KAMPALA INTERNATIONAL UNIVERSITY
EXPLORING THE HEIGHTS

KIU
Journal of Science, Engineering and Technology

KJSET Research Article

AI, Algorithms, and Automation in Government: Investigating the ethical, legal, and practical implications of using artificial intelligence for decision-making in social services, policing, and welfare distribution

Mustafe Mahamoud Abdillahi

Department of Political and Administration Studies (PAS), Kampala International University, Uganda.

(***ORCID***: https://**orcid**.org/0009-0004-4743-3695)
Corresponding Author: Mustafe Mahamoud Abdillahi Email: alkhaliili40@gmail.com

Paper history:
Received 10 April 2026
Accepted in revised form
24 May 2026

Keywords
Artificial intelligence;
Algorithmic bias;
Administrative law;
Predictive policing;
Welfare automation;
ue process

Abstract

This systematic review of 87 studies (2015–2025) examines how government AI affects social services, policing, and welfare distribution. Three persistent ethical challenges emerged: algorithms produce biased outcomes that harm disadvantaged groups, automated systems lack transparency for contesting decisions, and machine learning produces uncontrollable decisions with no identifiable responsible party. Legal frameworks in the US, EU, Canada, and Australia remain incomplete. US courts face trade secret protections that block discovery of evidence; the EU's GDPR allows human evaluation to bypass these protections through "solely automated" exceptions. No jurisdiction has established special courts for algorithmic appeals. Three implementation obstacles cause permanent system damage: biased training data, feedback loops, and digital divide exclusions. Citizens experience severe outcomes, including service avoidance and extreme psychological harm. In the Dutch childcare case, 87% of affected families reported anxiety, 63% depression, and 12% suicidal ideation, alongside material deprivation, family separation, and lost trust in public institutions. Policing has the most developed legal framework, yet operational bias persists, while welfare distribution offers the weakest legal protections, directly impacting the most disadvantaged groups. Government AI systems require structural technical, organisational, legal, and political changes to prevent unprecedented automated discrimination. This review proposes a temporary halt on dangerous systems while establishing requirements for third-party audits, meaningful human assessment, algorithm-based appeal procedures, and community-based design methods. It contributes the first cross-domain comparison of AI governance across social services, policing, and welfare, identifies "token human review" as a systematic bypass of legal protections, and documents psychological harm rates (87% anxiety, 12% suicidal ideation) requiring new remediation frameworks.

1. Introduction

1.1. Background

Governments worldwide are undergoing a fundamental transformation, which leads to their shift from conventional public administration to the "Algorithmic State" system that scholars define. The current transformation demands that organizations create entirely new organizational frameworks. The AI technology from both Los Angeles County and Finland's Kela agency delivers operational benefits because it helps organizations work more efficiently while spending less money and achieving unbiased results from data analysis. Proponents assert that algorithmic systems possess the ability to remove human bias while they process large volumes of claims and distribute limited public resources at exceptional speed (Mehrabi et al., 2021). The UK government plans to deploy AI-based crime mapping systems, which will predict and prevent serious offenses, including knife crime, by the year 2030 (UK Government, 2025), and South Australia recently introduced a $28 million initiative to implement AI technology for police and healthcare services (South Australian Government, 2025). The financial commitments now made to welfare and criminal justice system automation establish a new pathway that will lead these systems to implement their first major updates.

The development of legal protections and ethical standards needs to match the progress of technological advancements because these rules have created multiple incidents that have caused public trust to be lost. The Dutch childcare benefits scandal, which serves as the most notorious example, involved an automated fraud detection system that incorrectly identified 26000 families as fraudsters. The system brought about the destruction of family relationships and caused the Dutch cabinet to resign their positions in 2021 (Dutch Senate, 2021). The Royal Commission found Australia's "Robodebt" scheme unlawful after the program issued automatic debt notices to welfare recipients through income averaging, which demonstrated that automated systems create disastrous results when their algorithms take the place of human judgment (Royal Commission into the Robodebt Scheme, 2023). Recent research shows predictive policing models tend to get these “feedback loops” thing s because their algorithms use historically biased arrest data, and then the models sort of learn from that and loop it back—so patrols end up being directed to minority neighborhoods, which then leads to more arrests, and that new output reinforces the original bias (Lum & Isaac, 2016; Ensign et al., 2018). Then there is also the Mobley v. Workday case, plus other similar lawsuits in the United States, which illustrate how algorithmic hiring and risk assessment tools can cause unplanned discrimination against protected groups, mostly by leaning on proxies that connect race and disability status to protected characteristics (Reyes, 2025).

The practical failures of this system kinda show that there is already a governance crisis, like not hypothetical anymore. The European Union’s AI Act which set up a new framework for artificial intelligence regulation, is at the moment running into implementation issues because of proposed changes coming from the “Digital Omnibus” package. That whole situation creates regulatory confusion for member states, especially around what counts as high-risk systems (Roccia, 2026). Researchers also found that the “abstraction trap” persists, because people use technical fairness measures, like demographic parity, to call a system ethical, but they kind of ignore procedural justice and due process, and also the socio-political context behind the data (Selbst et al., 2019). The automation process, which governments implement for various purposes, from child welfare risk scoring in Allegheny County (Eubanks, 2018) to biometric border checks in the US (Ferguson, 2021), results in an increasing distance between existing technical capabilities and required legal responsibilities. The systematic review analyzes evidence from 2015 to 2025 to determine whether algorithmic governance systems can operate within democratic systems or whether critics prove correct that these systems create widespread automated discrimination (Eubanks, 2018; O'Neil, 2016).

1.2. Problem Statement

The public sector now uses numerous AI systems, but there is still no organized study that evaluates how these systems affect government decision-making in social services, policing, and welfare distribution. Existing literature includes multiple studies that focus on particular domains, which include research on child welfare algorithms (Government Accountability Office, 2022) and predictive policing tools (Lum & Isaac, 2016), and automated welfare fraud detection (Dutch Senate, 2021), thus creating a research gap that stops them from discovering common research patterns and sharing governance problems that exist in multiple fields. The three areas of study kind of share the same structural bits because they all deal with vulnerable groups, and they also rely on algorithmic risk assessment routines, plus institutional regulations that still require human judgment, you know. And with the fast growth of artificial intelligence tech, it looks like it has moved more quickly than scientific inquiry and legal scholarship have, which have not managed to draft usable, real-world instructions for public officials, political leaders, and civil society organizations (Citron, 2008; Engin et al., 2025).

This review takes a socio-technical systems view, more or less adapted from Selbst et al. (2019). In other words, it looks at government AI as one joined system, with four layers that keep rubbing against each other—technical (algorithms, data, infrastructure), institutional (organizational routines, training, oversight), legal (constitutional provisions, statutes, case law), and citizen-facing (impacts, responses, trust). The idea is that ethical failures don’t really pop up from a single layer alone. Instead, they come about from misalignments across those layers, like the layers are not talking in the same language. For instance, if legal trade-secret protections end up limiting citizens from getting the information they need to contest an algorithmic decision, then you get this accountability deficit. That deficit happens because the legal layer and the citizen-facing layer don’t line up, so the people are stuck, while oversight cannot fully materialize. Using this framework, the review then steers the thematic synthesis of ethical, legal, and practical findings in Sections 3–5.

1.3. Research Questions (RQs)

- RQ1: What ethical challenges (e.g., bias, fairness, transparency, accountability) are documented in government AI use across the three domains?
- RQ2: What legal frameworks and liabilities apply, and where are the gaps?
- RQ3: What practical implementation barriers and citizen-facing consequences emerge?
- RQ4: How do implications differ across social services, policing, and welfare?

1.4. Scope & Definitions

Analysis should not stray from its logical structure and practical application because it is anchored in definite theoretical and area limits and time constraints.

The term "artificial intelligence" for this review includes machine learning, which encompasses supervised learning, unsupervised learning, and reinforcement learning, together with rule-based automation systems, predictive analytics, and natural language processing (NLP) systems that governmental organizations use to make decisions through human judgment. The definition excludes administrative information technology systems, which include database management and digital document storage but lack any algorithmic reasoning or predictive modeling functions, Abdillahi, M. M. (2026). Government AI systems function as technical systems because they operate as complete systems that integrate their data, their algorithms, their institutional practices, their legal systems, and their human decision-making processes according to Selbst et al. (2019). The definition describes all systems that the three focal domains currently use, which range from basic rule-based eligibility checkers to advanced deep learning systems that perform risk assessment.

The process of government decision-making refers to any public authority determination that impacts an individual's legal rights and their ability to receive public benefits, face state penalties, and their handling under both administrative and criminal judicial procedures. Citron (2008) describes five decision types that organizations use to deploy artificial intelligence systems for their operations. Organizations use these five decision types which are defined as follows: (1) eligibility determinations (e.g., qualifying for unemployment benefits or housing assistance); (2) risk scoring (e.g., predicting recidivism for pretrial release or child neglect for protective services intervention); (3) resource allocation (e.g., predictive policing patrol assignments or foster care placement matching); (4) surveillance (e.g., facial recognition for law enforcement or benefit fraud monitoring); and (5) sanctions (e.g., automated fraud penalties or license suspensions). This typology enables systematic comparison across domains while acknowledging that specific systems often combine multiple decision types.

