societies

MDPI

Article

AI in Everyday Life: How Algorithmic Systems Shape Social Relations, Opportunity, and Public Trust

Oluwaseyi B. Ayeni [1,*], Isabella Musinguzi-Karamukyo [2], Oluwakemi T. Onibalusi [3] and Oluwajuwon M. Omigbodun [4]

[1] Department of Law, The University of Buckingham, Hunter St, Buckingham MK18 1EG, UK
[2] Department of Management and Entrepreneurship, Imperial College Business School, Imperial College London, London SW7 2AZ, UK; musinguzikaramukyoisabella@gmail.com
[3] Department of Government and Public Policy, School of Humanities and Social Sciences, University of Strathclyde, Glasgow G1 1XQ, UK; oluwakemionibalusi@gmail.com
[4] Department of Law, Afe Babalola University, Ado Ekiti 362101, Nigeria; omigbodun.juwon@gmail.com
* Correspondence: 1103117@buckingham.ac.uk

Abstract

Artificial intelligence is often framed as a neutral technical tool that enhances efficiency and consistency in institutional decision-making. This article challenges that framing by showing that automated systems now operate as social and institutional actors that reshape recognition, opportunity, and public trust in everyday life. Focusing on employment screening, welfare administration, and digital platforms, the study examines how algorithmic systems mediate social relations and reorganise how individuals are evaluated, classified, and legitimised. Drawing on regulatory and policy materials, platform governance documents, technical disclosures, and composite vignettes synthesised from publicly documented evidence, the article analyses how automated judgement acquires institutional authority. It advances three core contributions. First, it develops a sociological framework explaining how delegated authority, automated classification, and procedural opacity transform institutional power and individual standing. Second, it demonstrates a dual logic of inequality: automated systems both reproduce historical disadvantage through patterned data and generate new forms of exclusion through data abstraction and optimisation practices that detach individuals from familiar legal, social, and moral categories. Third, it shows that automation destabilises procedural justice by eroding relational recognition, producing trust deficits that cannot be resolved through technical fairness or explainability alone. The findings reveal that automated systems do not merely support institutional decisions; they redefine how institutions perceive individuals and how individuals interpret institutional legitimacy. The article concludes by outlining governance reforms aimed at restoring intelligibility, accountability, inclusion, and trust in an era where automated judgement increasingly structures social opportunity and public authority.

Keywords: algorithmic decision-making; automated judgement; social inequality; datafied citizenship; classification and visibility; procedural justice; public trust

Check for updates

Academic Editor: Cristiano Felaco

Received: 7 January 2026
Revised: 2 February 2026
Accepted: 5 February 2026
Published: 12 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction: Everyday Encounters with Automated Judgement

Across many ordinary decisions, people increasingly find themselves responding to outcomes produced by systems rather than by other human beings. A job applicant is removed from a shortlist before anyone reads their application. A family seeking social support receives an automated denial with no explanation. A content creator watches their

audience disappear overnight without knowing what rule was applied. These experiences may feel isolated or personal, yet together they reflect a wider shift in how institutions classify, recognise, and act upon individuals. Automated systems have become routine intermediaries of social life, shaping access to opportunity, structuring visibility, and influencing whose claims or aspirations are treated as legitimate.

Early scholarship often framed automation as an efficiency upgrade. Automated systems were presented as tools capable of reducing human inconsistency and improving procedural regularity, particularly in complex urban governance and administrative settings [1]. From this perspective, automation appeared as a continuation of bureaucratic rationalisation. However, research in platform governance and digital labour reveals a more complicated reality. Rather than merely applying rules, automated systems actively shape who becomes visible, influential, or marginal within digital and institutional environments [2,3]. At the same time, studies in organisational and AI governance show that assumptions embedded at the design stage frequently acquire institutional authority long before systems are deployed, embedding power into technical infrastructures that remain largely unexamined [4,5]. These findings challenge the idea that automated systems are neutral extensions of existing procedures. Instead, they suggest that such systems increasingly operate as institutional actors in their own right.

