ROYAL STATISTICAL SOCIETY

DATA | EVIDENCE | DECISIONS

POST-ELECTION BRIEFING: BALANCING AI INNOVATION AND REGULATION

RSS manifesto ask

A public register of cases where a complex algorithm or artificial intelligence (AI) tool is used in the public sector – with risk assessments carried out in cases where the tool directly impacts on citizens (eg, its use in facial recognition or in informing decisions around welfare payments).

Summary

Complex algorithms and AI tools have the potential to make public service delivery more productive. However, if the use of these tools becomes widespread without public scrutiny (as seems to be happening in, eg, the justice sector) there is a risk that the public will lose trust in the technology. Public trust is required if the tools are to be used in decisions that impact on individuals. Transparency is vital in building public trust. The starting point should be a register of all cases where complex algorithms or AI tools are used in the public sector. Risk assessments should be conducted and made public in cases where the tool directly impacts citizens, so the public can see the decisions that have been made and what steps have been taken to de-risk the use of such tools.

What's the problem?

Complex algorithms and AI tools have a wide range of possible applications in the public sector – and there is huge potential for the technology to improve productivity and help make better decisions. Used well this technology can have positive impacts – potentially helping in a wide range of areas from crime detection, to healthcare diagnoses to assisting welfare entitlement decisions.

The risk is that the technology gets deployed in these areas and used to make decisions that impact people's lives without proper scrutiny – that the practice of AI-assisted decision-making proliferates across the public sector without proper oversight. Indeed a recent House of Lords Select Committee report, *Technology Rules?*, suggests that this has already happened in the justice system where they highlight "a new Wild West, in which new technologies are developing at a pace that public awareness, government and legislation have not kept up with" (p3).

There is currently a lack of transparency in how complex algorithms and AI tools are being used in the public sector. Finding out about how the technology is used by government currently relies on piecemeal individual investigation in areas of interest – a recent example is Big Brother Watch's work

to uncover the Department of Work and Pensions's use of algorithms in assessing potentially fraudulent cases. This approach is not satisfactory, and risks undermining public trust.

We saw the importance of transparency in algorithms during the pandemic, when Ofqual attempted to use an algorithm to assign grades to students in the absence of exams. There was a lack of transparency around the algorithm in advance of people receiving their grades, which meant a missed opportunity to engage with experts. To their credit, however, they did release the details of the algorithm on results day – this enabled experts (including the RSS) to assess the algorithm and determine that the algorithm was not robust enough to bear the weight being put on it. It was subsequently abandoned.

The root issue is that using complex algorithms and AIs for the public good is only partly about the quality of the models themselves. You could have an algorithm or AI tool that works exactly as intended and that is based on full and representative data – but that is only part of the story. Decisions will be made in the design of the systems that embed political judgements. In the case of Ofqual's algorithm, these judgements were around how much grade inflation to allow and how to be fair to individual students. The public need to be able to know when complex algorithms and AI tools are being used in a way that affects them and what decisions have been made in how they are developed and used.

How to fix it

Transparency is key. After the Ofqual algorithm fiasco, the RSS asked the Office for Statistics Regulation to conduct a review of the case and to draw lessons for future use of statistical modelling in policy making. Their report, *Ensuring statistical models command public confidence*, drew three key lessons for organisations developing algorithms (p.62):

- Be open and trustworthy: this means being transparent about aims of the model and the model itself, including limitations, and acting on feedback.
- Be rigorous and ensure quality: this means ensuring that there are clear governance processes and accountability, involving subject matter and technical experts and ensuring that both data used as inputs and any outputs are quality assured.
- Meet the need and provide public value: in this context it is particularly important to engage with affected groups to test and ensure the acceptability of any new approach.

These are all important, but we wish to emphasise the first point as fundamental. Without transparency about what models are used and what they are aiming to do, we risk a situation in which the public lose faith in the technology.

A key starting point is the creation of a publicly available register that sets out where complex algorithms and AI tools are being used. In cases where citizens are directly affected by the technology risk assessments should be conducted – and made available – setting out decisions that have been made, limitations of the model, the possible impact this will have and what mitigations have been put in place.