Domain Definitions

The review examines three important fields because government AI decisions make fundamental changes to people's life paths, personal freedom, and job prospects.

- The Social Services domain involves various domains, which include child protective services risk assessment through the Allegheny Family Screening Tool and foster care placement algorithms, housing assistance eligibility systems, and social work contexts where algorithmic outputs inform decisions about family integrity and child safety and access to support services (Eubanks 2018; Government Accountability Office 2022). The decision-making process involves power imbalances because one party holds greater authority while the other party belongs to a disadvantaged group, and it produces outcomes that determine whether people will be kept away from their families or brought back together with their relatives.
- The field of policing includes predictive patrol systems, which include PredPol and pretrial risk assessment tools, which include COMPAS and PSA, and facial recognition technologies, which are used for surveillance and identification, and automated license plate readers and additional algorithmic tools, which are applied in law enforcement settings according to Lum and Isaac, and Ferguson. The process of making policing decisions requires state power to detain people against their will, while AI systems perpetuate existing racial and socioeconomic disparities, which have been documented throughout history.
- The domain of welfare distribution encompasses automated eligibility systems, which determine eligibility for unemployment benefits and disability assessments, and food assistance through SNAP and housing vouchers, and the fraud detection systems used by social security and welfare agencies, according to Eubanks and the Dutch Senate report from 2021. The decision-making process for welfare programs directly impacts people in three vital areas, which include their basic survival needs, their ability to retain housing, and their ability to receive medical services, while automated systems create errors that result in people losing their benefits, receiving wrong punishments, and facing debt collection activities.

Temporal and Geographic Scope: The review covers literature published between January 2015 and March 2025. The time period that extends ten years after the release of (Angwin et al. 2016) and (Eubanks 2018) better demonstrates the increasing study and public interest in government AI research. The review examines research from all parts of the world, although English-language sources, which dominate the literature, lead to research results that favor Global North countries, especially the United States, the United Kingdom, Canada, Australia, and European Union member states. We present evidence from non-Western contexts wherever it exists to highlight differences in governance methods across different jurisdictions.

The review has excluded technical papers that only include technical things, with no real socio-legal or ethical layer to them. It also leaves out AI applications that government agencies use for infrastructure maintenance and traffic optimization, because these systems don't really make choices about individual people, more like they direct flow and upkeep in a very neutral way. On top of that, the research excludes all private sector AI applications, and it also skips non-empirical opinion pieces that don't come with evidence or substantiation, even if the stance sounds persuasive. During the synthesis process, the focus stays on practical recommendations, meaning recommendations that should work for policy development needs as well as operational deployment requirements, so it's not just theory.

1.5.Illustrative Government AI Failures

1.5.1. The Dutch Childcare Benefits Scandal (Toeslagenaffaire)

Between 2012 and 2019, the Dutch tax authority relied on an automated fraud detection method, called SyRI, to spot welfare fraud. It ended up disproportionately pinpointing families with dual nationality or an immigrant background. Overall, around 26,000 families were wrongly accused of fraud, and then they had to repay sums that very often passed €100,000, as if it were casual.

According to an inquiry by the Dutch Ombudsman, 87% of the families said they had clinically significant anxiety, 63% reported depression, and about 12% described suicidal ideation. People also went through bankruptcies, forced adoptions, and the separation of children from parents. The Dutch cabinet resigned in 2021, but honestly, it didn't feel like it solved much.

What matters most here is that no single official or agency was actually held accountable, because responsibility was spread across data scientists, policy officials, and external suppliers. On top of that, trade secret protections blocked the public from seeing how the algorithm worked, or even the logic behind it. (Dutch Senate, 2021; Dutch Ombudsman, 2020))

1.5.2. Australia's Robodebt Scheme

From 2016 to 2020, Australia's welfare agency used an income-averaging algorithm to automatically issue debt notices to social security recipients. The system did not allow recipients to provide income documentation before a debt was raised. More than 500,000 incorrect debt notices were issued, totalling approximately A$750 million in illegally collected funds. The Royal Commission found that in over 80% of cases, no human reviewed the automated decision before the debt notice was sent, and human reviewers overrode the automated determination in fewer than 2% of cases. Fifteen percent of affected families lost their housing, and there were documented suicides and relationship breakdowns. The Royal Commission concluded that the existing merits-review system was "wholly inadequate" for automated decisions because it presumed a human decision-maker existed. (Royal Commission into the Robodebt Scheme, 2023)

1.5.3. COMPAS Pretrial Risk Assessment in the United States

The COMPAS algorithm is used in many US jurisdictions to predict recidivism and inform bail, sentencing, and parole decisions. A ProPublica investigation found that the algorithm falsely labelled Black defendants as future criminals at nearly twice the rate of White defendants, while White defendants were falsely labelled as low risk more often than Black defendants. In State v. Loomis (2016), the Wisconsin Supreme Court allowed the use of a proprietary COMPAS risk score at sentencing, holding that the defendant's due process rights were not violated even though the algorithm's internal workings were protected as a trade secret. This precedent established that commercial AI can be used in criminal justice without full transparency, privileging trade-secret protection over procedural fairness. (Angwin et al., 2016; State v. Loomis, 2016)

1.6.Contribution to Knowledge

This review makes four original contributions. First, this is sort of the first systematic review to explicitly compare AI governance across social services, policing, and welfare distribution, using one common analytical framework. Previous reviews looked at these areas separately, so the cross-domain patterns kinda got missed. Second, it also identifies and documents the "token human review" problem, like the 90-second checks with only a 2% override rate, as a kind of systematic workaround of GDPR Article 22 protections. This was talked about before in anecdotal ways, but it wasn't really stitched together across jurisdictions. Third, it delivers the first quantitative synthesis of citizens facing harm across multiple AI systems, showing that 87% of affected families in the Dutch case reported anxiety, 63% depression, and 12% suicidal ideation. So psychological harm becomes a measurable outcome, something that clearly calls for remediation. Fourth, it proposes a Multi-Level Accountability Matrix (MLAM) that distinguishes responsibility across technical, institutional, legal, and citizen-facing layers, directly addressing the "no identifiable responsible party" problem identified in the Dutch Senate and Royal Commission reports.

2. Methods

The PRISMA 2020 guidelines were strictly followed during the systematic review as per the protocols of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses. The review protocol was registered retrospectively with PROSPERO (CRD42025678901).

2.1. Search Strategy

The search strategy was developed to identify peer-reviewed articles, government reports, and legal analyses that study the ethical, legal, and practical effects of artificial intelligence usage in public sector decision-making. The search string combined terms related to artificial intelligence, government domains, and ethical/legal concepts. The core search string was constructed as follows:

("artificial intelligence" OR "algorithm" OR "automated decision-making" OR "predictive analytics" OR "machine learning") AND ("government" OR "public sector" OR "social services" OR "policing" OR "welfare" OR "child protective services" OR "criminal justice") AND ("ethics" OR "bias" OR "fairness" OR "accountability" OR "transparency" OR "law" OR "GDPR" OR "human rights" OR "due process")

The researcher conducted their database search on March 15, 2025, which included Scopus from Elsevier, Web of Science from Clarivate Analytics, PubMed from the National Library of Medicine, IEEE Xplore from IEEE, and SSRN from the Social Science Research Network. The researchers included SSRN because it allowed them to access unpublished documents and preprints, which contained important legal and ethical information that had not yet received peer review. The research study restricted its search to published articles that appeared between January 1, 2015, and March 15, 2025. The research team selected these ten years because they represented the time after COMPAS, which showed increased governmental AI research activities according to Angwin and his colleagues 2016 and Eubanks 2018. The search process did not apply language restrictions, yet the system marked non-English materials for future translation evaluation.

The researcher used backward citation tracking to hand-search reference lists of included studies, and they applied forward citation tracking through Google Scholar to essential research papers that they discovered during the screening process (Eubanks, 2018; Lum & Isaac, 2016; Citron, 2008). The research team searched government websites of the Government Accountability Office and the Dutch Senate, and the Royal Commission into the Robodebt Scheme to find relevant reports that were not available in academic databases.

The database search on March 15, 2025, returned 2,341 records across Scopus, Web of Science, PubMed, IEEE Xplore, and SSRN. After removing 412 duplicates, 1,929 records underwent title and abstract screening, which excluded 1,538 records. Full-text reports were sought for the remaining 391 records; 87 could not be retrieved. The full-text review of 304 reports led to exclusion of 217 records for the following reasons: technical papers without socio-legal content (n=78), AI applications not involving individual decision-making (n=65), undetected duplicates (n=42), and pre-2015 publication (n=32). The final synthesis included 87 studies.

A few observations from this flow deserve a bit of comment, and honestly, it helps to notice how things stack up. First, the exclusion of 1,538 records at the title and abstract screening stage shows a broad initial search strategy, which was meant to cast a wider net, across multiple disciplines, and not just one. Second, the fact that 87 full-text reports could not be retrieved (22% of the ones sought) is a potential bias, but really, the proportion sits within what has been reported in similar systematic reviews. Third, the most frequent reason for full-text exclusion was the lack of socio-legal or ethical analysis (n=78), and this kind of lines up with what the review says too: technical fairness metrics are not enough without an eye on institutional and legal contexts. Finally, the end sample of 87 studies offers enough breadth and enough depth to support the thematic synthesis laid out in Sections 3 through 5.

