Algorithms to live by

The challenge of practically integrating an ethical and social approach in the development and implementation of AI needs to be urgently addressed, to help restore public trust in technology.

We may look back on 2020 as the year that exposed several long-standing structural and complex problems in society and the damage caused by attempts to address them with quick technological and algorithmic solutions. A prime example of a [misguided approach](#) to address a complex social issue is the attempt of the UK government to standardize the outcome of A-level exam predictions of individual students by implementing a formula based on past school performance. The unsurprising result was that students from disadvantaged backgrounds were affected most by downgraded exam marks, highlighting once again how algorithms used for predictions affecting individual lives tend to perpetuate and amplify entrenched divisions.

The pandemic caused by the novel coronavirus has presented huge challenges for secondary school pupils who were preparing for a transition to higher education or the job market. They had to quickly get to grips with digital technology and online-only learning. But most couldn't sit exams, which posed a problem for countries like the UK, where exam grades weigh heavily in university admissions. Teachers' predicted grades had to be used instead, but the government's exam regulator Ofqual, which ensures that grades are not subject to yearly inflation, decided to use an algorithm that downgraded 40% of the results. The process caused despair among many students who looked set to miss out on university places conditionally offered to them. After a vigorous day of protest, the government abandoned the downgrading, silently acknowledging that the algorithm unfairly affected the most disadvantaged students and schools.

While the goal of standardizing exam results was in itself not a bad idea, the expectation that a simple algorithm can make fair predictions on individuals based on data of past school performance was misguided. Everything from the choice of which data are used (and what data are available in the first place) to the specific questions addressed (and by whom) are subject to societal biases that were not examined. Instead of turning to an algorithm straight away, the government could have taken on the task, long overdue, to review and overhaul the exam system, addressing inherent problems with inequality.

By postponing work that is needed to address deep structural problems in society and by trying to patch them over with technological and algorithmic fixes, society accrues what Shannon Vallor calls [moral debt](#). The term is related to the concept of technical debt in software applications, where imperfect software is quickly brought to market, making use of patches and shortcuts, with the result that for years after deployment, additional resources have to be spent to repair and eventually replace the software. Similarly, in the decades leading up to 2020, substantial environmental and social costs have been accumulating with technological and digital transformations. 2020 may be a year of reckoning as societies can no longer put off work to address issues such as underinvestment in healthcare, environmentally unsustainable practices, and racial, gender and other social injustices.

There is a risk that AI technology is on its way to add to moral debt, either by bad fixes of problems or by producing new problems such as the rise of social media platforms,

which threaten to undermine the democratic process, facial recognition in surveillance, which infringes on human rights, and an epidemic of disinformation with deepfakes. The only way forward is to recognize that there is a moral debt to pay off and to focus on developing AI that does not contribute to or wash over societal problems, and ideally helps to address them.

In the development of new AI applications in society, and in deciding whether they are sufficiently beneficial, fair and transparent to be deployed at all, it will be necessary to turn to what McLennan et al. call in this [issue](#) 'embedded ethics'. The group of researchers in AI, robotics, philosophy and social sciences argue that it is imperative to address ethical challenges as early as possible in the development of AI technology. Developers should not only call upon ethicists when they perceive ethical problems; rather, it should be standard practice to directly involve ethicists as members of a development team or to organize regular exchanges with ethicists at every step of the process, to collaboratively identify, anticipate and address ethical, moral and legal challenges.

This year's events call for confronting long-standing societal problems. AI technology has a role to play but not with quick algorithmic fixes. New technology can only be deployed with clear ethical processes in place and transparency in motivations and impact, to restore public trust in technology — and this will be essential when wide acceptance is crucial, as with COVID-19 vaccination programs. □

Published online: 14 September 2020
<https://doi.org/10.1038/s42256-020-00230-w>