JOURNAL OF CRIMINOLOGY AND CRIMINAL JUSTICE

Vol.2, No.1 2026 E-ISSN. 3137-0543. P-ISSN. 3137-0535 OPEN ACCESS

RESEARCH ARTICLE // DOUBLE-BLIND PEER REVIEWED

Artificial Intelligence in Public Administration: Efficiency, Accountability and the Rule of Law

Faizan Ali[1*] **Melanie O Brien**[2]

[1] Master's International Relations: European and Asian Studies, Faculty of World Economy and International Affairs, Higher School of Economics, Moscow, Russia. Email: faizanali@edu.hse.ru

[2] State Administrative Tribunal of Western Australia, Perth, Australia. Email: melanieobrien654@gmail.com

ABSTRACT

Governments everywhere are turning to artificial intelligence (AI) to modernise how they work. Automated systems now process benefit claims, flag suspected fraud, decide which businesses get inspected, answer citizens' questions and, increasingly, help draft the decisions themselves. The appeal is mostly about efficiency: faster processing, lower costs, and a way to cope with caseloads that have long outgrown the staff available to handle them. But the qualities that make AI attractive to administrators (its speed, its scale and its air of objectivity) also put pressure on some of the oldest commitments of administrative law: that power is exercised for reasons, that someone answers for it, and that it stays within the law. This article explores that tension. Using a qualitative doctrinal and policy-analytical approach, it draws on administrative law scholarship, comparative regulatory instruments and a set of well-documented cases: Australia's "Robodebt" scheme, the Dutch childcare-benefits scandal (the *toeslagenaffaire*) and the related SyRI litigation, Michigan's MiDAS unemployment-fraud system, and data-driven welfare targeting in Pakistan and India. I argue that AI accelerates a long-running shift from discretionary, street-level bureaucracy to rule-bound, system-level bureaucracy, and that this shift changes the character of administrative error. Mistakes stop being individual and become systemic, and legal doctrines built to catch errors one case at a time struggle to keep up. The article examines the main uses of AI across government and analyses the resulting problems under three headings: efficiency and its limits as a governing value; accountability and the erosion of meaningful human oversight; and the rule of law, including legality, equal treatment, legal certainty and the right to be heard. It closes with practical recommendations to help public administrations capture the real benefits of AI without giving up the accountability and legal safeguards on which their legitimacy depends.

01 ARTICLE HISTORY

Received 01-09-2026
Accepted 02-10-2026
Published 10-10-2026

02 KEYWORDS

Artificial intelligence; public administration; automated decision-making; administrative law; accountability; rule of law; algorithmic governance; Robodebt; digital welfare state

03 CORRESPONDING AUTHOR

Faizan Ali
faizanali@edu.hse.ru

CC BY-NC 4.0 © 2026 The Author(s). Published by JuriSocial Research Center Private Limited. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International Licence (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial use, distribution and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.68740/jjccj22

1. Introduction

Public administration has always been shaped by the technologies available to it. Max Weber's model of bureaucracy, with its hierarchy of offices, written files and rule-bound procedures, was itself a technology of government. It was designed to replace personal favour and arbitrary whim with impersonal administration according to rules (Weber, 1978). In the twentieth century, attention shifted to what Michael Lipsky (1980) called "street-level bureaucracy": the caseworkers, police officers, teachers and benefits officers who turn general policy into individual decisions and who, in doing so, inevitably exercise a great deal of discretion. Something new is happening in the twenty-first century. AI systems increasingly do the work that street-level bureaucrats used to do. They assess eligibility for benefits, flag cases for fraud investigation and sort applications into priority queues. In some jurisdictions they issue administrative decisions with little or no human involvement at the point of decision.

When I first started reading about this topic, I assumed that the central story would be about technology: better models, bigger datasets, smarter software. The more I read, the clearer it became that the central story is really about law and institutions. The interesting questions are not mainly about what the machines can do. They are about who is answerable when the machines get it wrong, what "a decision" even means when no one in particular made it, and whether the legal tools we inherited from an era of paper files and human officials still work when decisions are made by systems.

The push towards automation is driven largely by efficiency. Public administrations face tight budgets, rising expectations of fast and convenient service, and caseloads that have grown much faster than staffing. AI, and machine learning in particular, offers a way of managing that pressure. An automated system can process a routine application in seconds rather than weeks. It can scan millions of records for anomalies that no team of human reviewers could ever check by hand. In principle, at least, it can apply eligibility criteria more consistently than a large and geographically dispersed workforce. International bodies, including the Organisation for Economic Co-operation and Development (OECD), have actively encouraged governments to adopt AI as part of wider digital-government and modernisation strategies (OECD, 2019, 2024).

At the same time, a growing body of evidence from academic research, investigative journalism, litigation and official inquiries shows how much harm these systems can do when they are deployed without proper safeguards. The best-known example is probably Australia's "Robodebt" scheme. Between 2016 and 2019, an automated data-matching and debt-calculation process raised debts against hundreds of thousands of welfare recipients, using an income-averaging method that the government eventually conceded, and a Royal Commission later confirmed, had no lawful basis (Royal Commission into the Robodebt Scheme, 2023). The debts were often wrong, the financial and emotional toll was severe, and testimony before the Royal Commission linked the scheme to serious psychological harm, including in cases where people took their own lives. A strikingly similar episode played out in the Netherlands. There, the tax authority's approach to policing childcare-benefit claims, including a risk-classification model that used nationality as an indicator, contributed to tens of thousands of families being wrongly treated as fraudsters and ordered to repay benefits, sometimes stretching back years (Autoriteit Persoonsgegevens, 2020; Parlementaire ondervragingscommissie Kinderopvangtoeslag, 2020). The scandal brought down the Dutch government in January 2021.

It would be comforting to treat these episodes as one-off failures caused by bad software or careless officials. I do not think that reading holds up. As I try to show in this article, they reveal structural tensions between the logic of automated administration and the legal and constitutional principles that are supposed to constrain public power. The same patterns appear in very different countries, under different political systems and with different technologies. That repetition is itself a finding.

The article is organised around three research questions:

RQ1: In what ways is AI transforming the practice of public administration, and which functions are being delegated, wholly or partly, to automated systems?

RQ2: What does this transformation mean for administrative efficiency, and are the efficiency gains as clear-cut as advocates often suggest?

RQ3: What challenges does automated administration pose for accountability and for the rule of law (understood to include legality, equal treatment, legal certainty and procedural fairness), and how might law and policy respond?

The article proceeds as follows. Section 2 reviews the literature, placing the discussion within administrative theory on bureaucratic discretion and within the newer scholarship on algorithmic governance. Section 3 explains the methodology. Section 4 sets out a conceptual framework built around the three values at stake. Section 5 surveys how AI is being used across government, including the recent arrival of generative AI. Section 6 presents the case studies in some detail. Sections 7, 8 and 9 analyse, in turn, the efficiency, accountability and rule-of-law dimensions. Section 10 compares regulatory responses across jurisdictions. Section 11 draws the threads together, Section 12 makes recommendations, Section 13 notes limitations and directions for future research, and Section 14 concludes.

The stakes are high. Public administration is where the abstract promises of law and policy become concrete facts in people's lives: a benefit granted or refused, a permit issued, a penalty imposed. As more of that work is handed to automated systems, the legitimacy of the administrative state depends on making sure that efficiency is not bought at the expense of the accountability and legality that separate lawful administration from the arbitrary use of power.

2. Literature Review

2.1 Efficiency, New Public Management and Digital-Era Governance

A second strand of the literature looks at the wider movement toward efficiency in government. The New Public Management (NPM) reforms of the 1980s and 1990s placed a focus on performance measurement, cost control and the use of private sector management techniques in the public sector (Hood, 1991). Later, Dunleavy, Margetts, Bastow and Tinkler (2006) argued that NPM was giving way to "digital-era governance", in which information technology would reintegrate fragmented services and reorganise government around citizens' needs. Governments and international organisations often frame AI as the natural next chapter in this story, a tool to speed up processing times, reduce costs and improve enforcement precision (Margetts & Dorobantu, 2019; Wirtz, Weyerer, & Geyer, 2019).

The empirical evidence, however, is more mixed. Studies of digital government have documented realised productivity improvements in routine, high-volume processes. But the evidence on service quality, citizen trust and responsiveness to individual circumstances is much less settled (Medaglia, Gil-Garcia, & Pardo, 2023; Zuiderwijk, Chen, & Salem, 2021). The literature on administrative burden provides an important correction. Herd and Moynihan (2018) show that the costs that citizens experience in their interactions with government (learning costs, compliance costs and psychological costs) are not incidental to administration, but are often the result of policy decisions. Automated systems can ease those burdens for some and exacerbate them for others, a point I return to in Section 7.