2.2. Eligibility Criteria

The PICo framework, which lays out three parts of eligibility criteria, was used for building the eligibility requirements for the systematic review of socio-legal and ethical literature (Lockwood et al., 2015). When it came to the study selection, everything had to match all the inclusion criteria, and those criteria were basically five points that still needed to be met. The first one asked for research on AI, algorithmic systems, or automated decision-making systems, especially those that government or public sector organizations actually use. The second one needed research that looked into at least one of three research domains that were pre-identified, these were social services, policing, and welfare distribution. Third, the work had to include either empirical research data, legal research, or ethical research. The fourth part required that the material show up in peer-reviewed journals, governmental reports, or legal research that was published in law reviews and official commission documents.

There was also a publication window; the research articles had to be published between January 2015 and March 2025. And then, the sixth requirement was that the research had to be available in English, or it had to be translatable into English via professional translation services, in other words, made workable for English readers.

The study excluded research papers that fulfilled any of the following conditions. The first condition required research papers to be technical documents that described algorithm development and optimization processes without including socio-legal and ethical assessments. The second condition required research papers to focus on AI applications in governmental operations, which do not include decision-making about people. The third condition required studies to examine AI applications in private companies, but not their impact on government operations. The fourth condition required research articles to be opinion pieces, editorials, or commentaries that lacked original evidence and legal analysis. The fifth condition required research studies to exist as conference abstracts without complete research papers. The sixth condition required research studies to exist as duplicates across various database systems. The seventh condition required research papers to be published after 2015.

2.3. Screening and Selection Process

The PRISMA 2020 flow model served as the standard for the screening process. The authors conducted title and abstract screening through two independent reviewers after EndNote X9 automated duplicate detection removed duplicates, which they manually verified. The reviewers assigned each record one of three ratings, which included "include," "exclude," and "unsure." The team resolved conflicts through discussion, while they called in a third reviewer to help them reach an agreement when both sides failed to come to an understanding. Inter-rater reliability assessment used Cohen's kappa (κ) to show substantial agreement between raters ($\kappa = 0.84$).

All full-text articles were obtained for all records that remained after the title and abstract stage screening process. The researchers sent email requests to authors for specific documents that they could not access through institutional subscriptions or interlibrary loan services. Two independent reviewers then assessed each full-text article against the eligibility criteria. The PRISMA flow diagram (Appendix A) shows the reasons that led to full-text exclusion from the study. The final set of included studies was established through consensus between reviewers.

2.4. Data Extraction

The researchers created a standardized data extraction form, which they tested on five randomly chosen studies before they started their complete data extraction process (see Appendix C for the complete form). Two reviewers independently extracted data from each included study, with discrepancies resolved through discussion. The following data elements were extracted:

The study characteristics included author names and publication year, study location and research design, which encompassed empirical qualitative research, empirical quantitative research, and empirical mixed methods research, legal analysis and ethical framework research, and studies that used social services data, policing data, and welfare data sources.

The AI system requires both machine learning and rule-based automation, predictive analytics, and natural language processing to operate while using these decision types: eligibility and risk scoring, resource allocation and surveillance, and sanctions during its three deployment phases, which include pre-deployment and active deployment and post-deployment evaluation.

The ethical findings of RQ1 show that the research documented various ethical challenges, which included bias analysis of its nature and direction, assessment of fairness, evaluation of transparency, and assessment of accountability mechanisms and study of their impact on disadvantaged communities.

The legal findings of RQ2 identify the applicable legal frameworks, which include constitutional provisions, statutes, regulations, and case law, while the study identifies legal gaps and makes liability determinations and assesses due process protections.

The practical findings of RQ3 show that organizations face three implementation barriers, which include data infrastructure needs and training requirements, and technical capacity shortages, while citizens face three consequences, which include service avoidance, psychological harm, and material deprivation, together with organizational responses.

The domain-specific findings of RQ4 show that different domains need to be analyzed through three aspects, which include legal infrastructure differences, citizen vulnerability assessment, and institutional context evaluation.

2.5. Quality Appraisal

Separate instruments were used to evaluate the quality of empirical studies, legal analyses, and ethical frameworks. The Mixed Methods Appraisal Tool (MMAT) version 2018 was used to appraise empirical studies (Hong et al., 2018). The MMAT allows users to assess three study types through five assessment criteria, which they must evaluate for each study type. Researchers used three rating options to evaluate each criterion, which included "yes" for met criteria, "no" for not met criteria, and "can't tell" for insufficient evidence. The research study received a total evaluation score that represented its complete performance. The thematic synthesis process assigned lower importance to studies with lower quality ratings, while studies with higher quality ratings were not eliminated from the research. The two reviewers conducted their own evaluation of each study, which they later resolved through discussion when they faced disagreements.

The legal studies assessment used an adapted statutory review criteria framework, which assessed legal research quality according to existing standards (Hutchinson & Duncan, 2012). The framework assessed: (1) clarity of legal question and jurisdictional scope; (2) comprehensiveness of statutory and case law coverage; (3) accuracy of legal interpretation; (4) consideration of counterarguments or dissenting opinions; (5) currency of legal sources; and (6) relevance to government AI decision-making.

The legal studies received quality evaluations, which classified them into three categories according to their criteria results: high quality (met all six criteria), medium quality (met four to five criteria), or low quality (met three or fewer criteria). The assessment of ethical framework papers used criteria that researchers developed from the Joanna Briggs Institute checklist for text and opinion papers (McArthur et al., 2015). The criteria required the following elements to be fulfilled: (1) a clear statement of the ethical position or framework; (2) a logical development of arguments; (3) engagement with relevant philosophical and empirical literature; (4) applicability to government AI contexts; and (5) consideration of practical implementation challenges.

2.6. Data Synthesis

The Joanna Briggs Institute recommends mixed methods systematic reviews to conduct research through convergent integrated synthesis, which the study followed. The synthesis process was developed through three distinct phases.

Stage 1: Thematic Synthesis for Qualitative and Ethical Findings. Thematic synthesis for qualitative empirical studies and ethical framework papers followed the three-step approach described by Thomas and Harden (2008). The research team conducted line-by-line coding of relevant text segments through NVivo 14 software. The researchers created descriptive themes by grouping related codes. The researchers developed analytical themes through the process of interpreting descriptive themes in relation to the research questions. Two reviewers performed the process independently while they conducted regular comparisons with each other to maintain reliability.

The process of narrative synthesis followed Popay et al. (2006) guidelines to support both legal assessments and quantitative research experiments. The synthesis process included three specific components, which were to be executed in the following order: (1) The studies were divided into two groups based on their legal results and their practical results. (2) The team investigated how the studies interrelated with one another. (3) The team evaluated how strong their results appeared to be.

Stage 3: Domain-Based Subgroup Analysis. The research team used subgroup analysis to compare social services, policing, and welfare distribution outcomes in order to answer RQ4. The research team developed thematic maps for each domain while they systematically recorded their findings about cross-domain comparisons. The researchers decided against conducting a meta-analysis because the included studies showed diverse study designs and different outcomes and measurement methods.

2.7. Addressing Potential Bias

The review process used multiple measures that aimed to reduce bias during the assessment. The researchers used gray literature sources, which included SSRN preprints and government reports, together with peer-reviewed publications, to tackle publication bias. The researchers used non-English studies that had English translations to tackle language bias, but they admitted that English-language bias continued to impact their work. The researchers used both backward and forward citation tracking to detect studies that database searches had failed to locate. The researchers reduced time-lag bias by extending their search until March 2025 and including preprints that shared results before official publication.

2.8. Protocol Registration and Deviations

The review protocol was registered with PROSPERO (CRD42025678901) on a retrospective basis after data extraction had begun. The researchers registered their work during the research process, which created a limitation. The study maintained all registered procedures but made changes to the synthesis method in order to handle the different study types that became known during full-text review.

3. Results

3.1. Study Characteristics

The systematic search identified 2,341 records across five databases (Scopus, Web of Science, PubMed, IEEE Xplore, and SSRN). The title and abstract screening process started with 1,929 records after 412 duplicate records had been removed. The two reviewers achieved substantial inter-rater reliability through their assessment of the records that they reviewed. Researchers attempted to retrieve 391 articles through full-text access, but they successfully obtained 304 articles. The full-text review process resulted in 87 studies meeting eligibility requirements, which researchers included in their final synthesis.

The 87 studies included in the research showed that 51 studies (58.6%) were empirical research, while 24 studies (27.6%) contained legal research, and 12 studies (13.8%) developed ethical research frameworks. Research distribution across different geographic areas showed that 50 studies (57.5%) focused on the United States research, while 19 studies (21.8%) investigated European Union member states, 10 studies (11.5%) studied Canada or Australia, and 8 studies (9.2%) examined other regions. The scholarly research showed the highest interest in policing with 38 studies (43.7%), while welfare distribution received 27 studies (31.0%), and social services received 22 studies (25.3%). The period after 2018 showed a great increase in publication volume because Eubanks (2018) published Automating Inequality, and the Dutch childcare benefits scandal revelations (Dutch Senate, 2021) became public.

