Learning from past mistakes

ECONOMIC JUSTICE AUSTRALIA
For a fair social security system

Principles for safe use of automation in social security

Automation is now integral to the functioning of the social security system. Automated decisions, made without human scrutiny, can have serious consequences for those they effect. Past mistakes have clearly demonstrated the need for a robust, principled approach to automation in social security to ensure those mistakes are not repeated, and that people are not harmed in the process.

The social security system uses automation for multiple purposes, including assessment of payment eligibility, calculation of payment rates, decisions about whether a person has a debt, and calculation of the amount of money owed. When automated decision-making (ADM) gets things wrong, it is frustrating, distressing and debilitating for the people impacted.

Principles for safe use of automation in social security is based on interviews with Economic Justice Australia (EJA) member centre casework staff, from the 22 community legal centres that provide specialist advice to people on their social security issues and rights. EJA members raised concerns about the roll-out of automated decisions, including the design and testing of systems, accuracy of decisions, lack of transparency about how decisions are made, questionable legality of some decisions, and the impact of decisions on people in vulnerable circumstances.

Such risks could be mitigated through the application of specific principles for the considered use of automation. EJA's *Principles for safe use of automation in social security* is intended to ensure automation in social security is fair, safe and carefully implemented. ADM should enhance rather than diminish the quality of services, protect the rights and dignity of recipients, accommodate the needs of diverse groups of people and those experiencing complex situations, and maintain the fundamental purpose of social security as a safety net for those who need it.

These principles have been developed at the same time as a number of related government consultations have been occurring. They are guided by the insights of earlier scrutiny of automation, ADM and AI by the Commonwealth Ombudsman, the NSW Ombudsman, the

Department of Industry and other bodies (please see references section for an indicative list of sources).

Vulnerable populations

These principles recognise that social security recipients may be vulnerable, and can be negatively and disproportionately impacted by automated systems.

Principle 1.1 Beneficiary-centric design: Prioritise the development of automated systems that enhance service provision to social security recipients

The needs of people using the social security system must be at the forefront of social services, recognising that vulnerable people often require personal interaction and support. The effects of automation on people who rely on the social security system should be considered at every stage of service delivery, including during decision-making.

> *Human services should involve humans — not just services. Vulnerable people need a person to latch on to. (Interview 10)*

> *The impact on the person is at the point the decision is made — not after the fact. It should not be about solving it later — because this is about income support. There are impacts that aren't curable later. (Interview 19)*

> *The Centrelink client base is vulnerable by definition. (Interview 6)*

Principle 1.2 Risk assessment: Conduct thorough risk analysis that considers the legal effects of automation and its impact on people's rights, health and safety

It is necessary to conduct careful assessment of the potential negative impacts of automated decision-making, especially where decisions affect vulnerable populations, because the consequences of system failures affect basic survival.

> *Because Centrelink recipients are vulnerable, the consequences of poor automated systems are worse than if they affected general members of the public. (Interview 6)*

> *People's ability to eat and feed their children is at risk. Their health is at risk. When these systems fail, they have profound impacts on people. These are humanitarian risks — the system can't be allowed to fail. (Interview 21)*

> *We shouldn't be testing experimental ADM systems on people that are very vulnerable. (Interview 17)*

Principle 1.3 Human rights impact: Mandate human rights impact assessments for automated decision-making systems including AI

Human rights must be protected in the design and implementation of automated systems. This can be achieved by undertaking assessments that focus on the impact of decisions on people who are vulnerable.

> *Automated systems don't have embedded within them an intuitive and methodical rights base. Automated systems don't take into account decisions that could be reviewable,*

decisions that are discretionary [and] decisions that should account for particular groups, such as people experiencing family violence. (Interview 15)

You can't have a one-size fits all process because of the people Centrelink are dealing with and the issues they are facing, [which] are often complex and diverse. (Interview 12)

Automated systems only work when everything is as it 'should' be. But any time you have something different or a bit left field, the system will break down. This leads to people having their payments rejected, stopped and so on. (Interview 2)

Human-centred services

These principles maintain human involvement and oversight of automated social security systems, recognising that social security often deals with complex situations that may not be easily handled by automated systems alone.

