

Algorithmic Decision-Making and Social Inequality: Assessing Legal Accountability, Transparency, and Fairness in Automated Public Services

Dr. Alok Kumar Bhargava^{1*}, Aprajita Singh², Dr. Sudarshan Nimma³,
Manish Sharma⁴, Supriya Chaudhary⁵

^{1*}Founder and Principal Researcher, TrayiVani Foundation, India; Developer, The Inner Engine Framework™ for Conscious Leadership Assessment; Author of TrayiVānī – Eternal Verses on Peace, Silence & Discernment and The Inner Engine of Leadership Trilogy; TrayiVani Foundation, Ghaziabad – 201016, Uttar Pradesh, India, Email Id: founder@trayivani.in, Website: <https://trayivani.in>, Orcid ID: 0009-0009-1805-3075, Google Scholar Id: XRzNh50AAAAJ, Scopus ID: 60698294500.

²Research Scholar, Faculty of law, Banaras Hindu University, Uttar Pradesh, 221005, Email ID: aprajitasingh1205@gmail.com

³Associate Professor, University College of Law, Kakatiya University, Telangana, Email ID: snimma26@gmail.com

⁴Academic Associate, Department of Humanities & Social sciences, IIM Indore, M.P., Indore, 453556, Email ID: manishs@iimdr.ac.in

⁵Assistant Professor of Law, University Institute of Legal Studies, Chandigarh University, 140413, ORCID ID: 0009-0004-4094-2622, Email ID: supriya.e16244@cumail.in

HOW TO CITE: ABSTRACT:

Alok Kumar Bhargava, Aprajita Singh, Sudarshan Nimma, Manish Sharma, Supriya Chaudhary, (2026). Algorithmic Decision-Making and Social Inequality: Assessing Legal Accountability, Transparency, and Fairness in Automated Public Services
International Journal of Special Education, 41(24s), 1121 - 1134

Algorithmic decision-making is further integrated into public services and questions are being raised whether algorithms enhance administrative efficiency at the expense of legal accountability, transparency, fairness, and equality in society. The research adopted a comparative empirical design using only secondary data from official and authoritative documents of the public sector, and covered 12 automated public-service systems recorded in the Algorithmic Transparency Recording Standard documents of UK public sector bodies published between 2022 and 2025. The legal accountability, transparency, fairness, and contestability were rated on standardized 0–100 scales developed by the researcher. The median scores for legal accountability, transparency, fairness and contestability were 93.75, 100, 50.0 and 75.0 respectively. 9 of 12 systems (75.0%) have completed a formal impact assessment and 5 of 12 (41.7%) have completed an equality-specific assessment and meaningful human reconsideration was clearly documented in 10 of 12 systems (83.3%). In ten cases (83.3%), vulnerability to inequality was found, but in no case was there evidence at the group level that met the strict criteria of the study for an inequality outcome.

COPYRIGHT STATEMENT: None of the legal accountability, fairness or contestability results

Copyright: © 2026 Authors.
Open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

were statistically significant in the cross-domain tests and transparency was tested and found uniform at 100. The results indicate that disclosure was not enough to create a socially responsible automation. Effective governance called for enhanced equality assessment, substantive review, easy challenges and ongoing monitoring of group level impact. The research offered an empirically informed framework for evaluating if responsible-AI principles were applied in practice as a tool to ensure safety in automated public services.

Keywords: algorithmic decision-making; automated public services; social inequality; legal accountability; transparency; algorithmic fairness; human oversight; contestability

1. Introduction

Algorithmic systems are no longer confined to experimental government use, but are now part of decisions that could affect access to benefits, education, the police, health services, financial aid, and other government services. It's not surprising they're popular. Automated tools are able to analyze vast amounts of information, detect patterns that would be hard to see when looking through the material manually, and streamline repetitive administrative tasks. According to OECD, AI may enable public servants to release almost one-third of their time for value-added tasks. These successes have led to the governments' use of AI in the design of policies, service delivery, citizens and interactions, prediction and support for administrative decision making (Berryhill et al., 2019). However, public sector recommendations have different consequences than those in the private sector and can impact on the provision of essential services, investigations or state opportunities.

The implications are more dangerous if automated systems are functioning in an environment of already existing social and economic inequality. Algorithmic tools in high-stakes contexts, such as criminal justice, education, benefits administration, have an impact on the rights and participation of citizens, and are used by public agencies. Who is responsible can also be hard to determine when a decision is made in response to a combination of administrative rules, historical data, software systems, public officials and private technology suppliers. The issues that go beyond prediction, however, to accountability, privacy, participation, and social inequity are raised by algorithmic decision-making (Levy et al., 2021).

A welfare administration report shows the extent of this worry. The increasingly digital nature of social-protection systems, covering technologies for assessing eligibility, calculating benefits, fighting fraud, risk scoring, and identity verification, was described in a United Nations investigation. A computerised social-assistance system in Ontario, Canada has been linked to

1,132 cases of identified errors in eligibility and payment decisions worth around CAD 140 million. The same report reported that more than 1.2 billion people in India were already covered by a biometric identification system, Aadhaar, which is used directly in the provision of social rights and public benefits and shows how digital infrastructures can be at scale, operating at the level of the population. These examples revealed how administrative efficiency and the potential for new constraints were in competition with technological systems for populations already reliant on public institutions.

