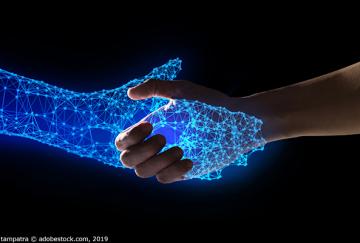
**Quality vital for data-driven artificial intelligence**

Rights areas:

* [Data protection, privacy and new technologies](https://fra.europa.eu/en/taxonomy/term/969)



A new FRA focus paper questions the quality of data behind automated decision-making and underlines the need to pay more attention to improving data quality in artificial intelligence.

**Publication**

[**Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights**](https://fra.europa.eu/node/36738)

Advances in technology are creating large pools of data. The lure of easy access to such data, reduced costs and faster data processing is driving the increased use of automated decision-making across many sectors such as finance, recruitment and policing.

Algorithms, used in machine learning and artificial intelligence, are responsible for analysing the data and making these decisions.

But flaws in the data can undermine the legitimacy of any decisions taken, as FRA’s focus paper on [Data quality and artificial intelligence – mitigating bias and error in algorithms](https://fra.europa.eu/en/publication/2019/artificial-intelligence-data-quality) seeks to highlight.

It raises awareness of the dangers of poor quality data as ‘garbage in’ can lead to ‘garbage out’ with a detrimental effect on people’s fundamental rights.

Algorithms based on poor data can negatively affect people’s rights to privacy, data protection, non-discrimination, gender equality and to justice.

This applies to both the data used for machine learning to train AI systems, as well as the data ultimately used when the systems go live.

Already, problems are arising.

For example, a hiring algorithm preferred men to women and a face recognition system worked well for white men but not for black women.

Common forms of errors result from using data to make decisions about people that the data do not fully represent. Another common error is using data that do not fully measure what they are meant to.

Drawing on data from internet and social media can be problematic, especially given the differences in rates of internet penetration in some countries and among some population groups. Across the EU, 11% of people overall, 27% of those aged 55+ and 30% of women with low formal education do not use the internet.

This underlines how large volumes of low quality internet data cannot validate some measurements, when they do not represent some people and they do not measure what they are supposed to.

Borrowing from long-established social sciences and survey practices of clearly stating the source of data and what is covered would help to trace errors. It would also allow corrective measures to be applied, if needed.

Improving the understanding of the importance of high quality data can help policy-makers, business and developers address such issues and avoid potential problems. Ultimately, it will also lead to better artificial intelligence and better automated decision-making, as this paper shows.