Principle 2.1 Human service preservation: Maintain direct service delivery and decision-making by capable staff where needed, particularly in cases where there are complex factors

The provision of social services requires human involvement in decision-making to ensure accessibility of services and nuanced consideration of complex factors.

You may have a human need, but you are not getting a human response. If social security is a human right you would expect human decency in the process. (Interview 11)

Centrelink are no longer staffed to assist people when they come in the door. Instead, they put people on the phone and wait to contact the relevant team and leave them sitting there. (Interview 10)

Centrelink personnel should know the users personally: who they are, what their communication needs are. The system is incredibly impersonal; I would not equate impersonal, detached approach with fairness. People need humanity, and they need their dignity. (Interview 21)

Principle 2.2 Human oversight: Maintain meaningful human control over decisions that grant or limit access to social security payments

In many cases, departmental staff are no longer able to oversee or intervene in automated processes. This has led to a loss of accountability and of appropriate discretion in the application of decision-making. Meaningful human involvement and oversight of automated decision-making is essential to ensuring that decisions are fair, safe and careful.

The computer is assessing eligibility in the first instance, so there is no oversight. Or if it is checked, we don't know who the decision-maker is. There is nothing to show that a human has looked at it. (Interview 9)

The [Centrelink worker] is not the delegate. The system is the delegate. (Interview 13)

Human beings have become like automated decision-makers. This is because the human decision-makers have been stripped of discretion and are going through such a prescriptive process, that a machine might as well have made the decision. (Interview 19)

Principle 2.3 System override capability: Enable human workers to override automated processes when necessary

Staff must be given the ability to intervene in automated processes when necessary, including where errors need to be corrected or where unique situations arise that an automated system may not handle appropriately.

Within Centrelink, there are a lot of processes happening that few people knew how to stop or intervene in. (Interview 12)

Even if workers can do that [i.e. understand how the decisions was made], they can't actually fix a problem once it is identified; you have to go back through an ARO. (Interview 17)

Staff say, 'it's the computer' and can't help because they don't know. (Interview 7)

Fairness and accuracy

These principles are intended to inform fair and accurate decisions, and advocate for mechanisms that can identify and correct errors made by automated systems.

Principle 3.1 Data-matching vigilance: Continuously monitor data-matching processes for mistakes

Automated data matching must be careful, accurate and able to identify incomplete or inaccurate data. Incomplete or inaccurate data can lead to incorrect decisions about eligibility or rate of social security payments, and therefore harm the people who those processes affect.

…[the] automated system meant: if the super fund changed their name at some point, there was a double up of information – so for some people it looked like they had received double the amount of super. Debt letters were being spat out automatically in response. (Interview 12)

Where there is an automatic assessment, there is no person to check if [someone] has an exemption from a child support requirement, or if a person is actually receiving the child support. (Interview 9)

Principle 3.2 Discretion preservation: Protect and restore the discretion of human decision-makers where interpretation of legislation is required

Human discretion in decision-making processes is needed, as automated systems do not have the ability to consider all relevant factors and apply discretionary criteria appropriately.

When decisions are made through automation, they are not necessarily legally incorrect, but it may be that there hasn't been consideration of all the things that are relevant. (Interview 7)

Computers can't pick up the risk factors at play that should refer people to onward services, crisis payments. This means the person has to figure out [themselves] what services to go to and what they need. (Interview 9)

Discretionary decisions are in the power of the Secretary. I query delegating authority to an algorithm... How can you make decisions and apply discretionary criteria as a computer? This may risk being an unlawful restriction of a discretion. (Interview 17)

Inclusivity and accessibility

These principles recognise that not everyone has equal access to, or comfort with, digital systems. They stress the importance of making systems accessible to all.

Principle 4.1 Digital inclusion: Consider how to accommodate the needs of people who are unable to use digital services

Many social security recipients are affected by the 'digital divide' — for instance, they may not have access to a smartphone, computer or internet service . There is a need for solutions that ensure everyone eligible for social security can access services, regardless of their technological resources or skills.