The legal and governance protection matters thus made a major focus of the discussion. The European Union Agency for Fundamental Rights (ELAF) analyzed the use of AI based on 101 interviews conducted in five countries and took into account applications in the fields of social welfare, law enforcement, health services and marketing. The investigation found that it had not received much in-depth focus on non-discrimination; had insufficient clarity about human review; and had encountered challenges with implementing effective remedies. It therefore highlighted impact assessment, accountability, transparency, non-discrimination and access to justice as interlinked requirements as opposed to as standalone safeguards (European Union Agency for Fundamental Rights, 2020).

Although these issues have received more and more attention, the empirical evidence was still sparse. Initially, 85 studies were identified and 26 of them were selected for further analysis in the systematic review as they explicitly addressed the implications of AI for public governance. The majority of this literature was exploratory, conceptual, qualitative or practice driven, resulting in calls for empirical, explanatory research on government AI, risk management, governance structure, and impact measurement (Zuiderwijk et al., 2021). There was a specific lack of comparative evidence of how well documented public-sector systems included substantive accountability, fairness protections, and

opportunities for appeal of decisions in addition to transparency.

To fill this gap, the present study examined the automated public-service systems with the help of systematically gathered secondary data from official records with an empirical comparative approach. It explored four questions: the strength of the evidence for legal accountability, transparency, fairness, and contestability across public-service domains and system characteristics; the frequency of inequality outcomes and inequality vulnerabilities documented; and whether secondary evidence was available to assess associations between inequality accountability with documented inequality outcomes and inequality vulnerabilities. This study was a first step in taking official documentary evidence and using it to create structured data at the case level to understand how responsible-AI principles were applied to real public-service governance.

2. Conceptual and Legal Background

2.1 Algorithmic Decision-Making in Public Services

Automated public-service systems included rule-based classification, tools that assisted in, ranked, or in part made administrative decisions based on machine learning. Their importance was not so much technical as in terms of their ability to influence access to state resources and

interventions. Other public-sector examples include migration, welfare and digital-service administration, revealing that governance decisions made about the deployment were as important as the model design (Kuziemski & Misuraca, 2020). Explanations were also required that were appropriate to institutional responsibilities, not to institutional technical transparency, as was the case with front-line professionals (Veale et al., 2018). More than 3.6 million referrals were made annually in the United States to child protection agencies, and algorithm-assisted screening could impact high-stakes decisions (Chouldechova et al., 2018). Allhutter, et al. (2020) segmented job seekers in Austria into three groups by three employment-service thresholds: (i) high short-term prospects (at least 66%), (ii) low long-term prospects (25% or less), and (iii) intermediate conditions. Risk was not a thing of the past once humans were involved as

officials could rely on algorithmic recommendations (Green, 2022). Likewise, the gaps between trustworthy-AI ideals and practice were captured by Swedish labour-market decision support (Berman et al., 2024).

2.2 Algorithmic Decision-Making and Social Inequality

Algorithmic inequality arose when existing social patterns were fed into a dataset, algorithmic modeling, rules in institutions, or when it was implemented. The difference in performance may be significant: For Gender Shades, error rates for darker-skinned women were 34.7% compared to 0.8% for lighter-skinned men across commercial classifiers tested (Buolamwini & Gebru, 2018). These differences demonstrated that the notion of fairness was more than just average predictive accuracy. Sociotechnical systems (Selbst et al., 2019) were embedded in organisational settings, and harms could occur during data collection, model development, deployment, or during post deployment use (Suresh & Gutttag, 2021). It is important to consider that statistical parity may be incompatible with social justice in child protection, especially when groups come into systems with different historical circumstances (Kedell, 2019). Fairness also demonstrated the competing philosophical notions of equality and desert (Binns, 2018). Legal protection continued to be complicated, as algorithmic discrimination could lead to proxies and differences that do not align with protected categories (Wachter et al., 2021). The decisions on fairness had to involve participation and democratic legitimacy, not just technical optimisation (Wong, 2020).

2.3 Legal Accountability, Transparency, Fairness, and Contestability

The empirical framework considered legal accountability, transparency, fairness and contestability as interrelated but separate safeguards. Legal responsibility related to the responsibility to explain, justify and be held accountable for algorithmic decision-making (Wieringa, 2020). Auditing, went beyond just checking model performance, but extended the responsibility across the system lifecycle (Raji et al., 2020). Transparency was seen as the need to provide information to people and oversight bodies about the role of AI in public decisions (de Fine Licht & de Fine Licht, 2020), whereas

comparisons with human decision-making raised the issue of how to avoid the expectation of total explainability (Zerilli et al., 2019). Transparency should be a system-level conception, not just transparency of source code or transparency of the model internals (Larsson & Heintz, 2020). Other requirements for public accountability were for “arrangements between the decisions and responsible authorities” (Loi & Spielkamp, 2021). The assessment of such processes had to be done in an integrated way, considering the three elements of the concept of Fair, Transparent and Accountable processes (Lepri et al., 2018). The framework was completed by the questions of whether decisions could be contested, challenged, corrected, and tracked throughout the decision-making process (Cobbe et al., 2021).

