Promising Trouble

Automating public services

A careful approach

Author: **Anna Dent**
Published: **July 2024**

Introduction

Automation is being introduced in public services at local and national levels, for administration, decision making, assessment and a whole host of other functions. Some automated systems are public facing, in which citizens interact directly with a system, while others are used behind the scenes by staff. Public sector organisations hope that automation will help to deliver their objectives, **improve efficiency, free up staff time, achieve savings, and deliver better quality, more consistent services**.

While these outcomes are potentially achievable through automation, they are by no means guaranteed. **Public bodies need to take a clear-eyed and informed approach, and avoid being swept up in hype or over-exaggerated claims**. The risks of getting it wrong can be significant, both for citizens subject to automated services, and for public bodies themselves.

This briefing provides a summary overview of where automation is already being used in UK public services, and discusses **the range of risks which must be considered when setting out to develop or commission a new automated system**. For the purposes of this briefing, automation / automated systems includes algorithms, artificial intelligence, machine learning and predictive analytics.

What is being automated?

A vast range of functions and activities are being automated across the public sector; this short list of examples highlights the variety of settings and tools available and in development:

- Local authorities are experimenting with chatbots, **automating social care assessments and using data analytics** to identify at-risk children and families, among many other digital transformation projects[1]

[1] Local Digital, Local Digital Fund projects; Redden et al, 2022, Automating Public Services: Learning from cancelled systems, Scoping Report

- GP practices are trialling **AI-powered triaging** and appointment management[2]
- The NHS is developing **AI diagnostic tools** to improve the speed and accuracy of diagnoses[3]
- The Department for Work and Pensions is using and trialling a number of **automated systems to identify claimants who may be committing fraud**[4]
- The Home Office uses automated decision making across a broad range of functions, including **identifying people at risk of modern slavery** and sharing immigration decisions.[5]

The sheer range of services and functions for which automation is being used can make it appear that the whole population is subject to automated decision making on a regular basis. However, a clear understanding of the populations that are actually most affected is essential: **some communities are more likely to experience automated decision making**, and the impacts will not be evenly distributed.

If you rely on support from the state, if your immigration status is uncertain, if you belong to any community which is already marginalised in some way or disproportionately targeted for enforcement or suspicion you are probably going to have more automated decisions made about you. Combined with the potential for things to go wrong, or for systems to be flawed in terms of bias and privacy, **the cumulative impacts and risk of harm cannot be overlooked**. When we consider the kind of high impact service areas noted above, the implications for individuals can be life changing.

[2] Deloitte UK, AI and Automated Decision Making
[3] NHS Transformation Directorate, NCCID case study: Setting standards for testing Artificial Intelligence
[4] Public Law Project, Tracking Automated Government Register
[5] Ibid

Considerations for public service automation

There are a range of issues to consider when designing and deploying automated decision making systems. Some relate directly to the automated systems themselves and their effectiveness; do they actually achieve what they are designed to do? Do they cut down on time spent by staff or administration costs? Are they more accurate than human decision-making?

Others stem from the impact of the systems on citizens, and the responses this provokes among politicians, civil society and the media. On occasion, the impacts are considered so serious that legal action is taken by those affected to seek redress.

Effectiveness

The question of effectiveness needs to be addressed at all stages of automating any element of public services. It is easy to be impressed by the claims of tech companies about the potential of automated systems to improve myriad aspects of service design and delivery. Before committing to the idea of automating something, **the potential for realising an effective solution needs to be interrogated thoroughly**, not just relying on evidence from those with a vested interest.

There is an understandable temptation to introduce automation to complex, deep-rooted and expensive social challenges; the impact on citizens and organisations of finding reliable automated solutions would be significant. However, **by their nature these challenges are not well suited to rules-based, automated processes** which struggle to take into account the interconnected nature of our lives, because they need to reduce complexity into data points. Social problems rarely exist in a vacuum, and we can't neatly package their solutions, but automation suggests that we can.