The digital divide is critical. Many people don't have access to technology because they are in poverty: they don't have reception, they don't have internet, they don't have a smart phone. They have no single device to access an interface other than through a library/friend. (Interview 15)

If you are on social security, you don't have much cash — so access to hardware such as a computer/tablet is hard. If you do have a device, you likely have an old one that doesn't cope with the technical requirements of the system. (Interview 16)

People in internet blackspots can't access online services at all. Sometimes you can pay for a satellite, and in some places you can pay per megabyte used. In many areas you can't get internet at all. Often there is no telephone service either. (Interview 8)

Principle 4.2 Human-centred inclusive design: Adopt genuine co-design processes and disability design standards in the development of automated social security systems

Systems development and design should accommodate diverse needs and circumstances, including cultural differences, geographic factors and the unique contexts of different communities.

The system is not built to listen to people — literally. Technology can't understand cultural differences or location-based needs. It can't do proactive engagement. Instead, it is responsive and punitive. (Interview 22)

Systems are 'metro-normative': They don't factor in remoteness and particular issues dealt with in that context. (Interview 21)

There are cultural communications issues at play: What does 'care of a child' mean? What is a 'reasonable excuse' for not keeping mutual obligations? What this means will be different in different communities. (Interview 19)

Continuous improvement and stakeholder engagement

These principles focus on ongoing improvement of systems and engagement with stakeholders to ensure they meet the needs of those they serve.

Principle 5.1 Stakeholder engagement: Maintain ongoing engagement with key stakeholders to develop appropriate safeguards

Relevant stakeholders must be involved in the design and implementation of automated systems to ensure their effective and safe operation. This includes people directly affected by social security systems as well as the organisations that provide legal assistance to those who encounter issues.

> *While Services Australia does consulting around the edges, they are not looking at the experiences of people who are the most vulnerable, relying on income support. They are not taking into account the lived experience of what it is like to deal with your livelihood through automated systems. (Interview 14)*

> *There is systemic discrimination against First Nations peoples in remote and very remote areas. It is justified on the basis of cost and population numbers, and both are unacceptable. (Interview 21)*

> *There is no effort on part of government to explain to civil society how automation works because they see it as a secret. (Interview 15)*

Principle 5.2 Continuous improvement: Use feedback from customers, Services Australia decision-makers and external legal bodies to continuously improve automated systems

Automated systems must adapt and learn from errors and inefficiencies, to create more effective and user-friendly systems over time. The bodies from which feedback is sought should include the Administrative Review Tribunal, Federal Court, and Commonwealth Ombudsman.

> *Automation probably leads to more appeals because decisions are less likely to be made right and explained properly the first time around. (Interview 10)*

> *You can't fix small mistakes because you don't know what they are. You just get a blanket 'no'. It could be an incredibly easy fix — but if you are never told, you won't know how simple it is. (Interview 60)*

> *It takes more time to undo an automated error than it does to just give someone an answer in the first place. (Interview 3)*

> *A human doesn't look at [a person's] whole application from beginning to end, and the system does things in little bits and pieces. Centrelink doesn't look at the whole picture of what's happening. (Interview 12)*

Principle 5.3 Complaint mechanisms: Establish accessible feedback and complaint mechanisms for reporting technical faults and other issues with digital services

Social security recipients and advocates need clear and easy-to-use mechanisms to raise issues and seek resolution of technology-related issues that arise from social security service

provisions. This should include the option of human interaction to address complaints and facilitate access to social security entitlements.

> *There are points of contact that we are told we need to go to and that gets lost in the ether... A lack of local connection point or understanding from staff means there is constant disruption in the flow of a matter that should be relatively easy to resolve. (Interview 22)*

> *If you report fraud, or are apparently a victim of fraud, your account may get frozen and you may lose access, which affects your ability to access the system overall. (Interview 14)*

> *People hate a system that doesn't allow them to talk to someone. The further we move towards automation, maybe it will become more normalised — but I don't see this being the case when you're dealing with vulnerable people and their payments. Fundamentally they want someone to hear them more than anything else. (Interview 2)*

Transparency and accountability

These principles aim to ensure automated systems are transparent and accountable, so that decisions can be understood and challenged if necessary and regularly audited.

Principle 6.1 Transparency and traceability: Ensure automated decisions are clearly documented, transparent and traceable to facilitate understanding and administrative review

Clear explanations of how and why decisions are made in automated systems are needed to provide transparency of payment calculations and decision-making so that people and their advocates can understand or challenge decisions affecting their benefits.