2.2 AI in the Public Administration Literature

Over the last decade, a more targeted literature has arisen around the ways in which AI differs from prior generations of e-government technology. This scholarship underscores the probabilistic nature of machine learning systems where the output is a prediction or classification based on patterns from data rather than a product of explicitly programmed rules (Veale & Brass, 2019). That feature can improve accuracy in some

settings, but it also introduces new kinds of error, opacity and bias that are different in kind from the errors of earlier, more transparent forms of automation.

Much of the most useful recent work has focused closely on how officials actually interact with algorithmic advice. Busuioc (2021) argues that the information that accountability forums need is often unavailable, unintelligible or contested, and this creates real difficulties for public accountability. In a series of experiments with Dutch respondents, Alon-Barkat and Busuioc (2023) found evidence of "selective adherence": decision-makers were more likely to follow algorithmic advice when it aligned with existing stereotypes about the people being evaluated. Peeters (2020) also shows that the relationship between officials and algorithms is not one of simple control; algorithms shape what officials notice and how they frame cases. Grimmelikhuijsen (2023) found that citizens may perceive the decision-making process as trustworthy if the rationale behind an algorithm's decision is explained, but this effect depends on the type of explanation given. Overall, these research results suggest that the common policy response of "keep a human in the loop" is far from reassuring (Green, 2022).

2.3 Administrative Law, Accountability and the Rule of Law

A third literature, grounded in administrative and public law, examines the relationship between discretion and rule of law. Traditionally, administrative law has restricted discretion through a series of well-known doctrines: decisions must be authorised by law (legality); they must be made for proper purposes and on relevant considerations; affected persons must be given notice and a chance to be heard; and decisions must be open to appeal or judicial review. Citron's (2008) groundbreaking article on "technological due process" argued that, in the United States, automated systems were already undermining each of these safeguards, often because the rules embedded in code were different from the rules written in statute and no one noticed. Later scholars have expanded on this analysis. Coglianese and Lehr (2017, 2019) suggest that machine learning can, in principle, be reconciled with American administrative law, provided that agencies can explain in general terms how their systems function and adequately validate them. Others are more sceptical. Zalnieriute, Bennett Moses and Williams (2019) highlight significant tensions between automated decision-making and the rule of law values of transparency, accountability, predictability and equality. Cobbe (2019) and Oswald (2018) explore the use of English public law grounds of review, including fettering of discretion and failure to consider relevant matters, in relation to algorithmic systems.

A recurrent question in this literature asks what happens to the moment of individual judgement. Even if human officials exercised that judgement imperfectly, the mere fact that it existed meant that someone was at least in theory looking at the particular case. Automated systems have the ability to silently erase that moment while the formal legal framework continues to assume its existence (Yeung, 2018). The risk, according to Danaher (2016), is a move toward 'algocracy', a governance where decisions are shaped by algorithms that citizens cannot meaningfully comprehend or challenge. From the perspective of those who have been subjected to automated eligibility systems in the United States, Eubanks (2018) documents how they have tended to be most intensively scrutinised and punished in poor communities.

2.4 Gaps in the Literature

This literature has three gaps that motivate the current article. First, much of the literature treats efficiency, accountability and legality as separate issues, to be analysed by separate disciplines. Public administration scholars write about efficiency and trust, lawyers write about legality and review and computer scientists write about fairness metrics. There is comparatively little work looking at how these values interact in practice, and in particular how failures of accountability and legality feed back into the efficiency case itself. Second, most of the well-known case studies are from the Global North: Australia, the Netherlands, the United Kingdom and the United States. But governments in South Asia, Africa and Latin America are introducing very large-scale data-driven systems, often in welfare targeting and identity management, and often with weaker institutional safeguards. Third, the regulatory landscape has changed dramatically since

2023, with the EU AI Act coming into effect, the adoption of the Council of Europe's Framework Convention on AI and a number of national policy shifts. The article seeks to make a modest contribution to each of these areas.

3. Methodology

The article employs a qualitative, doctrinal and policy-analytical approach, which is appropriate for a study that overlaps administrative law, public administration theory and technology policy. I used 3 different kinds of material.

The first is administrative law doctrine and scholarship on legality, procedural fairness and judicial review of automated and algorithmic decision-making. The second one is a collection of documented case studies on AI and automated decision-making in public administration. I focus mainly on Robodebt and the Dutch childcare-benefits scandal, because both have been the subject of detailed public enquiries, litigation and academic analysis. They therefore provide an unusually solid empirical basis for normative claims. I supplement these with litigation in the Dutch SyRI case, Michigan's MiDAS system, and data-driven welfare targeting in Pakistan and India, which help to test whether the patterns identified in the main cases travel across different legal and institutional settings. The third category of materials are comparative regulatory and policy instruments governing AI in government, such as the EU Artificial Intelligence Act, the Council of Europe Framework Convention, national algorithmic impact assessment frameworks and guidance from bodies such as the OECD.

The material was thematically analysed and selected on the basis of its relevance to the research questions. I read the sources to identify recurrent themes and then traced those themes through the cases and regulatory instruments. Three recurring themes were: erosion of discretion, diffusion of accountability and risk of systemic, legally significant error. They became the organising ideas of the analysis.

3.1 Limitations of the Approach

A few limitations should be stated bluntly. The method used does not permit statistical generalisation to all artificial intelligence systems in governments globally. Spectacular failures get investigated, and investigations generate documents. Quiet successes rarely do. The case selection is therefore liable to be biased toward over-representation of failure, and I have tried to keep that bias in mind when drawing conclusions. I also worked with English sources and English translations of Dutch material, so some of the subtleties of the original documents may be lost. Finally, as a graduate researcher writing from outside the agencies in question I do not have access to internal documentation of the system. Where I describe how a system worked I use what has been released by enquiries, courts and investigative reporting. Nevertheless, the approach lends itself well to the article's main purpose: to identify the recurring structural legal and administrative dynamics across the cases documented and to link them to the wider literature on accountability and the rule of law.

4. Conceptual Framework: Three Competing Values

Before going on to applications and cases, it is helpful to be clear about the three values that frame the analysis. All are contested, and the conclusions drawn depend on how one defines them.

4.1 Efficiency

In day to day policy talk, efficiency is often about doing more with less: more cases, faster, at less cost per case. Most AI business cases are based on that narrow input/output understanding. But there is a wider understanding, closer to what economists call allocative efficiency, which asks whether resources are producing the outcomes that society actually values. In this view, a system that processes cases quickly but gets a large percentage of them wrong is not efficient at all; it simply shifts costs off the agency's books and

onto citizens, courts, and future budgets. In this article I use the broader meaning partly because I think it is more defensible and partly because the case studies show that the narrow version systematically overlooks the most important costs.

4.2 Accountability

Following Bovens (2007), I understand accountability as a relationship between an actor and a forum, in which the actor is obliged to explain and justify his conduct to the forum, the forum can pose questions and pass judgement and the actor may be subjected to consequences. This definition is useful because it breaks accountability down into parts, each of which can fail. With automated systems the problem is not usually a lack of forum (courts, tribunals and ombudsmen are still there) but the difficulty of identifying the actor, the unavailability or unintelligibility of the information needed to explain the conduct, and diffusion of consequences across contractors, departments and successive ministers. This is an accountability problem at the "information" stage according to Busuioc (2021), and the cases below illustrate it.

4.3 The Rule of Law

The rule of law is notoriously hard to define. I adopt a moderate version from Fuller (1969) and Bingham (2010). According to this view, the rule of law requires at least that public power is exercised on the basis of law (legality), that laws and their application are accessible, clear and reasonably predictable (legal certainty), that like cases are treated alike without unjustified discrimination (equality before the law) and that individuals are afforded a fair opportunity to know the case against them and to respond (procedural fairness). These are not odd requests. These are the everyday working assumptions of administrative law in almost every legal system, common law systems like those of Australia, the United Kingdom and Pakistan, civil law systems like that of the Netherlands.

4.4 Interaction of the Values

The relationship between these values is often framed in terms of a trade-off: more accountability and legal safeguards means less efficiency, and vice-versa. My argument is that this framing is misleading, at least in high-stakes administration. In the cases discussed below, failures of accountability and legality had implications beyond individuals, undermining the efficiency case for the systems in question. Add up compensation, litigation, remediation and political damage and the safeguards cut to save money prove to have been the cheapest part of the project. I elaborate this argument more fully in section 11.