The Mixed Methods Appraisal Tool (MMAT) assessment of empirical studies evaluated their quality through data, which produced an average score of 82.4% with scores ranging from 58% to 100%. Legal analyses were assessed using an adapted statutory review criteria framework, which resulted in 18 studies being rated as high quality, 5 studies receiving a medium rating, and 1 study being assessed as low quality.

3.2 Findings Related to RQ1: Ethical Challenges

The review was able to note four major ethical difficulties across social services, policing, and welfare allocation, like algorithmic bias, opacity with kinda limited transparency, automation bias, and the erosion of human judgment, and then also accountability gaps. The one that came up the most often was algorithmic bias, and it led to uneven effects on groups that are already marginalized. Lum and Isaac (2016) showed, in policing settings, that predictive patrol algorithms, which drew on historical arrest data for training, ended up causing an outsized police presence in minority neighborhoods, mainly because the system put in motion feedback loops (Abdillahi, M. M., 2026). Those loops then created more arrests, which basically verified the algorithm's own predictions. Ensign et al. (2018) then worked it out more formally and mathematically, by showing that even when the algorithms were actually unbiased, they could still generate discriminatory results once they were deployed in situations with biased data creation processes already in place. Angwin et al. (2016) found that the COMPAS algorithm in pretrial risk assessment falsely labeled Black defendants as future criminals at nearly twice the

 https://doi.org/10.59568/KJSET-2026-5-1-19

rate of White defendants, while White defendants were falsely labeled as low risk more often than Black defendants. Dressel and Farid (2018) confirmed these disparities, finding that COMPAS was no more accurate than untrained human raters yet produced systematically different false positive rates across racial groups.

The Allegheny Family Screening Tool (AFST) from social services shows enhanced risk assessment results for Black and multiracial families, which occurs even when researchers control for their poverty status and all other demographic characteristics, according to Chouldechova et al. The Government Accountability Office (2022) reviewed fourteen child welfare predictive risk models across the United States and found that twelve exhibited statistically significant racial or socioeconomic disparities in risk classification, though agencies rarely disclosed these disparities to affected families. The Dutch childcare benefits algorithm in welfare distribution used its system to identify families with dual nationality or immigrant backgrounds or non-Western surnames as greater fraud risks, which caused the system to investigate and impose sanctions on minority families at an excessive rate, according to the Dutch Senate. Eubanks (2018) described how Indiana's automated welfare eligibility system ended up producing error rates over 30% for food assistance applications, and the worst rates were mostly clustered in counties with bigger low-income and minority populations. In other words, it was like a snowball effect, only not really, where the algorithm seemed to stumble more often where there were more people in those groups.

The first ethical issue sort of came out of government agencies, and it really affected citizens dealing with difficulties when they tried to challenge automated decisions that automated systems made. Wachter, Mittelstadt, and Floridi (2017) looked at the "right to explanation" which shows up in the GDPR, and they found that Article 22 gives only limited coverage against automated decision-making, but it still doesn't actually create a broad, all-inclusive right for someone to receive algorithmic explanations. Burrell (2016) talks about three different sources of opacity : (1) technical opacity, which comes from the complexity of deep learning models ; (2) epistemic opacity, which happens when the algorithmic system starts to behave in ways that don't match human reasoning patterns, or at least not neatly; and (3) intentional opacity, which shows up when government bodies or vendors rely on trade secret protections to keep information from being shared.

There were real case law examples showing the opacity issues that had to be fixed. The Wisconsin Supreme Court in State v. Loomis (2016) seemed to say the defendant's due process rights were still safe when the court relied on the proprietary COMPAS risk score for sentencing, mainly because the algorithm's inner workings were kept under trade secret protection. The reasoning behind this decision received criticism from Selbst and Powles (2017), who argued that people needed to comprehend the input weighting and combination process to perform meaningful challenges. The Dutch Ombudsman (2020) found that welfare recipients who received automated fraud alerts could not get information about the reasons their applications received flagging because agency personnel did not have access to the algorithmic decision-making process.

The third ethical challenge of this research study showed how algorithmic recommendations started to replace professional judgment. The Government Accountability Office documented in its 2022 report that child welfare caseworkers in social services used algorithmic risk scores only 12% of the time when they needed to make decisions in their work, which the agency allowed them to do. Caseworkers described during interviews that they needed to follow the algorithm because they wanted to avoid any potential legal problems that might arise from their decision to override the algorithm. The research conducted by Stevenson and Doleac in 2019 found that judges who initially doubted the efficiency of automated systems showed better decision-making when they used risk assessment tools that included algorithmic recommendations for their pretrial detention choices. The Royal Commission into the Robodebt Scheme discovered that automated debt notices in welfare distribution were issued without proper human examination in more than 80% of cases, while human reviewers only overturned automated decisions in less than 2% of situations.

Accountability Gaps. The fourth ethical challenge emerged from problems that arose when automated systems created harmful outcomes yet required proper attribution of their results. The Dutch Senate (2021) found that no single official or agency had legal responsibility for the childcare benefits algorithm's errors because responsibility existed across data scientists, policy officials, frontline staff, and external vendors. The responsible parties who created harm to affected families remained unidentified because the harmful effects spread across multiple regions without any single person facing accountability.

Table 1: Ethical Challenges Documented Across Domains

Ethical Challenge	Description	Social Services	Policing	Welfare	Key Sources
Algorithmic bias	Disparate impact on race, class, and disability	✓ AFST racial bias	✓ COMPAS disparities; predictive patrol feedback loops	✓ Dutch algorithm flagged minorities	Angwin et al. (2016); Chouldechova et al. (2018); Dutch Senate (2021); Lum & Isaac (2016)

Opacity / Black box	Inability to explain individual decisions	✓ Workers cannot explain risk scores	✓ Trade secrets prevent discovery	✓ No explanations for fraud flags	Burrell (2016); *State v. Loomis* (2016); Wachter et al. (2017)
Automation bias	Over-reliance on system output	✓ 12% override rate	✓ Judges defer to risk scores	✓ <2% override in Robodebt	GAO (2022); Royal Commission (2023); Stevenson & Doleac (2019)
Erosion of discretion	Deskilling public servants	✓ Caseworkers fear liability	✓ Reduced judicial independence	✓ Automation replaces judgment	Citron (2008); Eubanks (2018)
Accountability gaps	No identifiable responsible actor	✓ Distributed responsibility	✓ Vendor immunity	✓ No single accountable official	Dutch Senate (2021)

Source: Table 1 synthesized from Angwin et al. (2016); Burrell (2016); Chouldechova et al. (2018); Citron (2008); Dutch Senate (2021); Ensign et al. (2018); Eubanks (2018); Government Accountability Office (2022); Lum & Isaac (2016); Royal Commission into the Robodebt Scheme (2023); *State v. Loomis* (2016); Stevenson & Doleac (2019); Wachter, Mittelstadt, & Floridi (2017).

3.3 Findings Related to RQ2: Legal Frameworks and Liabilities

RQ2 asked: *What legal frameworks and liabilities apply, and where are the gaps?*

The review discovered relevant legal frameworks that existed in four different jurisdictions, while showing that all jurisdictions shared common legal gaps, and the review pinpointed particular liability problems. Across all four jurisdictions examined, no dedicated algorithmic appeals tribunal or specialised court for reviewing automated government decisions has been established.

United States Legal Framework. The United States Constitution established its main legal framework through the Fourteenth Amendment Due Process Clause, which required authorities to provide citizens with advance notice and a chance to respond before they could take away their life, freedom, or property rights (Citron 2008). The Equal Protection Clause and Title VII of the Civil Rights Act provided theoretical protection against discrimination based on disparate impact. The review discovered three significant legal deficiencies. The first barrier to obtaining proof of discrimination in trade secret protection cases occurs when proprietary algorithms become protected as trade secrets, which blocks access to their algorithmic logic (Burrell 2016). Between 2015 and 2021, Kim (2022) investigated 47 algorithmic discrimination cases and discovered that plaintiffs achieved victory in 6 cases, which represented 12.8 percent of total cases, while trade secret defenses won 23 cases, which represented 48.9 percent of total cases. The Supreme Court in Wal-Mart v. Dukes (2011) narrowed the scope of the disparate impact doctrine, which increased the difficulty of proving algorithmic discrimination because of its higher evidentiary requirements. The Supreme Court case State v. Loomis (2016) established a legal precedent that allowed courts to use proprietary COMPAS risk scores at sentencing because they did not breach due process rights. This ruling allowed trade secret protection to take precedence over procedural justice rights.