3. Methodology

3.1 Research Design and Secondary Data Sources

The study was of an empirical nature, using a comparative design, and based only on secondary data about algorithmic decision making in public services. The primary case evidence consisted of records from the UK's Algorithmic Transparency Recording Standard (ATRS), which had a standardised disclosure structure, interpreted in accordance with the official public sector guidance (Government Digital Service, 2025). The sources of supplementary government, regulatory, audit and legal materials were used to check institutional or legal information as needed, and to provide conceptual context but not in place of official case documents. Peer-reviewed empirical studies were used for providing an empirical context to the issues raised, but not as a substitute for the official case documentation. Information was collected and checked against sources prior to being placed in the analytical dataset. The comparative element in the selected public-sector sample for the UK was with cross-system and cross-domain comparisons, but not with a cross-national comparison.

3.2 Case Identification and Eligibility Criteria

Cases were identified through targeted searches of official repositories, public-sector algorithm registers, regulatory records, audit reports, and relevant institutional publications issued between 2018 and 2025. A case was added where a public authority made use of an algorithmic/automated system and this had a material effect on access to

a service, benefit, sanction or administrative decision. Systems that are used for routine clerical automation were not included. Lacking adequate documentation, substantiated institutional ownership, or identifiable evidence of public-service use were discarded cases. The criteria applied resulted in the final analytical sample of 12 automated public service systems documented in recordings of the UK Algorithmic Transparency Recording Standard from 2022 to 2025.

3.3 Data Extraction and Variable Construction

A researcher developed data set was developed to provide a consistent way of comparing across cases. The information obtained included jurisdiction, domain of public service, purpose of the system, level of automation, number of people impacted, involvement of a vendor, human oversight, appeals, explanation mechanisms, independent audit, and assessment of fairness or bias. Features of governance were not included in documented inequality concerns. If the sources recorded differences between numbers, those errors were recorded, as were the number of people affected and the results of appeals. Records have been reviewed with their sources before coding.

3.4 Legal Accountability, Transparency, Fairness, and Contestability Measures

Four governance dimensions were measured separately: legal accountability, transparency, fairness safeguards, and contestability or human oversight. There was legal accountability for a clearly identified public body responsible for the system, a clear legal or administrative basis for the use of the system, formal oversight or review mechanisms and mechanisms for review, appeal, correction or redress were documented. Transparency included disclosure of the purpose of the system, the role of the decision, data or technical information, and public documentation. Fairness included equality/bias assessments, consideration of protected or vulnerable groups, disparity monitoring and documented mitigation measures. Contestability included notification of algorithmic involvement, meaningful reconsideration by human, authority to challenge, correct or override the algorithmic output or as its basis, and the practical capacity of the person affected and/or of the responsible human

decision-maker to challenge, correct or override the algorithmic output or the underlying evidence in the context of an operational or formal review process. The indicators were rated as 0 if the indicator is not documented, 1 if it is partially or not well documented, and 2 if it is well documented. Within each dimension, scores were totalled and normalised on a 0–100 scale to reflect the extent to which safeguards were well documented. There were four indicators for each

dimension and a maximum raw score of 8 not standardized. The fairness indicators were compared and contrasted with the UK review of bias in algorithmic decision making (Centre for Data Ethics and Innovation, 2020). The indicators of impact and accountability were also compared to the Canadian Algorithmic Impact Assessment framework (Treasury Board of Canada Secretariat, 2020).

Table 1. Operationalization and Coding of Governance Dimensions

Dimension	Four indicators	Coding rule
Legal accountability	Responsible public authority; legal/administrative basis; formal oversight/review; formal review/appeal/correction/redress	0 = absent; 1 = partial/unclear; 2 = clearly documented
Transparency	System purpose; role in decision; data/technical information; public documentation	0 = absent; 1 = partial/unclear; 2 = clearly documented
Fairness	Equality/bias assessment; protected/vulnerable groups considered; disparity monitoring; mitigation measures	0 = absent; 1 = partial/unclear; 2 = clearly documented
Contestability	Notification; meaningful human reconsideration; override/correction authority; practical challenge/correction route	0 = absent; 1 = partial/unclear; 2 = clearly documented

Note. Each dimension had a maximum raw score of 8 and was standardized as (observed raw score / 8) × 100.

3.5 Measurement of Social Inequality

Two categories have been used to measure social inequality. Inequality outcomes were only coded if an authoritative source documented unequal treatment, unequal error rates, disproportionate exclusion, discriminatory effects, or other measurable disadvantages to identifiable social groups. Inequality vulnerability was coded in the presence of one or more of the following equality safeguards: equality-impact assessment, a system that involves assessing the performance of groups or classes, a system that involves monitoring disparities, or a protective mechanism of a similar nature, but in the absence of, or with incomplete, these equality safeguards. A safeguard that was

not documented did not make a system vulnerable, unless there was no substantive individual decision consequence or the operational target of the system was a regulated institution or profession and not a population of equality-sensitive public services. This distinction allowed for limited disclosure not to be automatically deemed as proof of discrimination. When the group-specific quantitative disparities were reported as rates, counts, percentages, or comparisons of outcomes in the official sources, these disparities were recorded.