5. Applications of Artificial Intelligence in Public Administration

The phrase "AI in government" covers a very wide range of activities, and it is wrong to think of these as a single phenomenon. A chat bot telling you when the tax office opens raises very different questions than a model that decides whose benefits get stopped. In this section I classify current applications into five broad classes, roughly ordered in terms of the directness with which they affect individual rights.

5.1 Citizen-Facing AI and Service Delivery

The first category includes systems that are between the government and citizens. Chatbots and virtual assistants have become ubiquitous across government agencies. They answer routine questions, walk people through application processes, steer them to the right service, reduce call-centre volumes and provide 24/7 availability. Incoming letters and complaints are filtered with more advanced natural language processing to direct them to the right department or flag urgent issues. Generally these tools are lower stakes than the decision making systems discussed below as they tend to support substantive decisions rather than replace them. But they shape how people experience the state. A chatbot that dispenses confident but erroneous advice, or that cannot comprehend someone with an unusual situation or limited language skills, can leave that person worse off than if they had simply been handed over to a

human. The disadvantage tends to be borne by those who are least 'standard' in their circumstances, a thread running through this article.

5.2 Administrative Decisions and Benefits Administration

The second is much more consequential. This includes using AI or automated rules to inform or make substantive decisions, particularly in social welfare, immigration and taxation. Automated systems determine eligibility, calculate payments, identify apparent overpayments or fraud by cross-referencing data across government databases, and in some locations generate debt notices or terminate benefits with little or no individual human review beforehand. Robodebt is a perfect example. The scheme compared income people had declared to the tax office against income they had reported to the welfare agency. Where the figures did not match, it evenly averaged the annual tax figure over the fortnightly reporting periods and took the difference as a debt. The burden was then reversed to the individual to prove the debt was wrong, which would usually be the case if the agency had assessed the debt itself (Carney, 2019; Royal Commission into the Robodebt Scheme, 2023). Section 6 discusses the scheme in detail.

Note that many of the systems in this category are not "AI" in the narrow machine learning sense. Robodebt, for example, was based on relatively simple rules and data matching. I deliberately include such systems. What matters, from the point of view of the person affected, and from the point of view of administrative law, is not the sophistication of the technology, but the fact that a consequential decision was made by a system, and not by individual consideration. If anything, the legal problems that simple automation revealed are likely to be magnified by more complex and opaque machine-learning models.

5.3 Enforcement and Compliance Monitoring

The third category is using AI to support enforcement, target inspections, monitor compliance. Many tax authorities use predictive analysis to help determine which returns to audit. Regulators of food safety, workplace safety and the environment use risk models to decide which of the many businesses they regulate should be inspected first, because they can never inspect them all. Financial regulators use automated tools to detect patterns that could signal money laundering or market manipulation. These practices are typically defended as efficient ways to focus limited enforcement resources on the riskier cases and in many respects they are. But they do raise real questions, examined in Section 9, about the fairness and transparency of the risk scores that determine who has to bear the burden of investigation. Selection for scrutiny is not a neutral event, even if there is no penalty. It costs time, money and peace of mind. When the selection is skewed, so are the costs.

5.4 Internal Operations and Staff Management

The fourth category is the least visible to the public but is growing in importance: AI used inside government for things like scheduling the workforce, analysing procurement, managing documents, conducting internal audits and managing risk. These uses tend to be less directly rule-of-law problematic, because they are usually not deciding the rights or obligations of individual citizens. But they still matter, because they determine what public administration can and wants to do. For example, a system that automatically monitors the performance of caseworkers may never determine the fate of a citizen, but it can have a strong effect on caseworker behaviour. When the system rewards speed when closing cases, caseworkers close cases quickly, and the individual attention the law assumes is being provided may quietly disappear. This is one of the more 'hidden' ways in which automation changes discretion.

5.5 Generative AI and Government

The fifth category is the new one, and it is developing very fast. Since the public release of large language models in late 2022, governments have experimented with generative AI for drafting correspondence, summarising consultation responses, preparing briefing notes, translating documents and helping staff to search internal guidance (OECD, 2024). Some governments have issued guidance on how civil servants can

use these tools and some have developed their own in-house systems. The appeal is clear: most of what public officials do is reading and writing a lot of text.

Generative AI presents some unique concerns. Such systems can produce fluent, plausible text that is just wrong, a problem often referred to as "hallucination." In an administrative setting, a false statutory reference in a draft decision letter or a summary of public consultation which does not mention minority views can have real consequences. Then there is the subtler problem. Generative AI may not accurately capture the actual reasoning of the decision-maker where the reasons for a decision are drafted by generative AI. Administrative law requires reasons in part because they make the decision-maker think. If the reasons are generated afterward by a machine, that disciplining function may be lost even if the reasons seem perfectly adequate on paper. This area is still young and its legal implications deserve much more research than I can give them here, but it looks likely to be one of the central issues in the field over the next few years.

6. Case Studies

In this section we examine more closely the cases that support the analysis. I spend most time on Robodebt and the Dutch childcare-benefits scandal, and more briefly on MiDAS and on examples from South Asia. The purpose is not to repeat all the details, which are done much better in the official reports, but to highlight the features that matter most for efficiency, accountability and legality.

6.1 Robodebt (Australia)

Robodebt was the informal name for the Online Compliance Intervention run by Australia's Department of Human Services (later Services Australia) from 2016, building on an earlier pilot. Its basic mechanism was simple. Annual income data from the Australian Taxation Office was compared with the fortnightly income reported by welfare recipients to the welfare agency. Where there was a discrepancy the system averaged out the annual income evenly across the year's fortnights and worked out what the person 'should' have been paid. The difference turned into a debt.

The flaw was just as simple. Many welfare recipients don't have regular income. They work casually, seasonally or intermittently, often moving on and off benefits as their work changes. For such people the average annual income over the whole year gives a fictitious picture in which they seem to have been earning during the fortnights when they were actually unemployed. The 'debts' that emerged were often inaccurate or non-existent (Carney, 2019; Whiteford, 2021). In the manual process, staff would only have used averaging as a last resort and would have looked for actual payslips or employer records. The automated version removed most of that human checking and it moved the job of gathering evidence to the recipient, who might have to produce payslips going back several years.

Legal concerns were raised early on by welfare rights advocates, academics and the Commonwealth Ombudsman. The critical moment was 2019 in *Amato v Commonwealth* when the Commonwealth admitted in the Federal Court that a debt raised on the basis of averaging alone had not been validly made. This was followed by a class action and in *Prygodicz v Commonwealth of Australia (No 2)* (2021) the Federal Court approved a settlement involving the repayment or cancellation of debts and further compensation, a package valued at approximately A$1.8 billion. Approving it, Justice Murphy called the scheme 'a shameful chapter' in the administration of Australian social security. The Royal Commission into the Robodebt Scheme was established in 2022 and published its report in July 2023. The Commissioner, Catherine Holmes, called Robodebt "a crude and cruel mechanism, neither fair nor legal" (Royal Commission into the Robodebt Scheme, 2023, p. iii). The report contained 57 recommendations, including recommendations aimed directly at automated decision-making in government.

Three features of Robodebt are particularly pertinent here. First, the problem was not one of technical complexity; the method was easy to understand. The failure was one of legal and institutional judgement:

no one with the authority to stop the scheme took action on clear warnings that it was unlawful. Second, the scale of the scheme meant that one bad assumption led to hundreds of thousands of bad decisions. Third, the burden of proof was reversed, so that the costs of error were borne almost entirely by the people least able to bear them.

6.2 SyRI and the Childcare Benefits Scandal (The Netherlands)

The Dutch *toeslagenaffaire* took place over a longer term and is in some ways more disturbing, because it involved an aggressive anti-fraud policy, a rigid 'all or nothing' legal framework and discriminatory risk profiling. Since approximately 2013, the Dutch Tax and Customs Administration had adopted a very strict policy regarding applications for childcare benefits. Families who were suspected of being fraudsters could be ordered to pay back all benefits received, sometimes for several years, often for minor administrative errors like a missing signature. Many were labelled as having acted with intent or gross negligence, which excluded them from payment arrangements and other relief.

Algorithmic risk scoring was part of this picture. The Dutch Data Protection Authority (Autoriteit Persoonsgegevens, 2020) found that the tax authority had used nationality data, including whether an applicant was Dutch or had dual nationality, as an indicator in its risk-classification model and that this processing was both unlawful and discriminatory. In December 2021, the Authority imposed a fine of €2.75 million on the tax administration. In December 2020, the parliamentary interrogation committee's report, *Ongekend onrecht* ("Unprecedented injustice"), concluded that basic principles of the rule of law had been violated (Parlementaire ondervragingscommissie Kinderopvangtoeslag, 2020). The risk-classification system was described as a "machine of xenophobia" (Amnesty International, 2021). The government resigned in January 2021 and set up a compensation scheme for families affected, but implementation has been slow and contested.