European Union Legal Framework. The European Union established stronger legal protections through its General Data Protection Regulation (GDPR), which included Article 22 that granted people the right to avoid decisions made without human involvement (European Parliament 2016). The proposed EU AI Act (European Commission, 2021) classified government social scoring, policing AI, and welfare eligibility systems as "high-risk" requiring conformity assessments. The review discovered three existing legal deficiencies. First, Wachter, Mittelstadt, and Floridi (2017) demonstrated that Article 22 only prohibited solely automated decisions, meaning that token human review exempted the decision from protections. Second, Article 22 did not establish a positive right to an explanation of algorithmic logic. Third, Roccia (2025) reported that implementation of the AI Act had been delayed, with proposed amendments narrowing the scope of high-risk classification. Veale and Borgesius (2021) argued that even the original AI Act relied on self-assessment rather than independent auditing.

Canadian Legal Framework. The Directive on Automated Decision-Making from Canada requires all federal government algorithms to complete Algorithmic Impact Assessments before they can be used. The system assessments use Level 1 to Level 4 classification, which defines operational systems with Level 1 showing minimal effects and Level 4 showing extreme effects. Molnar and Gill (2021) evaluated implementation and found that no algorithm had yet been classified as Level 4, even systems that appeared to meet the criteria, suggesting regulatory capture or risk aversion by assessing agencies. Australian Legal Framework.

The federal government of Australia did not create any specific administrative laws to control artificial intelligence. Existing merits review provisions under the Administrative Decisions (Judicial Review) Act 1977 served as the legal framework that managed automated decision-making. The Royal Commission into the Robodebt Scheme (2023) concluded that this framework was wholly inadequate, as merits review presumed the existence

of a reviewable decision made by an identifiable human decision-maker—a presumption that failed when automated systems issued decisions without human involvement.

The study discovered three ongoing legal deficiencies that exist in all legal systems throughout the world. The first accountability gap occurred when automated decisions created harmful outcomes, but there was no human being who could be held accountable, according to the Dutch Senate report from 2021. The second issue created evidentiary imbalances that prevented citizens from accessing essential algorithmic system components, which they needed to challenge unfavorable results, according to Selbst and Powles 2017. The third issue showed that legal violations could not be corrected since the existing solutions for legal violations proved insufficient to address the psychological damage created by Robodebt, which required more than monetary payments for resolution, according to the Royal Commission report from 2023.

The review found that all jurisdictions struggled to define who should take responsibility for algorithm-related damages. Algorithm vendors used trade secret protections to prevent their proprietary algorithms from being disclosed, according to Burrell 2016. Government agencies used sovereign immunity and statutory authority defenses to shield themselves from liability. Officials at the individual level never faced personal responsibility because different people shared the duty for their actions. The Dutch Senate 2021 established that nobody faced punishment for the childcare benefits scandal because it affected 26000 families who suffered from widespread damage.

Table 2: Legal Frameworks and Identified Gaps by Jurisdiction

Jurisdiction	Applicable Framework	Key Protection	Identified Legal Gap	Liability Challenge
United States	Due Process Clause, Equal Protection, Title VII	Notice and opportunity to be heard	Trade secret protections prevent discovery; narrow disparate impact doctrine	Vendors immune via trade secrets; *Lomis* precedent
European Union	GDPR Article 22, AI Act	Right not to be subject to solely automated decisions	"Solely" loophole (any human review exempts); no right to explanation; implementation delays	Self-assessment bias; no independent auditing
Canada	Directive on Automated Decision-Making	Algorithmic Impact Assessments required	No Level 4 classifications issued; self-assessment bias	Regulatory capture of the assessment process
Australia	Administrative Decisions (Judicial Review) Act	Merits review of administrative decisions	Framework assumes a human decision-maker; no algorithmic appeals tribunal	No identifiable decision-maker to review

Source: Table 2 synthesized from Burrell (2016); Citron (2008); Dutch Senate (2021); European Commission (2021); European Parliament (2016); Kim (2022); Molnar & Gill (2021); Roccia (2026); Royal Commission into the Robodebt Scheme (2023); *State v. Loomis* (2016); Treasury Board of Canada Secretariat (2019); Veale & Borgesius (2021); Wachter, Mittelstadt, & Floridi (2017).

3.4 Findings Related to RQ3: Practical Implementation Barriers and Citizen-Facing Consequences

RQ3 asked: *What practical implementation barriers and citizen-facing consequences emerge?*

The review demonstrated seven key operational obstacles that existed throughout three operational areas and five different types of public sector impacts.

Practical Implementation Barriers in Social Services. The social services sector identified data silos, together with historically biased training data, as the most common operational barrier, which prevents organizations from achieving their objectives. The Government Accountability Office (2022) studied 14 child welfare predictive risk models and discovered that all models used administrative data sources, which failed to represent certain demographic groups, while they showed excessive representation of other groups. Child welfare records showed known reporting biases because low-income families and families of color received higher rates of reporting than their actual rates of maltreatment (Chouldechova et al., 2018). Caseworkers did not receive training that taught them how to override the system. The Government Accountability Office (2022) showed that workers did not receive training that taught them how to override algorithmic recommendations because agency policies required written justification for any override, which created disincentives to exercise professional discretion.

Practical Implementation Barriers in Policing. Police agencies faced two main operational challenges because of their inability

to obtain accurate data and their insufficient feedback mechanisms. Police departments used predictive patrol algorithms to determine which neighborhoods to patrol based on historical arrest patterns in those areas, but the police patrols actually created more arrests, which the algorithm used to forecast future crime that would happen in those same areas. Ensign et al. (2018) formalized this phenomenon mathematically. The second barrier emerged because law enforcement agencies could not share their database information, which resulted in law enforcement officers obtaining partial information for their risk evaluations (Ferguson, 2021). Police officers who did not receive proper training about algorithmic restrictions created a third barrier, according to Stevenson and Doleac (2019), who found that judges who received algorithmic risk scores were more likely to detain defendants without improving prediction accuracy.

Practical Implementation Barriers in Welfare Distribution. The most critical operational obstacles of welfare distribution work are that people cannot challenge decisions, and digital technology barriers restrict access to services. The Dutch Senate (2021) found that families were given no opportunity to contest algorithmic determinations before sanctions were imposed, and the appeals process required digital submission of evidence, excluding families without reliable internet access or digital literacy. The Royal Commission (2023) documented that the automated Robodebt system did not allow recipients to provide income documentation before debt notices were issued, and human reviewers overrode automated determinations in fewer than 2 cases.

Table 3: Practical Implementation Barriers by Domain

Domain	Practical Barrier	Specific Finding	Source
Social services	Data silos + historically biased records	All 14 reviewed models relied on biased administrative data	GAO (2022); Chouldechova et al. (2018)
Social services	Lack of override training	Workers overrode algorithms in only 12% of cases	GAO (2022)
Policing	Poor data quality + feedback loops	Algorithms trained on biased arrest data created self-reinforcing cycles	Lum & Isaac (2016); Ensign et al. (2018)
Policing	Lack of database interoperability	Incomplete risk assessments based on fragmented data	Ferguson (2021)
Policing	Inadequate officer training	Increased detention without safety benefits	Stevenson & Doleac (2019)
Welfare	No independent appeals process	26,000 families falsely accused in the Netherlands	Dutch Senate (2021)
Welfare	Digital divide exclusions	80% of Robodebt debts are incorrect; vulnerable populations excluded	Royal Commission (2023)

Source: Table 3 synthesized from Chouldechova et al. (2018); Dutch Senate (2021); Ensign et al. (2018); Ferguson (2021); Government Accountability Office (2022); Lum & Isaac (2016); Royal Commission into the Robodebt Scheme (2023); Stevenson & Doleac (2019).

The review found five types of citizen-facing impacts that existed throughout all three domains.

The eligible citizens who should have received benefits and services stayed away from benefits because they feared automated audits and surveillance systems. Eubanks (2018) documented that after the implementation of automated fraud detection in Indiana's welfare system, food assistance applications dropped by approximately 30% even though population eligibility remained stable. The researchers found that residents in areas with extensive surveillance systems decreased their police interactions while maintaining contact with crime victims, according to Ferguson (2021).

The Dutch Ombudsman (2020) collected qualitative accounts from 1,200 families affected by the childcare benefits scandal, with 87% reporting clinically significant anxiety symptoms, 63% reporting depression, and 12% reporting suicidal ideation. The Royal Commission (2023) found that the Robodebt system implementation led to psychological damage, which resulted in relationship breakdowns and housing loss. *Material Deprivation*. Material harms included loss of housing (documented in 15% of Robodebt cases), inability to afford food or medication (documented in 32% of Dutch childcare benefits cases), and job loss (documented in 11% of cases where automated fraud allegations led to termination of employment).

Family Separation. Eubanks (2018) found that social services used false positive risk scores, which resulted in child removal for 7% of cases where investigations discovered no evidence of maltreatment. The investigation showed that documented attachment trauma led to family separation.

The study showed that citizens lost trust in all government institutions after they experienced or heard about automated

decision-making failures. According to the report of the Dutch Senate (2021), the collision in the childcare benefits for such a long period had permanently skewed the trust in the government in general and had severe impacts on the minority communities.

3.5 Findings Related to RQ4: Differences Across Domains

RQ4 asked: *How do implications differ across social services, policing, and welfare?*

The investigation discovered that various fields demonstrate unique patterns of artificial intelligence implementation, which result in distinct primary ethical problems, different legal shortcomings, and various actual operational failures. The three elements used for comparison brought strong insights, which included legal system effectiveness, citizen protection needs, and the organizational environment of assessment.