A persistent tension follows from this shift. On one hand, experimental research shows that administrators and citizens often interpret algorithmic decisions as objective or impartial, even when they cannot access the reasoning behind them [6]. On the other hand, studies of digital platforms and welfare systems show that many individuals experience automated outcomes as arbitrary, exclusionary, or unjust precisely because the rules remain opaque and unchallengeable [2,3,7]. This tension raises a sociological question that technical AI ethics debates do not fully resolve: why do automated systems continue to command authority while simultaneously producing feelings of misrecognition, alienation, and distrust?

Much existing AI research focuses on technical solutions such as bias mitigation, fairness optimisation, explainable AI, and auditing mechanisms. These approaches are valuable and can improve decision accuracy and transparency. However, this article advances a complementary claim: even systems that are technically fair or explainable can still produce social and relational harm. The central argument is that automated judgement reshapes not only outcomes, but also the relationship between individuals and the institutions that evaluate them. In doing so, it alters how recognition, legitimacy, and trust are experienced in everyday institutional encounters.

This study examines how algorithmic systems reorganise three interrelated dimensions of social life. The first is social relations, understood as how individuals experience institutional recognition or disregard. The second is opportunity, shaped by automated sorting in employment, welfare administration, and digital participation. The third is public trust, particularly in contexts where decisions appear final even though the reasoning remains inaccessible.

To explore these dynamics, the article adopts a multi-source interpretive design, drawing on policy and regulatory documents, platform governance materials, public accountability reports, and composite vignettes constructed from recurring patterns documented in public inquiries, investigative journalism, and scholarly research. The study does not rely on interviews, surveys, or direct participant observation. The vignettes used here are not real cases; rather, they are synthetic narratives designed to illustrate typical ways people encounter automated systems, allowing the analysis to examine social meaning without claiming to measure frequency or lived experience directly.

The article makes two principal contributions. Conceptually, it argues that automated judgement should be understood as a form of institutional power that operates through classification rather than deliberation. By linking insights from public governance, digital labour, and welfare automation, it shows that classification is not only a technical function but a social process that shapes recognition, credibility, and access. Empirically, it develops a dual account of inequality, demonstrating that automated systems both amplify historical disadvantage through patterned data and produce new forms of exclusion through data-driven abstraction that detach individuals from familiar legal or moral categories.

The sections that follow develop a framework for understanding algorithmic authority, outline the methodological approach, present cross-domain findings, interpret these findings through sociological theory, and propose governance reforms aimed at strengthening intelligibility, accountability, inclusion, and public trust. The central claim is that artificial intelligence is no longer peripheral to social life. It has become an emerging institutional force that reshapes how people understand judgement, opportunity, and their place within systems of public value.

2. Conceptual Architecture: Locating Algorithmic Power in the Social Realm

Automated systems are commonly presented as technical instruments designed to make routine decisions faster, cheaper, and more consistent. While this framing highlights efficiency gains, it understates the broader social implications of automation. Contemporary research increasingly indicates that automated systems do more than execute predefined rules. They function as classificatory infrastructures that influence how institutions categorise individuals, how people understand their own social standing, and how opportunities are distributed across populations. For this reason, this article treats automated systems not merely as technological tools but as institutional actors whose classifications carry social, political, and moral consequences.

Prior scholarship suggests that encounters with automated systems are shaped as much by institutional contexts as by computational logic. Fang et al. [7] show that algorithmic outputs are often perceived as authoritative even when their reasoning remains inaccessible, reflecting the institutional credibility attached to automation. Research on digital welfare infrastructures further indicates that automated scoring systems embed normative assumptions about deservingness, risk, and compliance, producing categories that implicitly express moral judgement rather than neutral assessment [8]. Related work on algorithmic imaginaries demonstrates that classification systems also shape which identities, claims, and narratives appear legitimate in digital and public communication environments [9]. These effects are especially visible in studies of marginalised communities, where algorithmic misclassification or suppression can operate as a mechanism of exclusion rather than impartial governance [10].