Another similar case is the SyRI judgement. SyRI (System Risk Indication) was a Dutch government system that combined data from many public bodies to identify people who were thought to be at greater risk of benefit or tax fraud. It was mainly used in poorer neighbourhoods. In February 2020, the District Court in The Hague found that the legislation on SyRI violated Article 8 ECHR, the right to respect for private life, mainly because the system was not sufficiently transparent and verifiable and lacked adequate safeguards (*NJCM v. State of the Netherlands*, 2020). What makes the judgement significant is that this is one of the earliest instances in which a court has declared a welfare-fraud algorithm invalid on human rights grounds before a full-blown scandal had broken out.

What is striking about the Dutch case is how ordinary administrative logic, applied rigidly and at scale, became systemic injustice. The algorithm did not act by itself. It worked in a legal framework that allowed no proportionality, a culture of fraud within the organization and political pressure to be tough on benefit abuse. Technology magnified each one of these.

6.3 MiDAS (USA, Michigan)

Between 2013 and 2015, Michigan's Unemployment Insurance Agency employed an automated system, known as the Michigan Integrated Data Automated System (MiDAS), to detect and adjudicate unemployment-insurance fraud. The system made fraud determinations, often with no actual human involvement, and imposed severe penalties of up to four times the amount allegedly overpaid. A later review by the agency found that a very high percentage of the system's fraud determinations without human review were wrong. Widely reported figures suggest an error rate of around 93 per cent for the automated determinations examined. Some got their wages garnished, their tax refunds seized, and filed for bankruptcy. This led to litigation, including *Cahoo v. SAS Analytics Inc.* (2019) in federal courts and a state class action that settled for $20 million in 2022.

MiDAS confirms the pattern seen in Australia: high-volume automated adverse determinations, weak notice, a reversed practical burden of proof, and a long delay before the systemic nature of the problem

was identified. It also points to the fact that these failures are not confined to parliamentary systems or to a particular legal tradition.

6.4 Data-Driven Welfare Targeting in South Asia

The most discussed cases in the literature are from wealthy democracies with strong courts, active media and well-resourced civil society. It is worth asking what happens where those conditions are less strong. Two South Asian examples are instructive, though I treat them with more caution because the documentary record is thinner.

The Benazir Income Support Program (BISP) in Pakistan, one of the country's largest social protection schemes, employs a proxy means test based on data from the National Socio-Economic Registry and the National Database and Registration Authority (NADRA). In December 2019 the government announced that over 800,000 beneficiaries would be removed from the program after data analysis identified wealth indicators such as foreign travel, owning vehicles, higher telephone bills or working for government. Some of those removed likely did not qualify. But civil society groups and media reports also raised concerns that truly poor households had been removed based on incomplete or outdated data, and that the process for contesting removal was unclear and difficult for people with low literacy or limited access to government offices. The government then did another national survey to update the registry. The merits of that particular exercise aside, it is a lesson in how quickly data matching can lead to large-scale exclusion decisions, and how much depends on the quality of the underlying data and the availability of redress.

The linking of welfare entitlements in India, particularly subsidised food rations under the Public Distribution System, to Aadhaar biometric authentication has been the topic of extensive debate. In 2017, some deaths from starvation reported in the media in Jharkhand were tied by activists to exclusions based on authentication failures (worn fingerprints, poor connectivity, mismatched records). Researchers and journalists documented cases where authentication failures resulted in people being denied rations. In *Justice K.S. Puttaswamy (Retd.) v. Union of India* (2019), the Supreme Court upheld the Aadhaar scheme for welfare purposes but limited its use in other contexts. The Indian experience demonstrates that systems designed primarily to reduce fraud and leakage can lead to errors of exclusion, and that these errors are most likely to be experienced by the poorest and most marginalised.

These examples suggest that the problems identified in the Global North are not unique to it, and may be more severe where judicial review is slower, legal aid is scarce and civil society has fewer resources to bring systemic problems to the fore. They also point to a gap in the literature that has just begun to be filled.

6.5 Lessons from the Cases

Examining these cases reveals a number of common features. In each case, systems designed to promote efficiency or to curb fraud produced large numbers of adverse decisions. In all cases, those affected were disproportionately poor, precariously employed or from minority backgrounds. In each case the individual had to bear the practical burden of proving the system was wrong. In each instance, it took years, and often some combination of litigation, journalism, advocacy and formal inquiry, before the systemic nature of the problem was officially recognised. And in each the ultimate cost to the state in financial, political and reputational terms was far greater than any savings the system had achieved. The following sections are informed by these common features.

7. Efficiency and Its Limits

7.1 The Efficiency Argument for AI

The efficiency argument for AI in public administration rests on four distinct claims. First, automated processing is faster than manual processing, especially for routine, high-volume decisions. And the second

is that automation decreases the cost of each additional case so that administrations can absorb growing caseloads without hiring proportionally more staff. Third, data matching and predictive analytics can uncover errors, fraud or non-compliance that would never be found through manual review. Fourth, consistent application of eligibility criteria to a large caseload reduces the variation that inevitably arises among individual human decision makers.

There is good evidence for the first two, at least where systems are well designed. Automation of simple decisions, such as benefit renewals or permit applications, has in many documented cases reduced processing time from weeks to hours or minutes, freeing up human caseworkers to focus on the complex cases that really require judgement (Engstrom, Ho, Sharkey, & Cuéllar, 2020; Wirtz et al., 2019). I want to be clear that I do not write off these gains. For someone waiting for a payment they urgently need, speed is not a trivial value; delay can itself be a grave injustice. The question is not if efficiency is relevant or irrelevant but rather what we mean by efficiency and at whose expense.

7.2 When the Efficiency Case Fails

For the rest, the evidence is much more mixed, and the limits of efficiency are not always recognised in the AI-adoption literature.

First, the claim that automated data matching can reliably pick out fraud or overpayment depends entirely on the quality of data and on the validity of the method of drawing conclusions from it. Robodebt is an example of how something that appears to be efficient can systematically be wrong when the underlying assumption is wrong. The assumption was that income was evenly spread over the year. That assumption doesn't hold in the case of people with irregular or seasonal work, who are wildly over-represented among welfare recipients. The system was thus least accurate for the population it was designed to police.

Second, and more fundamentally, efficiency in the narrow sense of speed and cost is not the only value public administration is supposed to serve, and arguably not the main one. Accuracy, fairness and individual consideration have long been principles of administrative law that can justify slower or more expensive processes, particularly where a wrong decision would have serious consequences, as is often the case in welfare, immigration and law enforcement. If a system cuts average handling time but does so by increasing the number or severity of wrongful adverse decisions made against vulnerable people, then the system has not generated any meaningful efficiency gain. Once you add in the human and social costs of error, including financial hardship, psychological harm and the administrative and judicial costs of correction, the books look very different. Governments had to spend huge amounts of money remediating and compensating for both Robodebt and the Dutch scandal, and they also had to establish royal commissions, hold parliamentary enquiries and engage in years of litigation. None of those costs were in the original business case.

Third, gains in efficiency from automation are often spread unevenly among the population in ways that can deepen existing inequality. Systems built with the average case in mind tend to process the lives of people with stable, well-documented lives (regular jobs, conventional family structures, straightforward immigration status) more quickly and accurately. The more complicated your life (irregular work, non-traditional households, complex migration histories), the more likely you are to be misclassified. These circumstances were never well represented in the data or the decision rules. The effect is to place administrative burden on already disadvantaged groups (Herd & Moynihan, 2018; Eubanks, 2018). Similarly, Peeters and Widlak (2018) describe this as a 'digital cage', in which those with errors on file find themselves trapped as each system relies on the data held by every other system and no one has the authority or the means to change it.

7.3 Efficiency Revisited

My takeaway from this isn't to throw out efficiency as a goal, but to measure it over the full life of a system and across everyone it affects. A more honest efficiency assessment would include the expected costs of

error, the costs of correcting it, and the distribution of those costs among different groups. It would also understand that some kinds of "inefficiency," like requiring a human to check an adverse decision before it's sent, are insurance against catastrophic and costly failure. Thus conceived, safeguards are not a tax on efficiency. They are part of what makes efficiency gains real and lasting.

8. Accountability Problems

8.1 Discretion as Limited Human Judgement

The conventional notion of accountability in public administration has been based on a determinate official who exercises judgement in the application of law and policy to an individual case and who can be called to account for it, through internal review, through external bodies such as ombudsmen, or through the courts. Moving decisions into automated systems threatens the structure in at least two ways.