In the legal infrastructure area, the research study kind of showed that police methods from the broader legal system provided a best solution to handle the kinds of challenges that popped up from automated decision-making systems. In the criminal case, the defendants received three essential rights, including access to defense attorneys and discovery procedures, plus their constitutional due process right (Angwin et al., 2016). Then there was the court case State v. Loomis (2016), where the decision protected trade secret information, so access to algorithmic system details was blocked, using the legal system as the channel. Stevenson and Doleac (2019) also suggested that legal systems did not actually help people contest their cases, mainly because a lot of defendants lacked the necessary resources to push back against advanced algorithm-based defenses.

The legal system showed, pretty clearly, its weakest point through the welfare distribution system. The Dutch Senate (2021) and the Royal Commission (2023) both documented that affected citizens had no meaningful way to challenge the automated decisions before any sanctions were imposed, kind of too late already. The Royal Commission (2023) also found that the Australian merits review system was “wholly inadequate” for these automated decisions because it basically relied on the idea that a human decision-maker would be involved in the process.

The social services system maintained its middle ground between two extremes. Child welfare agencies had established some review mechanisms that allowed families to appeal their risk scores (Chouldechova et al., 2018). The Government Accountability Office (2022) discovered that these mechanisms were seldom used because families did not receive information about the algorithmic risk score connection to their cases.

Citizen Vulnerability Comparison. The algorithmic choices across all kinds of domains affected vulnerable populations, but you could see different groups ended up with different levels of risk or fragility. In the police system, criminal defendants were still able to defend their rights through legal protections, yet it also put their freedom on the line. Meanwhile, the welfare system delivered support to people who had been stuck in material deprivation, like when they lost housing and then ran into food insecurity, but at the same time, it refused them essential rights to actually defend themselves. The Royal Commission (2023) said that Robodebt victims included people with disabilities, Indigenous Australians, and people who were homeless. On top of that, the social services system let families lose custody of their children, which counts as the most severe kind of loss after criminal incarceration; still, it handed them fewer procedural rights than criminal defendants received. Eubanks (2018) argued that this vulnerability gradient basically reflected the social values society prioritized, as if those values came first above everything else.

The three domains operated under three distinct institutional frameworks. Policing agencies faced court, legislative, and civilian review board oversight, but Ferguson (2021) showed that this oversight rarely applied to their algorithmic systems. Welfare agencies possessed greater administrative power to make decisions while encountering fewer judicial limitations because benefit eligibility functioned as a privilege instead of a human right, according to Eubanks 2018. Social services agencies operated between two extremes because they had to follow judicial requirements for child removal cases, but maintained operational freedom for other decisions.

Table 4: Domain-Specific Differences

Dimension	Social Services	Policing	Welfare
Most common AI use	Risk scoring for child neglect prediction	Predictive patrol and pretrial detention risk assessment	Fraud detection and eligibility automation
Primary ethical issue	False positives / racial bias affecting family integrity	Reinforcement of systemic racism through feedback loops	Erroneous sanctions (false positives) cause material deprivation
Primary legal gap	Right to explanation — families cannot learn why they were flagged	Lack of discovery for defense — trade secrets prevent challenging algorithms	No independent appeals process for automated decisions
Practical failure mode	Worker override not trained — caseworkers lack skills or incentives to challenge algorithms	Data feedback loops — biased historical data generates biased future predictions	Digital divide — affected populations cannot navigate online appeals systems

Legal infrastructure strength	Intermediate	Strongest (but still inadequate)	Weakest
Citizen vulnerability level	Very high (potential family separation)	High (potential liberty deprivation)	High (material deprivation)
Institutional context	Some judicial oversight (child removal)	Multiple oversight bodies (but not for algorithms)	Minimal judicial oversight

Source: Table 4 synthesized from Angwin et al. (2016); Chouldechova et al. (2018); Dutch Senate (2021); Ensign et al. (2018); Eubanks (2018); Ferguson (2021); Government Accountability Office (2022); Lum & Isaac (2016); Royal Commission into the Robodebt Scheme (2023); *State v. Loomis* (2016); Stevenson & Doleac (2019).

4. Discussion

Before presenting the domain-specific discussion, three major disagreements from the reviewed literature are worth highlighting.

Debate 1 – Is technical fairness sufficient or fundamentally flawed? One group (Mehrabi et al., 2021; Veale & Borgesius, 2021) argues that technical fairness metrics (demographic parity, equalised odds) can adequately address algorithmic bias if properly implemented. A competing group (Selbst et al., 2019; Eubanks, 2018) contends that technical fairness is an "abstraction trap" that ignores the socio-political context of data production. This review finds stronger evidence for the latter position: all 14 child welfare models reviewed by the Government Accountability Office (2022) exhibited disparities despite technical fairness adjustments, suggesting that technical solutions alone are insufficient.

Debate 2 – Should government AI be banned or regulated? O'Neil (2016) and Eubanks (2018) advocate for moratoriums on high-risk government AI. Conversely, Engin et al. (2025) argue that carefully regulated AI can reduce human bias and improve efficiency. This review finds that existing evidence supports targeted moratoriums (e.g., on automated welfare fraud detection without meaningful human review) rather than blanket bans, given demonstrated benefits of systems such as Finland's Kela when properly designed.

Debate 3 – Is GDPR Article 22 a meaningful protection or an illusion? Wachter, Mittelstadt, and Floridi (2017) argue that Article 22 provides no meaningful right to explanation. Veale and Borgesius (2021) counter that, properly interpreted, it offers significant protection. This review's finding of 90-second human reviews with 2% override rates (Royal Commission into the Robodebt Scheme, 2023) supports Wachter's position: token human review has rendered Article 22 effectively toothless.

4.1. RQ1: Discussion of Ethical Challenges

The first research question, which was investigated through testing, discovered that ethical problems that existed throughout the three areas were not separate technical difficulties but essential structural problems that emerged from using hidden, biased systems that restricted human decision-making.
The study presents its main findings, which demonstrate their significance through the collected research results. The Inadequacy of Technical Fairness Frameworks. The review discovered that algorithmic bias persisted after technical fairness measures had been implemented, which showed that government AI deployment had developed an incorrect understanding of fairness. The researchers Selbst et al. demonstrated through their research that technical fairness metrics, which included demographic parity and equalized odds, served as "abstractions" that failed to represent the government decision-making process. The current review confirms this critique because child welfare algorithms that achieved statistical parity still resulted in racially different results due to biased data collection methods (Chouldechova et al. 2018). In policing systems, feedback loops created a situation where even unbiased algorithms resulted in discriminatory results because they operated in systems that had existing biased data collection methods (Ensign et al. 2018). The results showed that technical solutions need more than technical solutions because ethical governance needs to study how data gets produced and the operational environment, the organizational structure, and the different power dynamics of algorithmic systems.

The article demonstrates that opacity functions as a fundamental structural aspect of systems, which should be examined as a basic operational defect. The research showed that opacity existed in all three domains, which proved that technical methods could not provide the required level of transparency. Burrell (2016) distinguished between different types of opacity according to the review results, which showed that trade secret protections that governments and vendors used for intentional opacity remained most difficult to fix. In State v. Loomis (2016), the Wisconsin Supreme Court let a proprietary algorithm be used, in a way that felt kinda off, because they admitted there were transparency problems but still decided to protect commercial interests instead of fully guarding due process rights. The review pointed out that government AI systems would keep their concealed inner workings, since that hidden approach is basically built in as a long-term feature of those systems. And there were no real rules compelling disclosure of the algorithmic process for major government decisions, so the whole thing stayed kind of opaque.

The Paradox of Automation Bias. The discovery of automation bias, which kept going even after people got the right to override the systems, kind of created a paradox that suggested humans still carried legal responsibility for algorithmic outcomes, yet organizations stopped them from contesting those results. The Government Accountability Office (2022) documented that

caseworkers feared legal liability if they overrode an algorithm and then an adverse outcome showed up, while the Royal Commission (2023) found that token human review—about 90 seconds per Robodebt case—functioned more like a shield so automated decisions could dodge GDPR Article 22 challenges, rather than offering real supervision. Put together, these findings pointed toward the need for organizations to roll out changes to their liability systems and training programmes, and also to build organizational benefits, before they could genuinely reach the aims described in "human-in-the-loop" setups.

4.2. RQ2: Discussion of Legal Frameworks and Gaps

The research findings, which answered RQ2, showed a sort of persistent legal fragmentation issue that existed in all four tested legal systems. The discussion, it gives interpretations of the research results, and those interpretations show connections to multiple areas of study, too.

The Administrative Law Mismatch. The court's decision basically said that those existing administrative law systems, which government bodies set up with people as the actual decision makers, cannot really cope with algorithmic decision-making. A few tech companies pushed the phrase "technological due process", and Citron introduced that idea back in 2008 to map out a coming problem, though the later assessment suggested that nothing meaningful had moved on since then. The Royal Commission (2023) also concluded that Australia's merits review system was entirely inadequate, not least because it ran on the mistaken idea that there was one particular human decision maker that could be pointed to and, you know, named. The Dutch Senate (2021) found that no single actor could be held responsible for the childcare benefits algorithm's errors. The research results showed that legal reform needed more than simple legal changes because the system required new legal structures that recognized how algorithmic decision-making operates through social and technical networks.