There are multiple examples of automated systems introduced and then withdrawn because they were found to not be delivering what they set out to do. In London, **Hackney Council halted a pilot programme to use predictive analytics to identify at-risk families**. The system was found not to provide good enough information to continue using it, while at the same time it was attracting local criticism for its handling of privacy and consent.[6]

In Bristol, a risk-based verification (RBV) system for assessing the risk of fraud by council tax and housing benefit recipients was dropped because **it did not deliver on the anticipated time savings for staff.** Other RBV systems have been dropped by councils because they were misclassifying applicants as high risk, thereby delaying their eligibility for support, and because the cost of the system was not justified when its impact was considered.[7]

A DWP system designed to identify potentially fraudulent claims for housing benefit actually flagged genuine claims by mistake in two-thirds of the cases it identified. Not only did this mean **individuals and households were unnecessarily investigated**, with all the associated stress and anxiety, but millions were spent on investigations which did not cut down on fraud or save any money.[8]

When automated systems don't deliver as they are supposed to they can have life-changing consequences. The 'RoboDebt' system in Australia wrongly accused thousands of benefit recipients of owing the government money. **Correcting these mistakes was extremely difficult**, with the onus entirely on citizens, and many people suffered significant stress and hardship in the process. It was also extremely costly to government; when the system was ruled to be unlawful, the **government not only had to repay debts of over $721m, they also settled in a class action lawsuit for over $1bn.**[9]

[6] Redden et al, 2022, Automating Public Services: Learning from cancelled systems, Scoping Report
[7] Ibid
[8] The Guardian, 2024, DWP algorithm wrongly flags 200,000 people for possible fraud and error
[9] Redden et al, 2022, Automating Public Services: Learning from cancelled systems, Scoping Report

Bias

The potential for biased and otherwise unfair decision making by automated systems is increasingly well known.

It often happens because the data used to train the system is intrinsically biased. Predictive policing tools trained using data about previous arrests for example are highly likely to reflect racial and other biases in policing.[10] West Midlands Police stopped trialling a predictive tool meant to identify people most likely to commit violent crime, in part **due to concerns about the data which was being used to predict future crimes**.[11]

The A Level algorithm employed in 2020 to predict the results of exams that could not be taken due to the pandemic was highly criticised due to its inaccuracies, as well as the likelihood that it would unfairly disadvantage some schools and their pupils. Basing predictions on grades from previous year groups meant **pupils in 2020 could not outperform their predecessors, regardless of their actual attainment**.[12] Government was threatened with legal action and ultimately had to withdraw the algorithm, causing delays for students and disruption within government and schools.

The Home Office's 'streaming' algorithm, which operated from 2015 to 2020, was found to have racial biases built in. **The algorithm was informed by biased practices such as targeting particular nationalities for immigration enforcement**, thereby making the immigration process more difficult for people of those nationalities. Faced with legal action, and public protest, the Home Office stopped using the algorithm.[13]

Bias can also arise when particular groups or communities are disproportionately subject to processing via an automated system. In the Netherlands a fraud detection system

[10] European Union Agency for Fundamental Rights, Test algorithms for bias to avoid discrimination
[11] Redden et al, 2022, Automating Public Services: Learning from cancelled systems, Scoping Report
[12] Ibid
[13] Digital Freedom Fund, UK Home Office visa application streaming algorithm

called SyRI, set up to identify individuals likely to commit benefit fraud, was used most frequently in locations with high proportions of low income households, and those of non-Dutch nationality. Unsurprisingly, as a result **more households from these backgrounds were targeted for fraud investigations**. This targeting was one reason for the system's withdrawal, following public outcry and a lawsuit against the government.[14]

Privacy and data protection

Most automated systems rely on large amounts of data to power their decision making. The sources of this data are not always clear to citizens, and opportunities to give informed consent to its processing or to opt out are often limited, if available at all. Designers and commissioners of automated systems need to **consider whether the data required to power a system is proportionate**; does the outcome of the system justify the type and volume of data needed? Do citizens know that their data is being used in this way? Does it breach a reasonable expectation of privacy?

Before the Data Protection and Digital Information Bill failed to be passed into law (it did not get through parliamentary process before the start of the 2024 pre-election period), government was attempting to give the DWP the power to require third party organisations to share data they hold on claimants in order to tackle fraud and error.[15] The impact of this on claimants' privacy would have been significant, and the **open-ended nature of the powers could have enabled the DWP to obtain data from any number of additional sources** in the future, without needing to gain claimants' consent. The proposal attracted a great deal of controversy, and it has been suggested that its addition to the Bill is one reason it failed.[16]

In the Netherlands, the fraud detection system SyRI was found by their courts to not comply with the EHRC, as it did not strike a reasonable balance between the aims of the system and citizens' rights to a private life. The system used data from a wide range of

[14] Redden et al, 2022, Automating Public Services: Learning from cancelled systems, Case studies
[15] HM Government, 2024, Data Protection and Digital Information Bill - (Amendment Paper)
[16] Tech UK, 2024, Data Protection and Digital Information Bill falls ahead of the UK General Election

sources, including identity data, employment data, data about incomes, assets, debts, pensions and education.[17] The court considered that **the system effectively meant these citizens had no right to privacy**, and their data was analysed regardless of whether there was any indication of fraud at all.