The first is when you have a fully automated system where an adverse decision is generated and sent out without anyone reviewing it first. There is just no human judgement to answer for at the moment of decision. Accountability is pushed to a later stage of review or appeal and the onus is on the individual affected to identify and contest an error which may not be apparent to a non-specialist. This was compounded in Robodebt by the reversal of the onus of proof: the individual had to disprove a debt the computer had calculated, rather than the agency having to justify it.

The second occurs even when a human official formally reviews an automated recommendation. There is a large body of research now on automation bias, the tendency for people to trust what automated outputs tell them. This tendency is strongest when caseloads are heavy and time pressures are high – precisely the conditions that lead agencies to adopt AI in the first place. In such cases, the continuing formal authority of human beings may become largely symbolic. Alon-Barkat and Busuioc (2023) complicate the picture further: officials aren't deferential to algorithms across the board, but selectively, more willing to follow them when the output aligns with what they already believed. Green (2022) surveys policies that mandate human review of government algorithms, and finds that they often give a false sense of security, legitimising flawed systems without actually correcting them. Reports on Robodebt suggest that caseworkers had little practical ability to examine the income-averaging calculations behind the debt notices they were processing, given their workloads (Royal Commission into the Robodebt Scheme, 2023).

8.2 Explainability, Reasoning and Transparency

A second accountability problem concerns whether automated systems can satisfy the duty to give reasons that administrative law generally imposes upon public decision makers. Reasons serve several purposes. They enable affected people to understand a decision and, if appropriate, to challenge it. They allow review bodies to examine whether the decision was properly taken. And they discipline the decision-maker, who has to give a rational basis for what they have done.

Machine learning systems, especially those relying on complex statistical models, may generate outputs that are not easily transformed into the kinds of individual, interpretable explanations that the law demands (Pasquale, 2015; Selbst & Barocas, 2018). The problem is compounded when developers or vendors treat the model as a commercial secret and refuse to explain how it works. A part of this can be solved through technical work on explainability, e.g. through counterfactual explanations that tell a person what would have had to be different for the outcome to change (Wachter, Mittelstadt, & Russell, 2018), but this does not close the gap between a statistical explanation and a legal justification fully. "Which features most influenced the model's score?" is not the same question as "why was this decision reasonable and lawful?"

Both Robodebt and the Dutch case had serious transparency problems. People in each country said they had received notices, automatically generated, which gave little or no meaningful explanation as to how

the debt or fraud finding had been reached. Without that explanation it was impossible to identify errors or launch a proper challenge. It is worth stressing that the method in Robodebt wasn't technically opaque at all, it was simple averaging. The opacity was institutional, not algorithmic. No one was told what had been done and the agency did not respond to challenges to the method itself. This is a useful reminder that explainability is not just a technical problem to be solved with better software, but a question of administrative practice and legal obligation.

8.3 Oversight, Redress and the Limitations of Individual Review

A third problem concerns the extent to which existing oversight and redress mechanisms (ombudsmen, tribunals and courts) can identify and rectify systemic errors in automated systems before they affect large numbers of people (Hertogh & Kirkham, 2018). Such mechanisms are usually constructed around individual complaints and appeals. That model does a pretty good job of catching isolated errors of individual officials. It is not good at detecting a flaw in an automated approach that affects hundreds of thousands of people at the same time and in the same way.

There are two reasons behind this. Each affected person feels the error as a personal grievance, not as a sign of systemic fault. And automated decision-making can occur at a speed and scale that exceeds the ability of oversight bodies to conduct investigations before great harm has been done. Individual appeals can also be ruled on by tribunals in the person's favour, without this being taken as a signal that the whole method is wrong. In Robodebt, tribunals overturned many individual debts on review, but the scheme kept going for years. The scheme's unlawfulness was formally established and the program came to an end only through the combined effect of individual legal challenges, investigative journalism, advocacy by welfare rights organisations and ultimately a test case in the Federal Court. By then several hundred thousand people had debts registered against them.

8.4 The Many Hands Problem

One final matter of accountability should be noted. Automated systems are typically designed and operated by multiple actors: policy officials who define the goals of the system, lawyers who interpret the governing statute, data scientists and contractors who build the system, managers who set performance targets, and frontline staff who process the outputs. When things go wrong, each can reasonably claim that the crucial decision was made by someone else. This is a version of what accountability scholars refer to as the 'problem of many hands' (Bovens, 2007). A lot of time was spent at the Robodebt Royal Commission figuring out who knew what and when, and how legal advice that questioned the legality of the scheme was ignored. Automation does not create the problem of many hands, but it increases the number of hands and obscures their contributions. As Kroll et al. (2017) argue, accountability for algorithms must therefore be embedded in systems from the start, through documentation, logging and verifiable procedures, rather than reconstructed after the event.

9. Threats to the Rule of Law

9.1 Legality and Automated Ultra Vires Action

A fundamental condition of the rule of law is that administrative action must be lawful. The decision-maker, whether a human being or an automated process, shall act within the powers conferred on it by the relevant statute or regulation. Robodebt is a sharp example of an automated approach overtaking those powers in ways that aren't necessarily apparent until someone gets a chance to test the method against the law. The Commonwealth's concession in Amato, and more recently the findings of the Royal Commission, made clear that income averaging alone was not a lawful basis for calculating a debt under the social security legislation. In effect, the scheme had been generating illegal debts for years before it was formally determined.

This shows a unique hazard of automated delivery. When an algorithmic approach is embedded in software and deployed at scale, it applies that logic consistently and continuously, across huge numbers of cases. There is no point where a thinking official, looking at an unusual case, might stop and ask himself if the method is really lawful. An unlawful method can therefore generate unlawful results on a scale and in a manner far beyond what is possible through individual human decisions, even taking into account the fact that human decision-making is subject to its own systemic biases and errors. Zalnieriute et al. (2019) make a similar point, noting that the problem is often not that code applies the law incorrectly in a single case, but that the translation of law into code was flawed in the first place and that flaw is then repeated automatically.

9.2 Non-Discrimination and Equal Treatment

A second rule-of-law concern is equality before the law: like cases should be treated alike, and administrative classifications should not differentiate on prohibited grounds. The Dutch case illustrates how an automated risk system can lead to systematically unequal treatment. The Data Protection Authority found that the risk model of the tax authority had used nationality as an indicator, which caused a disproportionate targeting of families with a migration background for increased scrutiny and recovery of benefits (Autoriteit Persoonsgegevens, 2020). Later, the Dutch authorities admitted that the system had worked in a discriminatory way, in violation of Dutch and European non-discrimination law.

The broader algorithmic-fairness literature warns that discrimination does not require the explicit use of a protected characteristic. Variables correlated with protected characteristics such as postcode, language, surname and employment history or household composition can act as proxies and have the same effect (Barocas & Selbst, 2016). This implies that simply removing sensitive variables from a model is not sufficient. Automated systems may replicate and formalise patterns of discrimination that would almost certainly be recognised as unlawful if adopted openly as administrative policy. Worse, the formalisation can make the discrimination appear objective. If an official is a human being and acts unjustly, at least there is a possibility that someone will notice and object. When a model does this, the unfairness is masked by authority of numbers.

9.3 Legal Certainty and Predictability

A third rule-of-law value, legal certainty, requires that persons be able to reasonably predict the legal consequences of their conduct and circumstances, and that administrative decisions be taken on ascertainable and stable criteria. This is a particular problem for machine-learning systems, particularly those that are retrained from time to time on new data. The criteria that the system effectively uses may change over time as the model is updated, without those changes being announced, or in some cases even fully understood, by the agency. This is what the technical literature refers to as model or performance "drift". Where drift influences the practical criteria used to decide eligibility, enforcement risk or other matters, affected people and their advisers may have no reliable way of knowing what criteria will be actually applied to them. This is precisely the predictability that the rule of law is designed to provide (Fuller, 1969).

There is an irony of interest here. One of the main points in favour of automation is that it makes for a more uniform administration. That is true in a narrow sense: a model applies the same logic to every case at a given moment. But consistency between cases at one point in time is not the same thing as stability and transparency over time. A system with unknown, shifting criteria, even one that is perfectly consistent at any given moment,, can satisfy the letter but not the spirit of equal treatment or legal certainty.