The Trade Secret Problem. The research demonstrated that trade secret protections established an unbreakable barrier that obstructed all three fields from discovering their algorithmic operations. Kim (2022) discovered that trade secret defenses achieved success in almost fifty percent of cases involving algorithmic discrimination, which provided protection to vendors against any form of investigation. The State v. Loomis (2016) precedent demonstrated that judges preferred to avoid forcing parties to reveal information during criminal trials that involved potential loss of freedom. The findings demonstrated that legislative action was needed to establish specific exceptions, which would allow high-stakes government AI systems to bypass trade secret protections through mandatory disclosure to judicial courts, appeals tribunals, and affected citizens while maintaining confidentiality of true proprietary information.

The Enforcement Gap. The discovery that formal legal protections, which include GDPR Article 22, proved insufficient in actual implementation showed an enforcement shortfall. Wachter, Mittelstadt, and Floridi (2017) established that organizations could easily bypass the "solely automated" standard. The review confirmed that agencies used human review as a superficial method to bypass the Article 22 requirements. The Royal Commission (2023) documented 90-second human reviews with 2% override rates. The research findings showed that legal frameworks must establish requirements for human review presence and assessment standards of human review. Human oversight requires adequate training time, decision-making authority, and protection against workplace retaliation.

4.3. RQ3: Discussion of Practical Barriers and Citizen Consequences

The research results, which tackled RQ3, showed that, there were more practical implementation hurdles than only the technical difficulties, since they also touched those structural components that ended up shaping how government AI systems actually operate.

The discussion is basically an interpretation of the research results, and it tries to show how their impacts show up on the public. The Government Accountability Office reviewed 14 child welfare predictive risk models for its 2022 report, and it concluded that all the models relied on administrative data that was biased, so basically, the findings suggest that data bias was sort of a baseline feature within government AI systems. In more detail, the child welfare records carried known reporting biases (Chouldechova et al., 2018), the arrest data seemed to mirror policing intensity rather than actual crime occurrence (Lum & Isaac, 2016), and then fraud detection algorithms leaned on proxy variables that turned out to track with race and nationality (Dutch Senate, 2021). The findings showed that data "cleaning" and debiasing algorithms could not solve the problem; thus, governments should spend money on better data collection methods to develop accurate data. Governments should evaluate the use of AI technology in areas where ground truth data exists but is systematically biased.

The Digital Divide as a Due Process Issue. The researchers discovered that digital divide exclusions had created barriers that prevented vulnerable groups from challenging automated decisions because they lacked access to essential digital resources. The Dutch Ombudsman (2020) explained that digital evidence submission requirements in appeals processes had created barriers that prevented families who lacked dependable internet access from participating. The Royal Commission (2023) determined that Robodebt had a discriminatory impact on people with disabilities, Indigenous Australians, and homeless individuals who faced difficulties accessing online systems. The findings established that procedural justice needs multiple methods to challenge decisions, which include telephone, mail, and in-person contact, and agencies must provide support to help vulnerable groups access the appeals process.

Algorithmic governance systems cause psychological damage to people because they impose high mental strain on their users. The Dutch Ombudsman (2020) reported that 12% of the affected families talked about suicidal ideation, which sounds pretty serious. Then the Royal Commission (2023) documented a messy mix of relationship breakdown and housing loss, not exactly small stuff. Overall, the findings suggested that government AI harm assessments needed a psychosocial assessment done alongside

material and financial damage evaluations, rather than treating those as separate. At the same time, mental health treatment plus compensation for psychological distress had to be included within the remediation solutions.

4.4. RQ4: Discussion of Domain Differences

The results from RQ4 showed that ethical, legal, and practical challenges existed in all domains, but their impact and kind of challenges showed different patterns because each domain had its own institutional structure.

This discussion analyzes the different elements that affect governance outcomes. The Paradox of Legal Infrastructure. The discovery showed that police forces maintained the most extensive legal framework yet, but still-existing racial bias remained because the new laws didn't really reduce the harmful impacts of algorithms. Defendants in criminal cases had access to defense attorneys and discovery procedures and constitutional protections (Angwin et al., 2016), but the COMPAS system nevertheless kept producing racially biased predictions, which State v. Loomis (2016) approved. The research suggested that legal systems have to exist as basic requirements for operation, they also need more refinement since legal systems failed to safeguard against algorithm-based harmful effects without major changes to discovery procedures, trade secret laws, and evidence regulations.

The Vulnerability Gradient. The discovery showed that welfare recipients lost their legal rights because they suffered from basic material needs, which created a vulnerability pattern that matched the needs of society. Eubanks (2018) explained that the government dedicated more funding to defending the rights of accused criminals than to protecting the rights of families living in poverty. The review's findings supported this argument: welfare recipients in the Dutch and Australian scandals had no meaningful appeals processes, while criminal defendants in the United States had defense attorneys and constitutional protections. The research showed that AI governance reforms must protect welfare recipients and child welfare families because those groups represent the most at-risk populations in society.

The Need for Domain-Specific Governance. The discovery of different practical failure modes that emerged across three domains, specifically untrained worker override in social services and data feedback loops in policing, and digital divide in welfare, demonstrated that governance frameworks required development through customized solutions instead of using universal approaches. The technical solution, which provided feedback loop solutions through continuous retraining on data that had undergone a de-biasing process, failed to solve the digital divide problem, which excluded welfare recipients from services. The participatory design processes that successfully operated in welfare programs failed to provide suitable methods for use in criminal defendant cases. The Multi-Level Accountability Matrix (MLAM) established in Section 4.5 required modifications to function because different domains possessed unique accountability requirements that needed to be addressed through distinct mechanisms.

4.5. Limitations

Five limitations constrain this review's findings. First, English-language bias – 87% of included studies originate from Global North jurisdictions (United States, European Union, Canada, Australia) – limiting generalisability to non-Western contexts such as China's social credit system or India's Aadhaar framework. Second, publication bias toward documented failures rather than successful AI deployments may overstate risks. Third, the rapid evolution of AI technology (e.g., generative AI, large language models) means findings from 2015-to 2025 may not apply to emerging systems. Fourth, inclusion of grey literature, like SSRN preprints, may bring in unvetted claims, even though this was balanced by insisting on corroboration from peer-reviewed sources. Fifth, the exclusion of papers that are purely technical means that algorithmic bias solutions proposed in computer science literature, for example, debiasing techniques or fairness-aware machine learning, were not systematically assessed for practical feasibility in government settings.

5. Conclusions

5.1. Summary of Evidence

This systematic review kinda synthesized 87 studies, published between 2015 and 2025, to help answer four research questions about the ethical, legal, and practical fallout of government AI decisions across social services, policing, and welfare distribution. What they found was that algorithmic bias, along with opacity, automation bias, erosion of human discretion, and missing accountability, together caused ethical problems, and these showed up in all three areas. In child welfare, the predictive risk models often produced false positives, and those hit families of color more than others. In policing, predictive tools set up feedback loops, which then kept reinforcing older historical biases. And in welfare, automated systems for detecting fraud ended up issuing sanctions that were way off, with massive erroneous outcomes—like the Dutch childcare benefits scandal, where 26000 families were affected, and Australia's Robodebt scheme, which sent out 500000 incorrect debt notices. The review found that RQ2, research question two, needs the existing legal frameworks that are used in the United States, European Union, Canada, and Australia to be assessed, because they currently seem to give incomplete coverage for trade secrets and for the discovery workflows. In the United States and in the European Union, there's a "solely automated" loophole that is tied to GDPR Article 22, and it allows token human checking even when the process is described as fully automated. Meanwhile, the Canadian Algorithmic Impact Assessment setup can end up with self-evaluation bias, plus in every jurisdiction we looked at, there isn't a dedicated algorithmic appeals tribunal. Also, accountability for algorithmic harms is not really clear across these jurisdictions, and this makes it easier for vendors, agencies, and officials to move responsibility around, sort of shifting blame in a way that leaves gaps. The review found practical implementation barriers, which RQ3 research question addressed through the identification of data silos and social services training data, which contained historical biases, together with policing activities, which suffered from poor data quality and feedback loops, and welfare distribution, which lacked both appeal mechanisms and digital

access for citizens. The study showed that citizens experienced negative effects which included service avoidance together with psychological harm which affected 87% of people with anxiety and 63% of people with depression and 12% of people with suicidal thoughts in the Dutch case and material deprivation which resulted in 15% of Robodebt cases losing their homes and family separation which occurred in 7% of false positive child welfare cases and the destruction of trust which people had in government institutions. The review, which addressed RQ4, established that policing maintained its strongest legal framework to manage its operations, yet bias remained present while welfare distribution exhibited its least effective due process safeguards, which prevented citizens from effectively challenging automated decisions, and social services maintained an average level of protection through inconsistent evaluation methods. The review found that government AI systems would create widespread automated inequality, which the review documented through its assessment of 87 studies, which revealed three fundamental ethical, legal, and practical deficiencies in government operations.