Lack of control and transparency

Many automated systems are supplied by private companies. While this offers potential benefits in bringing technologies into public sector organisations which may not have the capabilities to do themselves, it also introduces challenges around ownership and control.

Many private suppliers consider information about systems to be commercially sensitive, and so are reluctant to release details about how they work, even to the organisations purchasing them. One council using a Risk Based Verification system was not given information about why different people were categorised as low or high risk as this information was considered confidential and therefore known by the supplier only. This **lack of transparency meant the council were unable to accept the system's risk ratings with any confidence**, which limited any benefits of deploying the system.[18]

Public bodies also set limits on what information is released into the public domain. The ensuing **lack of transparency limits citizens' ability to challenge decisions**, correct mistakes and understand why particular decisions have been made. For example, the Public Law Project's register of automated public sector systems notes many cases where public bodies refuse to release information on the grounds that to do so would make it more difficult for them to detect and stop fraud and crime.[19]

Again, this makes it extremely difficult for citizens subject to processing by these systems to hold them to account, and rectify any mistakes. Of the 55 systems currently on the

[17] Redden et al, 2022, Automating Public Services: Learning from cancelled systems, Case studies
[18] Redden et al, 2022, Automating Public Services: Learning from cancelled systems, Scoping Report
[19] Public Law Project, Tracking Automated Government Register

Public Law Project's register, **only one is considered to have a high level of transparency**. We only know about many of these systems as a result of Freedom of Information requests, research and campaigns by civil society organisations and groups of people adversely affected by them.

Trust

Trust is fundamental to effective and ethically sound automated public services. Without trust in the institutions using automated systems, and in the systems themselves and the outcomes they produce, **citizens will resist their data being collected and used, and may avoid engaging with services and support they need**.[20]

Layering an automated system onto a public service that already lacks public confidence will not restore it. Introducing an automated system without consultation, transparency and accountability is likely to generate further doubt and a breakdown of trust. Controversies surrounding data sharing and patient privacy, and the role of private sector provider Palantir, threaten the success of NHS digital transformation.[21] The Post Office / Horizon scandal illustrates how significant the breakdown of trust can be when an IT system is prioritised over its users.

Countries considered leaders in digital public services, both in implementation and take-up by citizens, demonstrate **the value of trust**. In Denmark, digitisation and automation have been introduced to public services which already enjoyed a high level of confidence from citizens. Estonia has prioritised openness and privacy in its digital government programmes.[22]

[20] Connected by Data, 2024, Connected Conversation: What would it take to secure full public support for NHS data re-use?
[21] The Guardian, 2023, NHS data platform may be undermined by lack of public trust, warn campaigners
[22] CBO Projects, 2015, eGovernment: Learning from Denmark and Estonia

Conclusion and Guiding Principles

If the potential benefits of automation are to be realised in public services without sacrificing citizens' rights and trust, these **risks must be actively considered at a general and specific level**. Public bodies should have a coherent overall strategy for understanding and mitigating the risks of automation, and apply a consistent approach to assessing and addressing the risk of each individual automation project they implement. **Preventing harm rather than having to deal with the fallout of a system gone wrong is infinitely preferable**, for public bodies just as much as the people affected.

The following principles will support public bodies in shaping their approach to automation in ways that balance risks and benefits, and ensuring that they focus on the human impacts:

- Interrogate the reasoning for introducing automation: **what are the underlying policy drivers and assumptions**? Does it risk replicating or exacerbating harms?
- Consider where automation can have the **biggest impact for the lowest risk**; this may be at the less headline-grabbing end of digital transformation, but the long-term benefits and reduced risk may be preferable
- **Consider introducing 'red lines'**; decisions which are too risky and have too great a potential impact on people's lives to employ automated systems
- Introduce more **thorough and consistent safeguarding and impact evaluation**, with real powers to make changes to automated systems that are causing harm
- **Improve transparency**, making public what systems are being used and how they work, and publishing impact assessments in all cases of automated decision making
- **Upskill public sector staff** to understand more about automated decision making systems, what they can do, how they work and the risks and benefits involved, to **create better informed purchasers and decision makers**
- **Involve more citizens** and other stakeholders such as civil society groups in decision making and monitoring of automated systems and their impacts.

For more information

Please contact

hello@promisingtrouble.net

hello@promisingtrouble.net
www.promisingtrouble.net

Registered address: 7 Bell Yard, London, WC2A 2JR

Registered in England & Wales No: 13334908