9.4 Fair Procedures and the Right to Be Heard

A fourth closely related concern is procedural fairness. In common law systems this is often expressed in terms of the principle *audi alteram partem*, "hear the other side": before a decision adversely affects someone they should know the case against them and have a fair opportunity to respond. The opposite is

often the case with automated systems. First the decision is taken, automatically, and only then, if at all, comes the chance to be heard, in the form of an appeal that the person has to initiate. In MiDAS, people were often found to have committed fraud on the basis of questionnaires they did not know they had been sent or did not understand. Robodebt involved the raising of debts and, in many cases, the referral of debts to debt collectors without recipients having any meaningful opportunity to challenge the calculation.

The courts have started to act. European data protection law gives individuals a qualified right not to be subject to decisions based solely on automated processing which produce legal or similarly significant effects (General Data Protection Regulation, art. 22). The Court of Justice of the European Union gave a broad interpretation to that provision in *SCHUFA Holding* (2023), holding that an automated credit score itself could amount to a "decision" if it was heavily relied on by a third party. In *Dun & Bradstreet Austria* (2025) the Court considered what a person is entitled to be told about the rationale of an automated decision. These are rulings on data protection, not on administrative law in the strict sense, but they are indicative of an emerging judicial acceptance that procedural rights need to be adapted to automated decision-making, and not simply assumed to carry over.

10. Comparative Responses to Regulation

The regulatory response to these challenges has varied widely across jurisdictions. Here I provide an overview of the most important developments, with special emphasis on changes since 2023.

10.1 European Union

The most comprehensive response to date is the EU Artificial Intelligence Act (Regulation (EU) 2024/1689) which came into force in August 2024. It expressly bans some practices, including social scoring by public or private actors resulting in unjustified or disproportionate adverse treatment. It identifies as "high-risk" those artificial intelligence systems used by or on behalf of public authorities to determine eligibility for essential public assistance benefits and services, or to grant, reduce, revoke or reclaim them (Annex III). High-risk systems are subject to requirements on risk management, data governance, technical documentation, oversight by humans, as well as on accuracy and transparency. Public bodies deploying high-risk systems are also required to conduct a fundamental rights impact assessment prior to deployment (art. 27) and affected individuals have a right to an explanation of the role played by the AI system in certain high-risk decisions (art. 86). The Act's particular approach to public benefits clearly reflects lessons learned from the Dutch scandal, which was within the EU and gave the legislation a political boost. The obligations are being phased in, with most high-risk requirements initially due to apply from August 2026, although the European Commission has since proposed to delay some of these deadlines. It remains to be seen how effective the Act will be in practice and it will depend immensely on enforcement and on national supervisory authorities.

10.2 Council of Europe

In May 2024, the Council of Europe adopted the Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, the first legally binding international treaty in this field. It was opened for signature in September 2024 and has been signed by various European and non-European states. The Convention requires parties to ensure that activities involving AI respect human rights, democracy and the rule of law, through transparency, oversight, accountability and access to effective remedies (Council of Europe, 2024). It is stated at a very high level of generality, but matters symbolically because it names the rule of law explicitly as something that AI governance must protect.

10.3 The Netherlands and Australia

There have been big national responses to the two major scandals so far. The Dutch government set up a compensation scheme for the affected families and established a national public algorithm register in December 2022 disclosing the algorithms used in government decision-making; it also gave the Dutch Data

Protection Authority a coordinating role in supervising algorithms from 2023. It also suspended or limited certain automated risk profiling methods in benefits administration, pending further review.

In Australia, the Royal Commission's 57 recommendations included calls for legislative reform to provide a consistent legal framework for automated decision-making in government, clearer statutory authority before automated debt-recovery methods are used, a body to monitor and audit automated decision-making, and changes to ensure that the onus of proof in debt-recovery matters does not fall unfairly on recipients (Royal Commission into the Robodebt Scheme, 2023). Most of these recommendations were accepted or accepted in principle by the Australian Government in its formal response later in 2023 (Australian Government, 2023) and work on a whole-of-government framework for automated decision-making has continued in the meantime.

10.4 Canada, France and the United Kingdom

Other jurisdictions have opted for a more targeted approach. Canada's Directive on Automated Decision-Making, first published in 2019 and subsequently amended, requires that federal institutions carry out an Algorithmic Impact Assessment before deploying an automated decision system, with safeguards proportionate to the level of impact assessed, including requirements for human involvement, peer review, notice and explanation (Government of Canada, 2019). The French Digital Republic Law of 2016 provides that individuals subject to an individual administrative decision taken on the basis of algorithmic processing shall be informed of this, and be able to obtain on request the main rules and characteristics of the processing. The UK adopted an Algorithmic Transparency Recording Standard that became mandatory for central government departments in 2024, and has since amended its data protection regime, with the Data (Use and Access) Act 2025 changing the rules on automated decision-making. Some of the UK changes loosen protections against decisions made solely by automated systems, and how they work in practice will need careful watching, critics have argued.

10.5 United States and Pakistan

Federal policy in the United States changes with each administration. In 2024, the Office of Management and Budget issued Memorandum M-24-10 to establish minimum risk-management practices for federal agencies' "rights-impacting" and "safety-impacting" uses of AI (Office of Management and Budget, 2024). It was replaced in 2025 by memorandum M-25-21, which emphasised the need to accelerate AI adoption while maintaining a category of "high-impact" AI subject to minimum practices (Office of Management and Budget, 2025). As the MiDAS cases show, though, most of the important regulation of algorithmic decision-making in the US still comes through litigation, state law and constitutional due process.

Pakistan can serve as a useful comparison as a developing country at an earlier stage of regulation. In 2025, the federal government adopted a National Artificial Intelligence Policy, primarily to build AI capacity, skills and infrastructure, and to encourage adoption across the government and economy (Ministry of Information Technology and Telecommunication, 2025). The policy acknowledges ethical and responsible use issues, but it does not yet provide binding rights-based safeguards for automated administrative decisions, like those in the EU or Canada. There are however constitutional guarantees in Pakistan that are directly relevant such as the right to a fair trial and due process (Article 10A) and to equality before the law (Article 25) and a well-developed body of administrative law jurisprudence on reasoned decisions and natural justice. Thus, a promising path for Pakistan may be to build on these existing constitutional and administrative law foundations rather than await a comprehensive AI statute.

10.6 Convergence Trends and Remaining Gaps

There is across these jurisdictions a discernible, though incomplete, convergence on a number of principles: mandatory impact assessment before deployment of high-stakes automated systems, disclosure of the existence and, to different degrees, the logic of automated decisions, human supervision requirements, and enhanced procedural protections, including clarified burdens of proof, for people

impacted by automated adverse decisions. But at the same time, major gaps remain. Formal transparency and appeal rights are only useful if people can actually use them, and those most often subject to high-stakes automated decisions (people living in poverty, migrants, people with disabilities, people with low digital proficiency) are often least able to do so. Most regimes also continue to focus on individual rights rather than on systemic oversight where the real weakness lies, as argued in Section 8.

11. Discussion

The analysis above implies that the challenges that AI poses to public administration cannot be adequately addressed by treating efficiency, accountability and the rule of law as separate concerns to be balanced against one another. But the cases reveal something more interesting: that failures of accountability and legality directly undermine the efficiency case for automation. The costs of fixing systemic errors – in compensation, litigation, political and reputational damage, and the diversion of administrative resources into remediation – can far outweigh the original savings claimed. Robodebt was designed to collect money. Other than the human cost, it cost the Commonwealth many times more than it ever legitimately recovered. Seen this way, accountability and rule-of-law protections are not efficiency constraints to be minimised. These are the conditions for real efficiency through the whole life of a system.

The analysis also indicates that the doctrines developed by administrative law to constrain individual discretion (procedural fairness, reasoned decision-making and case-by-case review) are necessary but not sufficient. These are doctrines made for errors and abuses that occur case by case. Automated systems can generate errors that impact huge numbers of people simultaneously, often before individual review mechanisms can detect the underlying flaw. Administrative law must develop doctrines and institutions calibrated to systemic review: mechanisms that review a system as a whole, not simply the individual decisions it makes. Some encouraging developments are already in place, including ombudsman own-motion investigations, audit office reviews and the pre-deployment impact assessments now required in several jurisdictions. But these are rarely associated with the courts or given the force of law. We can describe the challenge this way: Administrative law needs to shift some of its focus from the moment of decision to the moment of design.

A third point concerns timing. The comparative evidence suggests that so far, the regulatory responses have been largely reactive, appearing after high-profile failures rather than in anticipation of them. That pattern is disturbing, however, as it indicates that similar failures may be occurring, undetected, in areas of administration or jurisdictions that receive less public scrutiny. This concern is especially sharp in the Global South where the institutions that revealed the failures of Australia and the Netherlands – for example, independent courts capable of handling complex litigation, well-funded investigative journalism and organised welfare advocacy – may be weaker or slower. The South Asian instances discussed in Section 6 indicate that this is not a theoretical concern.