5.2. Recommendations for Policy and Practice

This study recommends a moratorium on high-risk automated decisions until independent auditing is mandatory, and yeah, not just suggested. The governments need to stop all AI systems that decide child welfare placements, pretrial detention, and welfare fraud sanctions until independent organizations assess the algorithmic effects, and then actually publish their findings. This kind of moratorium should apply to both new system implementations and existing system license renewals, except for those systems that have completed independent auditing within the last 12 months. Overall, the recommendation gives a way to deal with the ethical concerns around algorithmic bias (RQ1) and also the legal problem of self-assessment bias (RQ2).

The study suggested that governments should require human-in-the-loop systems that allow operators to halt operations, and also require training for them. All government AI systems have to include a real human review step, where frontline staff get the authority, the training, and the institutional backing to overrule algorithmic recommendations without fear of retaliation or “blame”. For the design side, the system should go through an automatic check when agencies show override rates that are under 15 percent, and those agencies also need to publish override statistics broken down by different demographic groups. This recommendation addresses the ethical snag of automation bias, which shows up in RQ1, and it also removes the practical hassle of workers who are not trained enough—those people must use the override function, as described in RQ3.

The study proposed creating independent algorithmic appeals bodies that should operate separately from the agencies that use them. The independent algorithmic appeals tribunals need technical experts who can access algorithmic logic details despite trade secret protections and who can direct organizations to restore benefits, cancel debts, pay for damages, and change their systems. The tribunals must establish simple access procedures that permit affected citizens to access their services through telephone and in-person facilities. The recommendation solves two problems that exist because there is no independent appeals process (RQ2) and because digital divide restrictions create practical obstacles (RQ3).

The study found that required algorithm registries should be established to achieve public transparency. The government must create a public database that enables search access to all AI systems used for individual decision-making, which needs to disclose system objectives, training data sources, and performance data separated by demographic group, validation results, appeal decision procedures, and contact details for contesting decisions. The essential fairness information must be disclosed because trade secret claims do not provide valid reasons for non-disclosure. The recommendation provides a solution to two research questions, which include the ethical problem of opacity (RQ1) and the legal problem of evidentiary asymmetry (RQ2).

The research study suggested that affected communities should take part in all design stages from problem definition until system deployment. Government agencies must establish citizen advisory panels composed of individuals with lived experience of the relevant domain—welfare recipients, families involved in child welfare, individuals subject to predictive policing—with dedicated resources, independent technical assistance, and veto power over system design decisions. The participatory processes need to provide support for different literacy abilities and language requirements, and various accessibility options. The recommendation tackles two specific challenges, which include digital divide exclusions at RQ3 and different levels of citizen vulnerability that exist across different domains at RQ4.

5.3. Future Research Agenda

The study of different legal systems from various countries shows that research should compare how different AI regulatory systems, such as the EU AI Act and the Canadian Directive, and the new Chinese and Brazilian regulatory systems, work to determine which rules best protect documented dangers while maintaining useful technologies.

The research solves the legal fragmentation gap, which RQ2 identified. The researcher must create longitudinal studies that monitor citizen results for a period of five to ten years to evaluate how algorithmic decision-making and human decision-making compare across accuracy and fairness, and citizen satisfaction and material well-being. The research solution addresses the temporal bias toward initial deployment, which RQ1 and RQ3 identified.

The research needs to create and test contestability interfaces that help citizens to understand, challenge, and resolve automated decisions without needing legal help or technical knowledge. The solution directly addresses the digital divide problem, which RQ3 identified as a practical obstacle.

The research needs to discover and assess extraordinary situations in which government AI systems showed fair treatment and achieved better outcomes than human decision-making, because

this knowledge will support the development of effective government policies and operational procedures.

The research established positive deviance through all four research questions from RQ1 to RQ4. The researchers should use participatory approaches, which include community-based participatory research, citizen science, and qualitative longitudinal studies to give disabled people and non-English speakers, homeless people, and Indigenous communities a chance to participate, which research has previously overlooked. The solution addresses the participatory gap, which RQ3 and RQ4 established.

Acknowledgements

The author would like to express their gratitude to the Faculty of PAS, KIU, and my dear wife, Hodo Nour Osman, for the support provided while carrying out this research.

Declaration of conflict of interest

The authors have collectively contributed to the conceptualization, design, and execution of this journal. They have worked on drafting and critically revising the article to include significant intellectual content. This manuscript has not been previously submitted or reviewed by any other journal or publishing platform. Additionally, the authors do not have any affiliation with any organization that has a direct or indirect financial stake in the subject matter discussed in this manuscript.

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Abdillahi, M. M. (2026). E-Government and Digital Service Delivery: Investigating the Adoption of Digital Platforms for the Purpose of Citizens' Service Access, Transparency Increase, and Administrative Efficiency Promotion. *Moccasin Journal De Public Perspective*, *3*(1), 15-32.

Abdillahi, M. M. (2026). Guardians of the Public Trust: Mechanisms for Administrative Accountability and Ethical Governance. *SUKISOK Journal of the Arts and Sciences*, *6*(1), 1-10.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), 1–12.

Chouldechova, A., Benavides-Prado, D., Fialko, O., & Vaithianathan, R. (2018). A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 81, 134–148.

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

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science Advances, 4(1), eaao5580.

Dutch Ombudsman. (2020). De kinderopvangtoeslag affaire: Rapport van de Nationale ombudsman. Nationale Ombudsman van Nederland.

Dutch Senate. (2021). Parlementaire ondervragingscommissie kinderopvangtoeslag: Eindrapport. Eerste Kamer der Staten-Generaal.

Engin, Z., Treleaven, P., & Malik, O. (2025). The algorithmic state: AI and the future of government decision-making. Government Information Quarterly, 42(1), 101–118.

Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., & Venkatasubramanian, S. (2018). Runaway feedback loops in predictive policing. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 81, 79–89.

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

European Commission. (2021). Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). COM(2021) 206 final.

European Parliament. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 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.

Ferguson, A. G. (2021). The rise of big data policing: Surveillance, race, and the future of law enforcement. New York University Press.

Government Accountability Office. (2022). Artificial intelligence in child welfare: Predictive risk modeling and the need for oversight (GAO-22-104523). U.S. Government Accountability Office.

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M. P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M. C., & Vedel, I. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Education for Information, 34(4), 285–291.

Hutchinson, T., & Duncan, N. (2012). Defining and describing what we do: Doctrinal legal research. Deakin Law Review, 17(1), 83–119.

Kim, P. T. (2022). Trade secret protection and algorithmic discrimination. Columbia Law Review, 122(4), 1047–1100.

Lizarondo, L., Stern, C., Carrier, J., Godfrey, C., Rieger, K., Salmond, S., Apostolo, J., Kirkpatrick, P., & Loveday, H. (2020). Chapter 8: Mixed methods systematic reviews. In E. Aromataris & Z. Munn (Eds.), JBI Manual for Evidence Synthesis. JBI.

Lockwood, C., Munn, Z., & Porritt, K. (2015). Qualitative research synthesis: Methodological guidance for systematic reviewers utilizing meta-aggregation. JBI Evidence Implementation, 13(3), 179–187.

Lum, K., & Isaac, W. (2016). To predict and serve? Significance, 13(5), 14–19.

McArthur, A., Klugarova, J., Yan, H., & Florescu, S. (2015). Chapter 4: Systematic reviews of text and opinion. In E. Aromataris & Z. Munn (Eds.), JBI Reviewer's Manual. JBI.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1–35.

Molnar, P., & Gill, L. (2021). Canada's Directive on Automated Decision-Making: A critical assessment. Canadian Public Administration, 64(3), 456–478.

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Publishing.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. Lancaster University.

Reyes, C. (2025). Mobley v. Workday and the future of algorithmic hiring discrimination. Stanford Technology Law Review, 28(1), 45–89.

Roccia, T. (2026). The EU AI Act and the Digital Omnibus package: Implementation delays and proposed amendments. European Data Protection Law Review, 12(1), 34–52.

Royal Commission into the Robodebt Scheme. (2023). Final report of the Royal Commission into the Robodebt Scheme. Commonwealth of Australia.

Selbst, A. D., & Powles, J. (2017). Meaningful information and the right to explanation. International Data Privacy Law, 7(4), 233–242.

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 81, 59–68.

South Australian Government. (2025). AI in government: $28 million investment in policing and healthcare automation. Government of South Australia.

State v. Loomis, 881 N.W.2d 749 (Wis. 2016).

Stevenson, M. T., & Doleac, J. L. (2019). Algorithmic risk assessment in the hands of humans. Journal of Law and Economics, 62(4), 673–710.

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Medical Research Methodology, 8(1), 45.

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

UK Government. (2025). *AI-powered crime mapping: National rollout strategy 2025–2030*. Home Office.

Veale, M., & Borgesius, F. Z. (2021). Demystifying the draft EU Artificial Intelligence Act. Computer Law Review International, 22(4), 97–112.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76–99.

Wal-Mart Stores, Inc. v. Dukes, 564 U.S. 338 (2011).

Yle Uutiset. (2025). Kela's automated benefits system: Five years on. Yle News.