The institutional consequences of automation extend beyond individual classification and into broader governance relationships. Parviainen et al. [11] argue that algorithmic decision systems alter how citizens relate to public institutions by replacing interpersonal judgement with data-driven representation. In media and communication contexts, algorithmic filtering has been shown to recalibrate perceptions of credibility and authority, influencing what information is amplified and what is disregarded [12]. Grimmelikhuijsen [13] further observes that automated infrastructures increasingly perform organisational gatekeeping functions, determining which claims enter formal administrative processes and which are filtered out before human review.

Taken together, these strands of research suggest that automated systems reshape three interrelated domains: institutional authority, social visibility, and the normative

foundations of legitimacy. The conceptual framework developed in this article builds on these insights by focusing on automation as a social and institutional phenomenon rather than solely a technical one. It does not claim to measure the prevalence of specific public attitudes or lived experiences. Instead, it synthesises documented evidence from policy materials, platform governance texts, regulatory reports, and prior scholarly research to interpret how algorithmic systems restructure recognition, classification, and trust.

To operationalise this framework, Section 2.1 examines delegated authority, referring to the institutional process through which organisations treat algorithmic outputs as binding judgements. Section 2.2 analyses classification as the mechanism through which automated systems construct visibility, credibility, and social presence. Section 2.3 explores opacity and procedural justice, focusing on how limited intelligibility alters normative expectations of fairness and accountability. Finally, Section 2.4 introduces a dual-pathway model of inequality, conceptualising how automated systems both amplify historical disadvantage and generate new forms of exclusion through computational abstraction.

2.1. Delegated Authority and the Rise of Automated Judgement

Delegated authority describes the institutional process through which organisations transfer decision-making power from human officials to automated systems and subsequently treat algorithmic outputs as binding judgements. This shift does not merely replace one decision-maker with another. It transforms how decisions are justified, how responsibility is distributed, and how individuals interpret institutional fairness and legitimacy.

Importantly, delegation often occurs well before a system is deployed in practice. Herrera Poyatos et al. [14] show that value judgements are frequently embedded during early design stages, including the selection of optimisation goals, feature weighting, and audit criteria. These choices shape how systems prioritise efficiency, risk, or compliance, yet they are rarely subjected to public scrutiny or democratic deliberation. Once implemented, these design decisions acquire institutional authority even though they remain largely invisible to those affected by them.

Evidence from recruitment systems illustrates the downstream consequences of this delegation. Fabris et al. [15] document how hiring algorithms trained on historical organisational data tend to reproduce existing employment patterns. Variables such as educational background, employment gaps, or career trajectories may appear neutral, but they often function as indirect carriers of long-standing social inequalities. In this context, delegation does not eliminate bias. Instead, it stabilises and scales existing patterns under the appearance of technical objectivity.

From a broader sociological perspective, Zajko [16] argues that automated judgement replaces deliberative reasoning with what can be described as statistical inheritance. Rather than evaluating individuals in context, automated systems infer future behaviour from past patterns, effectively converting historical inequalities into predictive evidence. Atoum et al. [17] further note that this shift complicates accountability. Because the internal logic of automated systems is typically opaque, individuals cannot easily challenge decisions, even though those decisions carry institutional weight.

Taken together, these studies suggest that delegated authority should not be understood as a neutral efficiency reform. It represents a redistribution of institutional power from identifiable human actors to computational infrastructures whose assumptions are difficult to inspect, contest, or explain. Automated decisions often appear objective because their underlying criteria remain hidden, and they appear final because institutions present them as final. As a result, delegated authority reshapes the moral and procedural foundations through which legitimacy has traditionally been constructed.

This conceptualisation does not assume direct access to the lived experiences of affected individuals. Rather, it synthesises documented institutional practices and prior scholarly analyses to interpret how authority is reorganised when organisations rely on automated judgement. The next section builds on this foundation by examining how classification functions as the primary mechanism through which automated systems construct visibility, credibility, and institutional recognition.