3.6 Statistical Analysis

Descriptive statistics were used to summarize system characteristics and the distribution of

governance safeguards. Mean scores of governance of the total sample and of the four public-service domains were standardized. The small sample size and ordinal-derived scores prevented the use of traditional domain differences analysis based on one-way ANOVA; thus, cross-domain differences were evaluated by tie-corrected Kruskal-Wallis H tests for dimensions with non-zero variance, and set the p value for determining statistical significance at .05. No tests were performed on dimensions that had no variance. Only where there was some variation in outcomes was it planned to investigate association between governance safeguards and documented inequality outcomes. When the outcome variable was constant, regression analysis was not estimated.

4. Results

4.1 Characteristics of Automated Public-Service

Systems

There were 12 systems identified that matched the inclusion criteria and were coded from the official UK ATRS records from 2022 to 2025. Five systems (41.7%) had to do with social protection or with health-related administration, four (33.3%) had to do with education or with funding for students, and three (25.0%) had to do with public-order or with regulatory decisions. There is 11 system(s) (91.7%) in production and 1 system (8.3%) that is still in pre-deployment. Table 2 lists the authorities that are included, deployment status, reported operating scale, and the official record that was used for coding. The reported scale varied from around 65,000 Advanced Learning Loan applications per year to energy-cost modelling of 27 million properties with support of around 3 million Warm Home Discount rebates per year.

Table 2. Automated Public-Service Systems Included in the Analysis

Authority and system	Domain	Phase	Reported scale	Official source
Department for Work and Pensions — ESA Online Medical Matching	Social protection/health	Production	Not quantified in ATRS	(Department for Work and Pensions, 2025b)
Department for Work and Pensions — CMS Predictive Analytics Compliance	Social protection/health	Production	~100,000 compliant cases	(Department for Work and Pensions, 2025a)
NHS Business Services Authority — NHS Prescriptions and Dental Exemption Check	Social protection/health	Production	Not quantified in ATRS	(NHS Business Services Authority, 2025)
Bristol City Council — Bristol NEET Model	Social protection/health	Production	Applied to ages 4–16	(Bristol City Council, 2025)
Department for Energy Security and Net Zero — Warm Home Discount Energy Cost Modelling	Social protection/health	Production	27m properties; ~3m rebates/year	(Department for Energy Security and Net Zero, 2025)
Student Loans Company — Advanced Learning Loans Assessment Rules	Education/student funding	Production	~65,000 applications/year	(Student Loans Company, 2025a)
Student Loans Company — Postgraduate Assessment Rules	Education/student funding	Production	~1m applications since 2016	(Student Loans Company, 2025b)

Authority and system	Domain	Phase	Reported scale	Official source
Standards and Testing Agency — Reception Baseline Assessment Routing	Education/student funding	Production	~600,000 pupils/year	(Standards and Testing Agency, 2025)
Office for Students — Teaching Funding Allocations	Education/student funding	Production	~350 providers; 1.25m student FTEs	(Office for Students, 2025)
The Probation Service — Effective Proposal Framework	Public order/regulatory	Production	All applicable PSR/pre-release cases	(The Probation Service, 2025)
Driver and Vehicle Standards Agency — MOT Risk Rating	Public order/regulatory	Production	~64,000 testers; 23,000 garages	(Driver and Vehicle Standards Agency, 2025)
Hampshire and Thames Valley Police — DARAT	Public order/regulatory	Pre-deployment	Tens of thousands of incidents/year per force	(Hampshire and Thames Valley Police, 2022)

Note. Scale values reproduce the quantities stated in the official ATRS records; “not quantified” indicates that no comparable operating count was reported in the record.

4.2 Legal Accountability, Transparency, Fairness, and Contestability Safeguards

The coded safeguards exhibited very strong transparency and lower levels of fairness and contestability coverage. Table 3 shows that the median scores for legal accountability, transparency, fairness and contestability were 93.75, 100, 50.0 and 75.0, respectively (Figure 1). Only one of twelve systems (75.0%) reported a completed formal impact assessment, one (8.3%) an assessment is ongoing or planned and two (16.7%) reported there is no assessment. Five

cases (41.7%) completed equality-specific assessments and one extra case (8.3%) planned to complete equality-specific assessments. Ten systems (83.3%) had clearly documented meaningful human reconsideration, one system (8.3%) did not have a complete review of human reconsideration but had planned to, and one system (8.3%) did not have human reconsideration at all.

Table 3. Standardized Governance Scores by Automated Public-Service System

System	Legal accountability	Transparency	Fairness	Contestability
ESA Online Medical Matching	100	100	75	100
CMS Predictive Analytics Compliance	100	100	62.5	75
NHS Prescriptions and Dental Exemption Check	87.5	100	37.5	87.5

System	Legal accountability	Transparency	Fairness	Contestability
Bristol NEET Model	75	100	37.5	75
Warm Home Discount Energy Cost Modelling	100	100	75	87.5
Advanced Learning Loans Assessment Rules	100	100	37.5	75
Postgraduate Assessment Rules	100	100	37.5	75
Reception Baseline Assessment Routing	87.5	100	62.5	25
Teaching Funding Allocations	87.5	100	37.5	75
Effective Proposal Framework	100	100	87.5	75
MOT Risk Rating	62.5	100	12.5	75
DARAT	62.5	100	62.5	25

Note. Scores range from 0 to 100; higher values indicate more completely documented safeguards under the coding framework in Table 1.