Lastly, I want to push back a little against a framing that is common in this field. The debate around AI in government is largely framed in terms of whether machines or humans make better decisions. That's a good question, but it's also a misleading one. The cases considered here are not, at bottom, stories of machines doing better or worse than humans. These are stories of institutions using technology to do things that would have been seen as unacceptable if done openly by people: raising debts without evidence, profiling families by nationality, punishing people without hearing them. The technology made these practices speedier, bigger and less visible. So the more useful question is not "human or machine?" but "what kind of institution is using this machine, under what legal constraints, and who is overseeing it?"

12. Recommendations

The following recommendations are made to public administrators, legislators and oversight bodies, based on the analysis presented above. They are intentionally pragmatic and most of them could be implemented within existing legal structures.

1. Independent legal clearance before deployment. Agencies must secure independent legal certification that the decision method is authorised by the governing statute prior to deploying any automated system that will materially affect benefits, penalties or other significant interests. This directly addresses the kind of *ultra vires* risk exemplified by Robodebt, where what appeared to be an efficient approach was found to be beyond statutory authority only after it had been operating at scale for years. Where legal advice throws up doubts about legality, that advice should be brought to the attention of the responsible minister and oversight bodies.

2. Mandatory algorithmic impact assessments including fairness testing. Agencies should be required to carry out algorithmic impact assessments before they roll out automated systems in high stakes contexts. These should include specific testing of disparate impact across protected characteristics and their plausible proxies, with particular attention to variables such as nationality, postcode or employment history, as the Dutch case demonstrated. Assessments should be published, if only summarised.

3. The state should retain the burden of proof. Laws and regulations should make clear that the burden of proof lies with the administering agency when it makes automated adverse decisions like debts, benefit terminations, and fraud findings, not with the individual. This is consistent with the view of the Robodebt Royal Commission that the usual evidentiary burden in agency-initiated debt recovery should not be reversed simply because a computer generated the initial finding.

4. Dedicated systemic-review functions. Agencies should establish and adequately fund internal systemic-review units, independent of individual appeals, to monitor automated systems for patterns of error or disparate impact across the entire affected population. These reviews should automatically include automated determinations overturned by tribunal decisions so that patterns of successful appeals raise questions about the underlying method.

5. Real, not symbolic, human oversight. Concrete safeguards should make human monitoring requirements real: caseloads that allow for real review rather than rubber-stamping, regular audits of override rates, and training that specifically addresses automation bias and selective adherence. A system of oversight where reviewers very rarely disagree with the machine is a red flag, not a victory.

6. Public algorithm registers. Public registers of algorithms used in significant administrative decisions should be more widely adopted, as are now in operation in the Netherlands, France and the United Kingdom. Registers should include information on purpose, legal basis, data sources, impact assessments and points of contact, so that they can be examined by researchers, oversight bodies and affected individuals.

7. Continual re-validation and disclosure of changes. Models drift, so agencies should be required to revalidate deployed systems periodically against current outcome data, rather than relying solely on testing prior to deployment. Material changes to the model or decision logic of the system must be publicly disclosed.

8. Clear rules for generative AI in decision making. Governments should set clear rules around the use of generative AI in decision-making in the administrative context, including a prohibition on the use of such tools to generate the reasons for a decision without real review by the relevant decision-maker, and requirements to verify any legal or factual assertions they make.

9. Accessible redress for those who need it most. Agencies must ensure that complaint mechanisms, appeals and legal aid or advocacy support are genuinely accessible to people with limited knowledge of the internet, language skills or financial means, as the burdens of automated error fall disproportionately on disadvantaged groups. Offline channels should still be there. Formal rights to challenge a decision are of little use if people cannot take advantage of them.

10. Building on existing law in developing countries. In countries such as Pakistan, which are currently working on building data-driven governance before comprehensive AI regulation, courts, ombudsmen and legislators should proactively apply existing constitutional guarantees of due process and equality, and established principles of natural justice, to automated administrative decisions. Social protection programs that use data matching should make their exclusion criteria transparent and provide simple, local, face-to-face processes for appeals.

13. Limitations and Future Research Directions

There are some limitations of this study that also suggest directions for future work. The case selection is biased toward well-documented failures, as discussed in Section 3, so the article may understate the number of automated systems in government that do reasonably well. It would be very useful to do comparative research on successful deployments and to identify the institutional characteristics which differentiate them from failures. Second, the evidence from South Asia presented here is limited and is largely based on secondary sources. There is a great need for much more detailed empirical work, including interviews with officials and affected citizens, on how data driven welfare targeting works in Pakistan, India and other developing countries. Third, the legal implications of generative AI in administrative decision-making are only just beginning to be understood; questions of reason-giving, record-keeping and the delegation of decision-making authority to such tools deserve sustained doctrinal attention. And finally, the regulatory landscape is changing rapidly. The real impact of the EU AI Act, the Council of Europe Convention and national reforms will only be seen as they are implemented and enforced in the coming years, and longitudinal studies tracking that implementation would contribute a great deal to the field.

14. Conclusion

Artificial intelligence offers public administrations real opportunities to manage growing caseloads, reduce delays and target scarce enforcement resources more effectively. The analysis in this article shows, however, that these gains cannot be assessed in isolation from their effects on accountability and the rule of law. When automation is pursued without a firm legal basis, meaningful human oversight and robust systemic review, it can cause serious harm on a very large scale. Robodebt and the Dutch childcare-benefits scandal both show this vividly, and MiDAS and the South Asian examples suggest the pattern is not confined to any one country or legal tradition. In every case examined, the eventual financial, human and political costs far exceeded whatever efficiency gains were originally claimed. Accountability and legality, in other words, are not obstacles to efficient public administration. They are what make efficiency gains genuine and durable rather than illusory and, in the end, reversed at much greater cost.

The transformation of public administration by AI is, in large part, an intensification of the long-running migration of discretion from individual officials to the systems within which they work. That migration began long before AI, but AI accelerates it and changes its character, because of the scale, speed and opacity of modern machine-learning systems. Meeting the resulting challenges will require administrative law and public policy to develop new doctrines and oversight institutions designed for systemic rather than purely individual error. It will also require humility on the part of governments about what automated systems can and cannot know about the lives of the people they govern. If that can be achieved, the substantial benefits AI offers public administration can be realised within, rather than at the expense of, the accountability structures and rule-of-law guarantees that give public administration its legitimacy.

References

Alon-Barkat, S., & Busuioc, M. (2023). Human–AI interactions in public sector decision making: "Automation bias" and "selective adherence" to algorithmic advice. *Journal of Public Administration Research and Theory, 33*(1), 153–169. https://doi.org/10.1093/jopart/muac007

Amnesty International. (2021). *Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal* (EUR 35/4686/2021). Amnesty International.

Australian Government. (2023). *Government response to the Royal Commission into the Robodebt Scheme*. Commonwealth of Australia.

Autoriteit Persoonsgegevens [Dutch Data Protection Authority]. (2020). *Belastingdienst/Toeslagen: De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag* [Tax and Customs Administration/Benefits: The processing of the nationality of applicants for childcare benefits]. Autoriteit Persoonsgegevens.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(3), 671–732.

Bingham, T. (2010). *The rule of law*. Allen Lane.

Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal, 13*(4), 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x

Bovens, M., & Zouridis, S. (2002). From street-level to system-level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review, 62*(2), 174–184. https://doi.org/10.1111/0033-3352.00168

Bullock, J. B. (2019). Artificial intelligence, discretion, and bureaucracy. *The American Review of Public Administration, 49*(7), 751–761. https://doi.org/10.1177/0275074019856123

Busch, P. A., & Henriksen, H. Z. (2018). Digital discretion: A systematic literature review of ICT and street-level discretion. *Information Polity, 23*(1), 3–28. https://doi.org/10.3233/IP-170050

Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review, 81*(5), 825–836. https://doi.org/10.1111/puar.13293

Cahoo v. SAS Analytics Inc., 912 F.3d 887 (6th Cir. 2019).

Carney, T. (2019). Robo-debt illegality: The seven veils of failed guarantees of the rule of law? *Alternative Law Journal, 44*(1), 4–10.

Citron, D. K. (2008). Technological due process. *Washington University Law Review, 85*(6), 1249–1313.

Cobbe, J. (2019). Administrative law and the machines of government: Judicial review of automated public-sector decision-making. *Legal Studies, 39*(4), 636–655. https://doi.org/10.1017/lst.2019.9

Coglianese, C., & Lehr, D. (2017). Regulating by robot: Administrative decision making in the machine-learning era. *Georgetown Law Journal, 105*(5), 1147–1223.