2.2. Classification, Sorting, and the Social Construction of Visibility

Classification is the primary mechanism through which automated systems organise social visibility and institutional relevance. Through sorting, ranking, and scoring processes, automated systems determine which individuals, behaviours, or claims become visible, marginal, credible, or dismissible within institutional and digital environments. Unlike traditional bureaucratic categorisation, algorithmic classification operates continuously, at scale, and without granting those subject to it access to the rules that shape their classification.

Research on platform governance illustrates how classification structures visibility. Kojah et al. [18] show that creators often experience ranking systems as dynamic authorities that influence whether their content is surfaced or suppressed. In this context, visibility is not merely the outcome of merit or effort but is shaped by engagement-driven metrics embedded in platform architectures. Duffy and Meisner [19] further demonstrate that these classificatory processes generate stratified attention economies in which certain content styles, identities, or communicative norms are systematically advantaged over others.

Public administration scholarship highlights a parallel dynamic in institutional settings. Fang et al. [20] find that individuals frequently interpret algorithmic classifications as procedural facts, even when they lack insight into how those determinations were made. This effect is reinforced by organisational narratives that frame automated systems as neutral or objective instruments, thereby masking the political, managerial, and normative choices embedded in their design.

Welfare automation research provides additional insight into the moral dimensions of classification. Larasati et al. [21] show that risk-scoring systems encode implicit assumptions about compliance, responsibility, and deservingness, producing categories that carry symbolic and normative significance beyond their administrative function. Parviainen et al. [22] extend this analysis by arguing that algorithmic assessments can influence how individuals anticipate institutional judgement, encouraging behavioural adaptation in response to expected sorting practices. In this way, classification shapes both institutional recognition and how individuals understand their own positioning within institutional systems.

Beyond administrative and welfare contexts, classificatory infrastructures also influence epistemic authority. Walter and Farkas [23] demonstrate that algorithmic relevance-ranking systems affect which knowledge claims gain legitimacy within digital public discourse. Decker et al. [24] similarly note that automated categorisation influences perceptions of fairness by shaping the informational environment in which institutional decisions are interpreted and evaluated.

The contribution of this section lies in conceptualising classification as an institutional process rather than a purely technical operation. The existing literature often treats algorithmic sorting as either a computational function or a cultural phenomenon tied to visibility. The framework developed here integrates these perspectives by interpreting classification as a relational mechanism through which institutions construct recognition, credibility, and social presence. In this sense, automated classification does not merely process information.

It performs institutional work by defining who is seen, how they are evaluated, and what forms of participation or opportunity become available.

2.3. Trust and the Perception of Procedural Justice

Trust in institutional decision-making has traditionally rested on three core expectations: that individuals can understand how decisions are made, that their circumstances are meaningfully considered, and that outcomes can be questioned or reviewed. Automated decision systems complicate each of these expectations by relocating judgement into computational processes that are typically opaque, standardised, and procedurally closed to those affected by them.

Research in public administration suggests that automation can sometimes enhance perceived legitimacy. Goldsmith et al. [1] show that some individuals interpret algorithmic outputs as impartial or neutral, even when they cannot access the reasoning behind them. In these contexts, the presence of technology itself can operate as a symbolic marker of objectivity, encouraging the acceptance of decisions without requiring transparency.

However, studies in welfare governance and platform regulation reveal a different pattern. Larasati et al. [25] document that, when automated determinations conflict with individuals' understanding of their own circumstances, opacity is often interpreted not as technical complexity but as institutional disengagement. In these cases, unexplained outcomes are experienced as evidence that decision-making processes failed to acknowledge personal context or social reality. Rather than reinforcing trust, automation can signal detachment or indifference.

This contrast highlights a broader sociological tension in how procedural justice is understood under automation. Some policy and technical frameworks assume that improving explanation or transparency will restore legitimacy. Zajko [26], however, argues that this framing underestimates the relational dimensions of trust. The challenge is not only that algorithmic systems obscure reasoning, but that they reduce opportunities for dialogue, contestation, and recognition. Automated processes typically provide limited space for individuals to clarify meaning, express disagreement, or feel heard within institutional procedures.