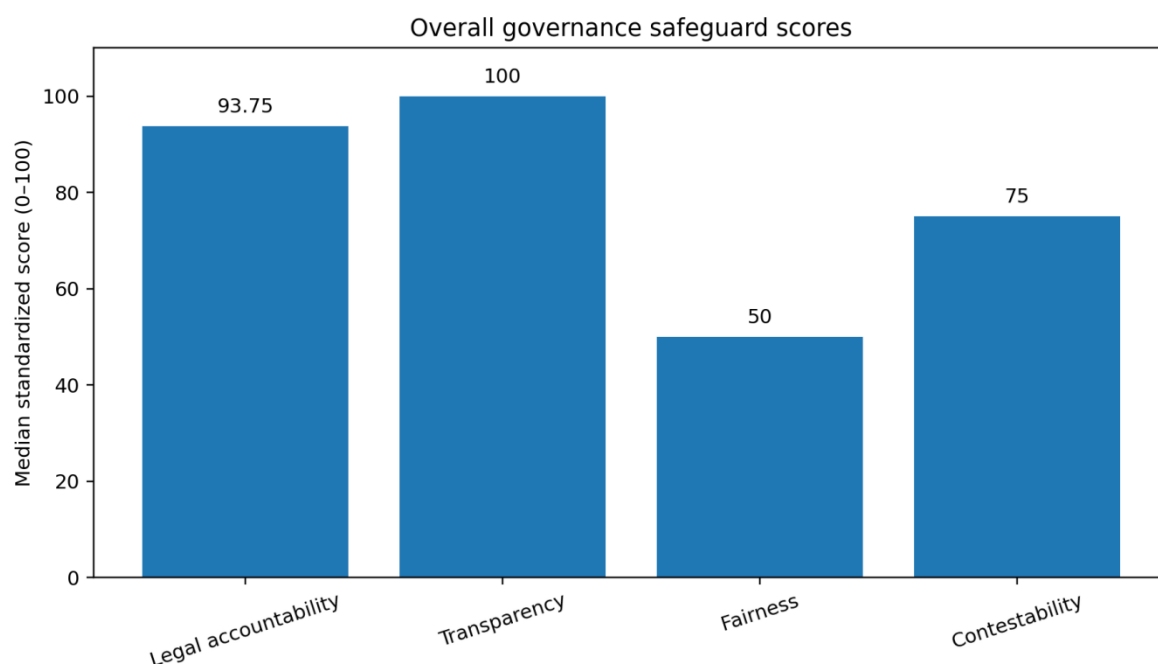


Figure 1. Median standardized governance safeguard scores across the 12 systems.

4.3 Variation Across Public-Service Domains and System Characteristics

Governance scores were more likely to differ with safeguard type than with service domain. For social-protection/health systems, the median

scores were 100 for legal accountability, 100 for transparency, 62.5 for fairness, and 87.5 for contestability. The levels shown were 93.75, 100, 37.5 and 75.0 for education/student-funding

systems and 62.5, 100, 62.5 and 75.0 for public-order/regulatory systems (Figure 2). There were no statistically significant differences in domain scores for legal accountability ($H = 1.95$, $p =$

.378) or fairness ($H = 1.16$, $p = .559$) or contestability ($H = 5.19$, $p = .075$) (Table 4). There was no testing for transparency as all cases received 100 points.

Table 4. Cross-Domain Kruskal–Wallis Tests

Dimension	H	p	Interpretation
Legal accountability	1.95	.378	Not significant
Transparency	Not tested	—	Zero variance; all cases = 100
Fairness	1.16	.559	Not significant
Contestability	5.19	.075	Not significant at $p < .05$

