational field or to another group of applicants.



▲ Algorithms can help quickly identify links among large volumes of data.

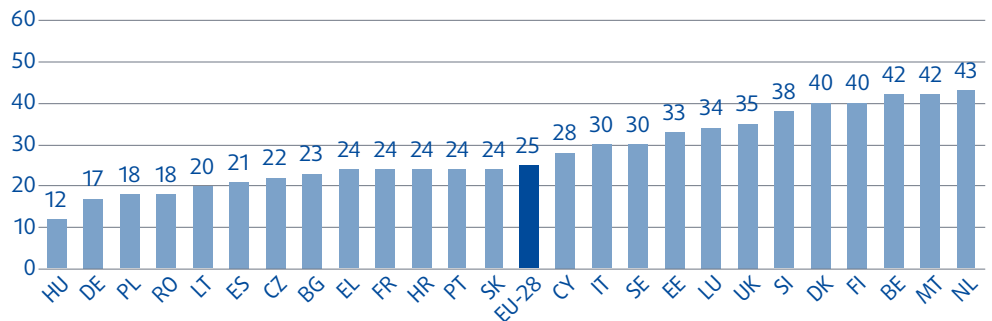
Charter corner

Article 21 of the EU Charter of Fundamental Rights forbids discrimination based on sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation.

In numbers

45%

of companies that use big data use social media data



Large enterprises (250+ employees) analysing big data from any source, by country, 2016 (%)

Source: FRA, 2018 (based on Eurostat data/no data for AT, IE and LV)

Uncovering discrimination

One way to test whether an algorithm is contributing to discrimination in employment, for example, consists of sending in two job applications for the same position, identical save for the applicant's gender or ethnicity. Whether so-called proxy information – such as names or postcodes for ethnicity – is causing discriminatory outcomes can also be checked.

Properly auditing algorithms can be extremely complex. Results may at first glance look discriminatory, but, upon further analysis, the issue may not be conclusive. A full review involves advanced statistical analysis. In addition, uniform standards for evaluating algorithms are lacking.

Copyright rules, companies' interest in protecting their business secrets, as well as privacy rules can all discourage openness about the data used, hampering meaningful reviews.

Building fair algorithms

As a result, truly making algorithms fair and non-discriminatory is a daunting exercise. But several steps can help move us in the right direction. These include:

- checking the quality of the data being used to build algorithms to avoid faulty algorithm 'training';
- promoting transparency – being open about the data and code used to build the algorithm, as well as the logic underlying the algorithm, and providing meaningful explanations of how it is being used. Among others, this will help individuals looking to challenge data-based decisions pursue their claims;
- carrying out impact assessments that focus on the implications for fundamental rights, including whether they may discriminate based on protected grounds, and seeing how proxy information can produce biased results;
- involving experts in oversight: to be effective, reviews need to involve statisticians, lawyers, social scientists, computer scientists, mathematicians and experts in the subject at issue.

In the spotlight

The new General Data Protection Regulation emphasises the need to prevent discrimination as a result of automated decision making, and gives people a right to 'meaningful information' about the logic underlying automated decisions.

The best starting point to test for bias is to review the attributes of people whose information is put into the system. But the GDPR restricts the use of sensitive personal data – important to keep in mind when trying to uncover bias.

Related reading



#BigData: Discrimination in data-supported decision making

FRA Focus
May 2018



Under watchful eyes: biometrics, EU IT systems and fundamental rights

Report
March 2018

For a more detailed discussion of this topic, see the agency's focus paper.

FRA – EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS

fra.europa.eu – info@fra.europa.eu

[facebook.com/fundamentalrights](https://www.facebook.com/fundamentalrights)

twitter.com/EURightsAgency

[linkedin.com/company/eu-fundamental-rights-agency](https://www.linkedin.com/company/eu-fundamental-rights-agency)

Printed by the Publications Office in Luxembourg

© FRA, 2018

CREDITS: © Stock.adobe.com (Scanrail)

PDF: ISBN 978-92-9474-241-4

doi:10.2811/315832

TK-01-18-907-EN-N



Publications Office