> *You are arguing against a decision that you don't understand because it hasn't been explained to you. You don't get information about how and why a decision is made. (Interview 14)*

> *Debt calculations are impossible for a lawyer to understand, let alone a user. (Interview 2)*

> *Often you cannot tell which decisions are automated and which are not. (Interview 19)*

Principle 6.2 Independent auditing: Conduct regular, independent audits of ADM systems to ensure compliance with standards and legislative requirements

There is a need for external, independent auditing of automated decision-making systems, noting the current lack of public access to information about how these systems work.

> *There should be regular auditing and testing of ADM systems. The most vulnerable people should be included in the testing process. (Interview 8)*

> *The oversight body should be able to halt systems when things go wrong and have enforcement powers. It should be headed by people of influence who are entirely independent of Ministerial pressure. (Interview 17)*

Privacy and ethical use of data

The sensitive nature of information held in social security systems requires informed consent regarding use of data. Additionally, data needs to be ethically used and protected.

Principle 7.1 Informed consent: Ensure informed consent is genuine and can be revoked when appropriate

Privacy law generally requires that an organisation needs consent to collect, use or disclose personal information. Yet Services Australia's handling of customer data is not well understood by those providing their information. More attention is required to the notion of consent.

> *People have no idea the use to which the information they are providing will be put. (Interview 13)*

Principle 7.2 Data protection and ethical use: Protect collected data from potential biased and discriminatory uses

There is a need for robust data protection measures to ensure social security data is not used in ways that perpetuate discrimination or undermine rights. The role that AI and machine learning can play in perpetuating discrimination needs further attention.

> *On the system are people's driver's license, passport etc. This creates risks regarding identity theft and breaches. (Interview 15)*

> *Data breaches are a huge risk... Think about millions of people who are getting some form of Centrelink payment: your CRN follows you forever [and] data is always stored for years. This can have catastrophic implications. (Interview 16)*

System integrity

These principles underpin robust governance, with a focus on rights protection and the enforcement of ethical behaviours.

Principle 8.1: Legislative framework: Establish strong governance structures and robust protections to oversee automation in social security

Robust governance and accountability structures must underpin automated decision-making systems, especially given the lack of transparency and difficulty identifying responsible parties when issues arise.

> *Robodebt was a government policy implemented at the department level, designed by people who don't like social security recipients. They wouldn't have tried it out on people that they did like. (Interview 6)*

> *Legislation gives a discretion that then that gets delegated. Those discretions are important: They are there to protect people's rights and the rule of law. It may be convenient to just skip this, but there is no good reason why a policy should be able to operate outside the legislation — which is the end result if proper steps aren't taken along the way. There needs to be protection of law all the way through development of policy and the use of discretion. Legislative discretion mustn't be fettered through automated*

decision-making but must be treated in law as a policy and subject to similar things as policies. (Interview 17)

Principle 8.2 Checks and balances: Develop appropriate regulations and extend digital protections/rights to establish a solid framework of checks and balances

Regulation is needed to ensure the lawfulness of automated systems is secured, and to provide protections for people who are vulnerable.

Currently [Centrelink] take the position that it is always the client's fault and it's nothing to do with the system. People need to understand what is happening. Explanations need to be given [as to] what their responsibilities are, what their rights are and, where there is a problem, where they can go. (Interview 9)

An ideal system would have people in power who recognise that Centrelink is a client-focused organisation that needs to assist the vulnerable; they wouldn't just tick boxes and issues like Robodebt wouldn't arise. The culture should be one of service and not one of policing. Unless that happens, Robodebt-type things will happen. (Interview 10)

References

- Attorney-General's Department (Cth), *Review Report 2022* (February 2023)
- Attorney-General's Department (Cth), *Use of automated decision-making by government Consultation paper* (November 2024)
- Australian Human Rights Commission, *Human Rights and Technology* (March 2021)
- Department of Industry, Science and Resources, *Safe and Responsible AI in Australia* (June 2023)
- NSW Ombudsman, The New Machinery of Government: Using Machine Technology in Administrative Decision-Making (November 2021)
- Commonwealth Ombudsman, *Automated Decision-Making Better Practice Guide* (March 2020)
- *Royal Commission into the Robodebt Scheme* (July 2023)

About These Research Briefings