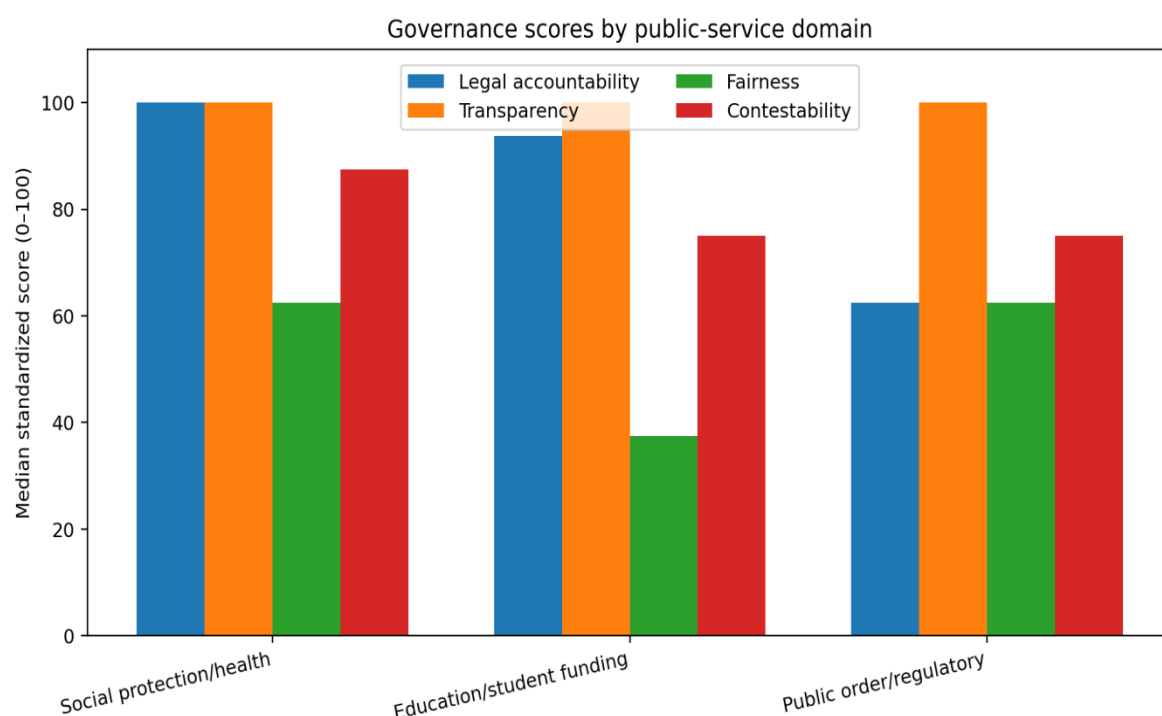


Figure 2. Median governance safeguard scores by public-service domain.

4.4 Governance Safeguards and Social Inequality

The study's strict criterion for a documented inequality outcome was not fulfilled in any of the cases as the official records did not show verified group-level disparities that could be attributed to the system deployed. In ten of 12 cases (83.3%), inequality vulnerability was found using a predefined inequality classification rule (Table 5). The Reception Baseline Assessment routing record was not identified as vulnerable, as it contained no consequential individual decision from the routing process; the MOT Risk Rating

was more about regulated testers and garages and not an equality sensitive beneficiary population. A total of 5 systems (41.7%) reported completing a formal equality-specific assessment and one more system reported planned (8.3%). There was no association or regression modelling possible due to the outcome variable being documented-inequality being fixed at zero, thus any unsupported causal claims were not possible from the available secondary evidence.

Table 5. Equality-Safeguard and Inequality Classification Summary

Indicator	Cases (n = 12)	Percentage
Documented inequality outcome	0	0.0%
Inequality vulnerability	10	83.3%
Completed equality-specific assessment	5	41.7%
Equality-specific assessment planned	1	8.3%

5. Discussion

The findings revealed that there was a variable level of governance maturity in the 12 automated public-service systems. The legal accountability and transparency showed the highest medians while accountability and transparency that protect fairness were less reported. Although ten systems (83.3%) had at least one inequality vulnerability, equality-specific assessments were found in only five of these (41.7%). This distinction implied that there did not necessarily have to be extensive research to examine the possible unequal effects of the automated system published. In a similar manner the OECD framework considered transparency, fairness, robustness, and accountability as separate, but interrelated, elements of trustworthy AI and not mutually exclusive (OECD, 2019).

The uniform transparency score was tricky to interpret, as inclusion was based on the quality of ATRS records which were developed specifically to reveal the algorithmic use in public-sector operations. The outcome, therefore, was a positive finding on documentation within the corpus as opposed to offering evidence of transparency in the broader picture of public-sector systems in the United Kingdom. There was no consistent disclosure of testing equality or strong challenge procedures. The U.S. Government Accountability Office structured AI accountability into four complimentary domains—governance, data, performance and monitoring—and asserted that disclosure was just one aspect of responsible system management (U.S. Government Accountability Office, 2021). NIST also arrived at a similar conclusion in their four functions of AI risk management: Govern, Map, Measure, and Manage, with a focus not just

on transparency, but on ongoing identification, measurement, and treatment of risks (National Institute of Standards and Technology, 2023).

The lack of inequality inequality outcomes on the group level does not mean that the systems were socially neutral. Of the 12, 10 systems fulfilled the inequality-vulnerability criterion of the study, and only 5 systems (41.7%) completed an assessment of equality-specific systems. The results thus reflected a more general picture of algorithmic inequality as an institutional and legal challenge that is influenced by administrative procedures, procurement processes, data selection, judicial and human discretion, oversight, and remedies. The Council of Europe also took responsibility over the procurement, design, development, deployment and supervision of algorithmic systems and made the connection between those stages and already existing human rights obligations (Council of Europe, Committee of Ministers, 2020).

The results also identified the need for more robust protections in the entire decision-making process. Pre-deployment impact assessment should be done in tandem with regular disparity testing, with human override record kept, and available independent review, understandable explanations and a route for appeal and correction, by public authorities. The United Nations Educational, Scientific and Cultural Organization (UNESCO) adopted the Recommendation on the Ethics of Artificial Intelligence by 193 Member States, which puts human dignity, equality, non-discrimination, transparency and human oversight at the heart of responsible AI governance (UNESCO, 2021). The European Union's Artificial Intelligence Act

(AIA) further expressed this risk-based principle by adding supplementary requirements on systems listed as being high risk, such as those that are related to fundamental public and private services (EU Regulation 2024/1689, 2024). The empirical results indicated that visible algorithms were not enough for accountable automated public services. They needed institutions that could identify unequal effects, explain decisions that have consequences, determine accountability and offer substantive avenues for those who were affected to seek review and remedy.

