

A-level results: why algorithms get things so wrong – and what we can do to fix them



When A-level grades were announced in England, Wales and Northern Ireland a few weeks ago, nearly 40% were lower than teachers' assessments. The grades were formed using an algorithm based on the schools' past performance and the grading process had "significant inconsistencies", as [acknowledged by the Education Secretary Gavin Williamson](#). Following numerous [protests against the unfairness of the system](#) biased against students from poorer backgrounds, the government abandoned its decision and students received grades based on their teacher's assessments. Still, [algorithms affect decision-making in many aspects of our lives](#) including healthcare, policing, banking, insurance, or social media. In this blog republished from [the Conversation](#), [Amany Elbanna](#), [Royal Holloway](#) and [Jostein Engesmo](#), [Norwegian University of Science and Technology](#) discuss the pitfalls of

algorithmic decision-making and the possible social implications related to fairness and justice.

The [scale of public anger](#) over the automated downgrading of thousands of students' A-level results highlights how much social and political power algorithmic decision-making now has. As well as students' grades, algorithms [are now deciding](#) all sorts of things that hugely impact ordinary people's lives, from loan applications to job interviews to which neighbourhoods are targeted by police.

Too often the outcomes of these decisions are what most people would consider unfair, as was the case for the students whose results were downgraded despite having strong academic records or based on their school's past performance not their own. How are these algorithms going so wrong, and how can we ensure they produce fairer outcomes in the future?

In computer science, an algorithm is a set of instructions based on a mathematical model that tell a computer how to perform a calculation. The model is usually built from data about past decisions and some of the factors used to make them.

The algorithm can then automate decision-making, so large amounts of data can be processed efficiently in a short period of time. Machine learning algorithms will improve their models as they process more and more data.

It is in this data used to build and train algorithms that many of the problems lie. First, algorithms typically need relatively large data sets to work well. So in the case of the A-level results, small classes of fewer than 15 students [still had their teachers' assessments taken into account](#) but larger classes didn't.

Another key issue is that data about the past doesn't necessarily help you make adequate decisions about the present or the future. It blocks any chances of change and development – like when a school improves its teaching or one year group of students performs better than their peers in previous years.

This might not matter when Google or Amazon tries to work out what ads or recommendations might be useful for you based on what other people of similar profile have liked. But determining your future based on someone else's past has much greater implications.

The kind of social data that is involved in these critical life decisions is inherently unpredictable. Building a model of how a tumour will react to treatment is grounded in well-established laws of nature about molecules and cells. But people don't behave according to similar laws. This increases the chances that test data used to build algorithms could be different from the real data they process, and that the decisions of the algorithm will be inaccurate or unfair.

On top of this, all social data holds biases that an algorithm can end up replicating. For example, the A-level algorithm adjusted results to try to replicate the previous overall achievements of different [ethnic groups](#), which are likely to reflect racial inequality. Again, relying on historical data to train an algorithm locks in the problems of the past, preventing changes in society or efforts to address these biases from showing up in the way the system works.

Finally, social data also carries a political and social meaning. For example, a cohort of less than 15 students that is excluded from being subjected to the algorithm is probably either a class in a private school or studying a less popular subject. So, a convenient decision based on the functional working of the algorithm will have serious social ramifications, in this case advantaging private school students or students studying less popular subjects.

This means that you can't remove the systematic discrimination against certain characteristics that can be found in biased algorithms simply by avoiding using those characteristics in the calculation because other data can act as a proxy.

There's also a broader problem. Algorithms supported by machine learning aim not to replicate the decisions of experts but rather to replicate the average decision-making from past data.

This logic of averaging society is dangerous for a society that values individual creativity and achievement. It prevents distinction and excellence as the algorithm systematically pushes people towards the average.

All this means that algorithmic fairness is a multifaceted problem that a technical solution alone cannot solve. Instead, the way to make sure people aren't unjustly disadvantaged by an algorithm is to involve them closely in its development.

People power

Our research has shown that people using an algorithm can guess how it works and detect changes in it simply by being on the receiving end of its decisions. For example, [we found](#) that workers using digital labour platforms such as Uber and Fiverr can work out how to manipulate the data that goes into the system in order to receive more favourable decisions.

Another of [our studies](#) showed that people working in an organisation using AI-enabled decision-making could detect when its decisions are incorrect. This means they can act as early detection system for unfair and biased decision.

In one successful case, the organisation developed its algorithms in close consultation with the people who used to be in charge of the decision-making and with different types of users. It created a way for workers to register their observations and highlight any problems to be corrected.

The organisation also recognised that it still held responsibility for the decisions made by the machine. So it created a mechanism for explaining the algorithm and its decisions so that different workers would have trust in the system and be able to report when it went wrong.

When algorithms have such power over our lives, it's vital these systems are the result of political debate and deliberation between everyone who is affected by them. Such debate ensures that the algorithm is transparent, explainable and accepted.

The A-level fiasco is a strong lesson in why we need to reconsider the importance of fairness in algorithms, their data and the mathematical models that govern them. Algorithms cannot be allowed to make social decisions without having an understanding of their social implications at their heart.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

This post gives the views of the authors and does not represent the position of the LSE Parenting for a Digital Future blog, nor of the London School of Economics and Political Science.

Featured image: photo by Gerd Altmann on Pixabay. In-text image: photo by energetic.com from Pexels