Coglianese, C., & Lehr, D. (2019). Transparency and algorithmic governance. *Administrative Law Review, 71*(1), 1–56.

Council of Europe. (2024). *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* (CETS No. 225). Council of Europe.

Danaher, J. (2016). The threat of algocracy: Reality, resistance and accommodation. *Philosophy & Technology, 29*(3), 245–268. https://doi.org/10.1007/s13347-015-0211-1

Dun & Bradstreet Austria, Case C-203/22 (Court of Justice of the European Union, 27 February 2025).

Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2006). New public management is dead—Long live digital-era governance. *Journal of Public Administration Research and Theory, 16*(3), 467–494. https://doi.org/10.1093/jopart/mui057

Engstrom, D. F., Ho, D. E., Sharkey, C. M., & Cuéllar, M.-F. (2020). *Government by algorithm: Artificial intelligence in federal administrative agencies*. Report submitted to the Administrative Conference of the United States.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Official Journal of the European Union*, L 119, 1–88.

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*, L, 2024/1689.

Fuller, L. L. (1969). *The morality of law* (Rev. ed.). Yale University Press.

Government of Canada. (2019). *Directive on automated decision-making* (as amended). Treasury Board of Canada Secretariat.

Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review, 45*, 105681. https://doi.org/10.1016/j.clsr.2022.105681

Grimmelikhuijsen, S. (2023). Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Administration Review, 83*(2), 241–262. https://doi.org/10.1111/puar.13483

Herd, P., & Moynihan, D. P. (2018). *Administrative burden: Policymaking by other means*. Russell Sage Foundation.

Hertogh, M., & Kirkham, R. (Eds.). (2018). *Research handbook on the ombudsman*. Edward Elgar Publishing.

Hood, C. (1991). A public management for all seasons? *Public Administration, 69*(1), 3–19. https://doi.org/10.1111/j.1467-9299.1991.tb00779.x

Justice K.S. Puttaswamy (Retd.) v. Union of India, (2019) 1 SCC 1 (Supreme Court of India).

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review, 165*(3), 633–705.

Lipsky, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public services*. Russell Sage Foundation.

Margetts, H., & Dorobantu, C. (2019). Rethink government with AI. *Nature, 568*(7751), 163–165. https://doi.org/10.1038/d41586-019-01099-5

Mashaw, J. L. (1983). *Bureaucratic justice: Managing social security disability claims*. Yale University Press.

Medaglia, R., Gil-Garcia, J. R., & Pardo, T. A. (2023). Artificial intelligence in government: Taking stock and moving forward. *Social Science Computer Review, 41*(1), 123–140. https://doi.org/10.1177/08944393211034087

Ministry of Information Technology and Telecommunication. (2025). *National Artificial Intelligence Policy 2025*. Government of Pakistan.

NJCM et al. v. State of the Netherlands (SyRI), ECLI:NL:RBDHA:2020:1878 (District Court of The Hague, 5 February 2020).

OECD. (2019). *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449; amended 2024). OECD Publishing.

OECD. (2024). *Governing with artificial intelligence: Are governments ready?* (OECD Artificial Intelligence Papers, No. 20). OECD Publishing.

Office of Management and Budget. (2024). *Advancing governance, innovation, and risk management for agency use of artificial intelligence* (Memorandum M-24-10). Executive Office of the President.

Office of Management and Budget. (2025). *Accelerating federal use of AI through innovation, governance, and public trust* (Memorandum M-25-21). Executive Office of the President.

Oswald, M. (2018). Algorithm-assisted decision-making in the public sector: Framing the issues using administrative law rules governing discretionary power. *Philosophical Transactions of the Royal Society A, 376*(2128), 20170359. https://doi.org/10.1098/rsta.2017.0359

Parlementaire ondervragingscommissie Kinderopvangtoeslag. (2020). *Ongekend onrecht* [Unprecedented injustice]. Tweede Kamer der Staten-Generaal.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Peeters, R. (2020). The agency of algorithms: Understanding human-algorithm interaction in administrative decision-making. *Information Polity, 25*(4), 507–522.

Peeters, R., & Widlak, A. (2018). The digital cage: Administrative exclusion through information architecture – The case of the Dutch civil registry's master data management system. *Government Information Quarterly, 35*(2), 175–183. https://doi.org/10.1016/j.giq.2018.02.003

Prygodicz v Commonwealth of Australia (No 2) [2021] FCA 634.

Royal Commission into the Robodebt Scheme. (2023). *Report*. Commonwealth of Australia.

SCHUFA Holding (Scoring), Case C-634/21, ECLI:EU:C:2023:957 (Court of Justice of the European Union, 7 December 2023).

Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham Law Review, 87*(3), 1085–1139.

Veale, M., & Brass, I. (2019). Administration by algorithm? Public management meets public sector machine learning. In K. Yeung & M. Lodge (Eds.), *Algorithmic regulation* (pp. 121–149). Oxford University Press.

Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology, 31*(2), 841–887.

Weber, M. (1978). *Economy and society: An outline of interpretive sociology* (G. Roth & C. Wittich, Eds.). University of California Press.

Whiteford, P. (2021). Debt by design: The anatomy of a social policy fiasco – Or was it something worse? *Australian Journal of Public Administration, 80*(2), 340–360. https://doi.org/10.1111/1467-8500.12479

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—Applications and challenges. *International Journal of Public Administration, 42*(7), 596–615. https://doi.org/10.1080/01900692.2018.1498103

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance, 12*(4), 505–523. https://doi.org/10.1111/rego.12158

Young, M. M., Bullock, J. B., & Lecy, J. D. (2019). Artificial discretion as a tool of governance: A framework for understanding the impact of artificial intelligence on public administration. *Perspectives on Public Management and Governance, 2*(4), 301–313. https://doi.org/10.1093/ppmgov/gvz014

Zalnieriute, M., Bennett Moses, L., & Williams, G. (2019). The rule of law and automation of government decision-making. *The Modern Law Review, 82*(3), 425–455. https://doi.org/10.1111/1468-2230.12412

Zouridis, S., van Eck, M., & Bovens, M. (2020). Automated discretion. In T. Evans & P. Hupe (Eds.), *Discretion and the quest for controlled freedom* (pp. 313–329). Palgrave Macmillan.

Zuiderwijk, A., Chen, Y.-C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly, 38*(3), 101577. https://doi.org/10.1016/j.giq.2021.101577

// DECLARATIONS

▸ **Funding**	This research received no specific grant from any funding agency.
▸ **Conflict of Interest**	The authors declare that they have no known financial or personal conflicts of interest that could have influenced the work reported in this paper.
▸ **Data Availability Statement**	No new data were created or analysed in this study. The article is based on legislation, regulatory and policy instruments, reported judicial decisions and published scholarship cited in the references.
▸ **Ethical Statement**	Ethical approval was not required because the study is a doctrinal and comparative policy analysis of publicly available legal texts, judicial decisions, policy documents and published literature and did not involve human participants.
▸ **AI Use Disclosure**	Generative artificial intelligence tools were used in a limited and assistive capacity during the preparation of this manuscript, primarily for idea development, language refinement, and grammatical correction. The authors retain full responsibility for the originality, accuracy, analysis, interpretation, and final content of the manuscript.

// JOURNAL POLICIES *JuriSocial Journal of Criminology & Criminal Justice*

01 OPEN ACCESS & LICENCE
All articles are freely available on publication under CC BY-NC 4.0. Readers may share and adapt the work for non-commercial purposes with proper attribution.

02 COPYRIGHT
Authors retain copyright and grant JJCCJ the right of first publication.

03 PEER REVIEW
Every research article undergoes double-blind peer review by at least two independent reviewers.

04 PUBLICATION ETHICS
JJCCJ follows the Committee on Publication Ethics (COPE) Core Practices and guidelines.

05 PLAGIARISM POLICY
All submissions are screened with similarity-detection software; plagiarism in any form leads to rejection.

06 GENERATIVE AI POLICY
AI tools cannot be listed as authors. Any substantive use must be disclosed; authors remain fully responsible for the content.

07 CORRECTIONS & RETRACTIONS
Errors are corrected through published corrigenda; retractions follow COPE guidelines.

08 ARCHIVING & INDEXING
Content is preserved through [PKP Preservation Network / LOCKSS]. DOIs are registered with Crossref.

09 PUBLICATION CHARGES
No article publishing fee charged.

10 DISCLAIMER
The views expressed are those of the authors and do not necessarily reflect the views of the editors or the publisher.

SUBMISSIONS & ENQUIRIES [editor@journal-domain.org] ◆ [https://journal-website-url/]