6. Limitations and Future Research

The study was limited by its reliance on publicly disclosed secondary information and by the relatively small, UK-focused sample of 12 systems. Because eligibility required sufficiently complete ATRS documentation, the design likely selected for comparatively transparent systems and may overstate transparency relative to the wider population of public-sector algorithmic tools. Differences in documentation quality also meant that an unreported safeguard or disparity could not always be interpreted as genuinely absent. The lack of variation in documented inequality outcomes prevented regression analysis and restricted causal interpretation. Future research should expand the dataset across jurisdictions and public-service domains, link transparency records with audit findings, appeals, complaints, and demographic performance data, and examine systems longitudinally. Such evidence would allow stronger testing of whether improved governance safeguards were followed by measurable reductions in unequal outcomes.

References

- Allhutter, D., Cech, F., Fischer, F., Grill, G., & Mager, A. (2020). Algorithmic profiling of job seekers in Austria: How austerity politics are made effective. *Frontiers in Big Data*, 3, Article 5. <https://doi.org/10.3389/fdata.2020.00005>
- Alston, P. (2019). Extreme poverty and human rights (A/74/493). United Nations General Assembly.
- Berman, A., de Fine Licht, K., & Carlsson, V. (2024). Trustworthy AI in the public sector: An empirical analysis of a Swedish labor market decision-support system. *Technology in Society*, 76, Article 102471. <https://doi.org/10.1016/j.techsoc.2024.102471>
- Berryhill, J., Kok Heang, K., Clogher, R., & McBride, K. (2019). Hello, World: Artificial intelligence and its use in the public sector (OECD Working Papers on Public Governance No. 36). OECD Publishing. <https://doi.org/10.1787/726fd39d-en>
- Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of Machine*

7. Conclusion

Algorithmic decision-making in public services offered clear administrative benefits, but its use also introduced important concerns about legal accountability, transparency, fairness, and social inequality. Formal transparency and institutional responsibility were more frequently reported than safeguards for equality, and meaningful assessment of disparate effects was less frequently reported. Vulnerability inequalities were found in the majority of the systems examined, suggesting that automated decision making might impact socially important settings even in the absence of public reporting of discriminatory outcomes. Such findings confirmed that algorithmic systems should be part of a broader administrative and legal framework and not be viewed as standalone technical tools. Good governance needed to be based on a clear responsibility, clear disclosure, periodic fairness assessment, meaningful human oversight, independent scrutiny and appeal and correction processes. There was also a need for public authorities to be aware of the impact of automated decisions on different social groups across the lifespan of the system rather than through pre-deployment assurances. Simple improvements to documentation of demographic outcomes, audit findings, complaints and reviews would help to assess in the future the impact of governance in protecting against the disproportionate effects in practice. The key to responsible automation in public services was ultimately to bring in technological efficiency and institutional protections that could foster respect for rights, public responsibility, and that digital transformation did not exacerbate existing forms of social disadvantage.

Learning Research, 81, 149–159.

- Bristol City Council. (2025, January 28). Not in Education, Employment or Training Model [Algorithmic Transparency Recording Standard record]. GOV.UK. <https://www.gov.uk/algorithmic-transparency-records/bristol-city-council-not-in-education-employment-or-training-model>
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 77–91.
- Centre for Data Ethics and Innovation. (2020). Review into bias in algorithmic decision-making. UK Government.
- 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 Machine Learning Research*, 81, 134–148.
- Cobbe, J., Lee, M. S. A., & Singh, J. (2021). Reviewable automated decision-making: A framework for accountable algorithmic systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 598–609). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445921>
- Council of Europe, Committee of Ministers. (2020). Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems.
- de Fine Licht, K., & de Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making. *AI & Society*, 35(4), 917–926. <https://doi.org/10.1007/s00146-020-00960-w>
- Department for Energy Security and Net Zero. (2025, March 2). Warm Home Discount Eligibility – Energy Cost Modelling [Algorithmic Transparency Recording Standard record]. GOV.UK. <https://www.gov.uk/algorithmic-transparency-records/desz-warm-home-discount-eligibility-energy-cost-modelling>
- Department for Work and Pensions. (2025a, October 30). Child Maintenance Service (CMS) – Predictive Analytics Compliance [Algorithmic Transparency Recording Standard record]. GOV.UK. <https://www.gov.uk/algorithmic-transparency-records/dwp-child-maintenance-service-cms-predictive-analytics-compliance>
- Department for Work and Pensions. (2025b, February 10). Employment and Support Allowance Online Medical Matching [Algorithmic Transparency Recording Standard record]. GOV.UK. <https://www.gov.uk/algorithmic-transparency-records/dwp-employment-and-support-allowance-online-medical-matching>
- Driver and Vehicle Standards Agency. (2025, February 10). MOT Risk Rating [Algorithmic Transparency Recording Standard record]. GOV.UK. <https://www.gov.uk/algorithmic-transparency-records/dvsa-mot-risk-rating>
- European Union Agency for Fundamental Rights. (2020). Getting the future right: Artificial intelligence and fundamental rights.
- Government Digital Service. (2025). Algorithmic Transparency Recording Standard: Guidance for public sector bodies. UK Government.
- Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review*, 45, Article 105681. <https://doi.org/10.1016/j.clsr.2022.105681>
- Hampshire and Thames Valley Police. (2022, July 6). DARAT (Domestic Abuse Risk Assessment Tool) [Algorithmic Transparency Recording Standard record]. GOV.UK. <https://www.gov.uk/algorithmic-transparency-records/hampshire-and-thames-valley-police-darat>
- Keddell, E. (2019). Algorithmic justice in child protection: Statistical fairness, social justice and the implications for practice. *Social Sciences*, 8(10), Article 281. <https://doi.org/10.3390/socsci8100281>
- Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy*, 44(6), Article 101976. <https://doi.org/10.1016/j.telpol.2020.101976>
- Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review*, 9(2), 1–16. <https://doi.org/10.14763/2020.2.1469>
- Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes: The premise, the proposed solutions, and the open challenges. *Philosophy & Technology*, 31(4), 611–627. <https://doi.org/10.1007/s13347-017-0279-x>

- Levy, K., Chasalow, K. E., & Riley, S. (2021). Algorithms and decision-making in the public sector. *Annual Review of Law and Social Science*, 17, 309–334. <https://doi.org/10.1146/annurev-lawsocsci-041221-023808>
- Loi, M., & Spielkamp, M. (2021). Towards accountability in the use of artificial intelligence for public administrations. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 757–766). Association for Computing Machinery. <https://doi.org/10.1145/3461702.3462631>
- National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.100-1>
- NHS Business Services Authority. (2025, September 25). Prescriptions and Dental Exemption Check [Algorithmic Transparency Recording Standard record]. GOV.UK. <https://www.gov.uk/algorithmic-transparency-records/nhsbsa-prescriptions-and-dental-exemption-check>
- Office for Students. (2025, February 27). Teaching Funding Allocations [Algorithmic Transparency Recording Standard record]. GOV.UK. <https://www.gov.uk/algorithmic-transparency-records/office-for-students-teaching-funding-allocations>
- Organisation for Economic Co-operation and Development. (2019). Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449).
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). Association for Computing Machinery. <https://doi.org/10.1145/3351095.3372873>
- Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). (2024). Official Journal of the European Union.
- Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59–68). Association for Computing Machinery. <https://doi.org/10.1145/3287560.3287598>
- Standards and Testing Agency. (2025, October 30). Reception Baseline Assessment Routing [Algorithmic Transparency Recording Standard record]. GOV.UK. <https://www.gov.uk/algorithmic-transparency-records/standards-and-testing-agency-reception-baseline-assessment-routing>
- Student Loans Company. (2025a, September 25). Advanced Learning Loans Assessment Rules [Algorithmic Transparency Recording Standard record]. GOV.UK. <https://www.gov.uk/algorithmic-transparency-records/slc-advanced-learning-loans-assessment-rules>
- Student Loans Company. (2025b, February 27). Postgraduate Assessment Rules [Algorithmic Transparency Recording Standard record]. GOV.UK. <https://www.gov.uk/algorithmic-transparency-records/student-loans-company-postgraduate-assessment-rules>
- Suresh, H., & Gutttag, J. (2021). A framework for understanding sources of harm throughout the machine learning life cycle. In *Equity and Access in Algorithms, Mechanisms, and Optimization* (pp. 1–9). Association for Computing Machinery. <https://doi.org/10.1145/3465416.3483305>
- The Probation Service. (2025, February 10). Effective Proposal Framework [Algorithmic Transparency Recording Standard record]. GOV.UK. <https://www.gov.uk/algorithmic-transparency-records/the-probation-service-effective-proposal-framework>
- Treasury Board of Canada Secretariat. (2020). Algorithmic Impact Assessment [Risk assessment tool]. Open Government Portal.
- U.S. Government Accountability Office. (2021). Artificial intelligence: An accountability framework for federal agencies and other entities (GAO-21-519SP).
- United Nations Educational, Scientific and Cultural Organization. (2021). Recommendation on the ethics of artificial intelligence.
- Veale, M., Van Kleek, M., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Paper 440, pp. 1–14). Association for Computing Machinery. <https://doi.org/10.1145/3173574.3174014>

- Wachter, S., Mittelstadt, B., & Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review*, 41, Article 105567. <https://doi.org/10.1016/j.clsr.2021.105567>
- Wieringa, M. (2020). What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 1–18). Association for Computing Machinery. <https://doi.org/10.1145/3351095.3372833>
- Wong, P.-H. (2020). Democratizing algorithmic fairness. *Philosophy & Technology*, 33(2), 225–244. <https://doi.org/10.1007/s13347-019-00355-w>
- Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). Transparency in algorithmic and human decision-making: Is there a double standard? *Philosophy & Technology*, 32(4), 661–683. <https://doi.org/10.1007/s13347-018-0330-6>
- Zuiderwijk, A., Chen, Y.-C., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly*, 38(3), Article 101577. <https://doi.org/10.1016/j.giq.2021.101577>