Related findings reinforce this interpretation. Fang et al. [7] observe that people often treat automated decisions as final, even when they do not understand them, suggesting that authority is being relocated from interpersonal judgement to system-based classification. Platform governance research further indicates that when visibility or ranking shifts occur without explanation, users may interpret these changes as punitive or arbitrary rather than procedural [18]. Across these settings, perceptions of fairness are shaped less by technical accuracy and more by whether institutional processes appear attentive, responsive, and accountable.

The contribution of this section lies in reframing trust in automated systems as a relational rather than purely informational phenomenon. Trust is not determined solely by whether rules are disclosed or models are explainable. It is shaped by whether individuals perceive that institutions meaningfully recognise them within decision-making processes. Automated judgement can weaken this perception by replacing interpersonal evaluation with abstract classification, thereby limiting the social cues through which fairness, care, and accountability have historically been communicated.

This perspective helps explain why automation can generate acceptance in some institutional contexts while producing scepticism or resistance in others. It also clarifies why technical interventions such as explainable AI, fairness auditing, or transparency tools—while valuable—may be insufficient on their own to restore legitimacy. The challenge is not

only to make automated systems more interpretable, but to embed them within institutional practices that preserve recognition, responsiveness, and procedural voice.

2.4. Inequality Through Automated Pathways

Discussions of algorithmic inequality often frame automated systems as mirrors of society, reproducing historical disparities when training data reflect entrenched bias. While this account captures an important mechanism, it does not fully explain the new patterns of inequality emerging in automated recruitment, welfare assessment, and platform visibility systems. Automated systems do not merely inherit inequality; they actively reorganise how inequality is produced, classified, and experienced.

Figure 1 illustrates this reorganisation by conceptualising two interconnected pathways through which automated systems generate inequality: amplification and novelty. These pathways arise from the combined effects of delegated institutional authority, automated classification, and procedural opacity.

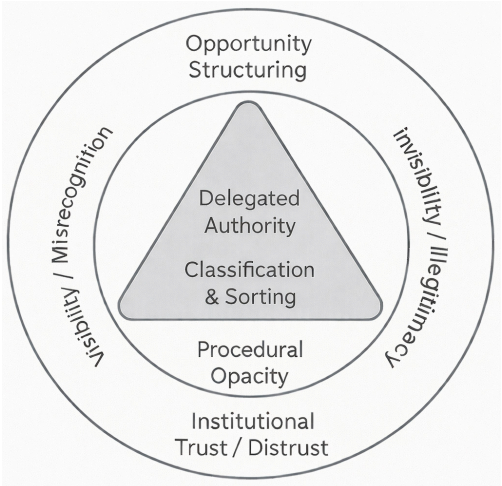


Figure 1. Automated Judgement as Institutional Power: A Relational Architecture of Delegated Authority, Classification, Opacity, and Inequality Pathways. Source: Author's conceptual model based on synthesis of policy, governance, and scholarly sources analysed in this study.

The first pathway, amplification, describes how automated systems reinforce existing social hierarchies by encoding historical patterns into predictive models. Recruitment research shows that algorithmic hiring tools often treat past employment trends as indicators of suitability, thereby stabilising gendered and racialised disparities within computational decision-making routines [27]. Zajko [28] argues that this is not simply a design flaw but a structural feature of predictive analytics, which rely on historical regularities to forecast future outcomes. As shown in Figure 1, amplification operates through statistical inheritance: past inequalities are translated into present classifications, even when sensitive demographic attributes are formally excluded.

The second pathway, novelty, refers to inequalities produced through computational abstraction rather than historical replication. Automated systems increasingly generate categories that have no direct counterpart in legal, social, or moral frameworks. Welfare scoring systems, for instance, assign individuals to newly constructed classifications such as predicted non-compliance or administrative risk—labels that citizens may not recognise as meaningful descriptions of their lived circumstances [8]. Parviainen et al. [11] further demonstrate that algorithmic profiling can assign people to categories defined entirely within the internal logic of a model, detached from conventional identity markers or institutional norms. As represented in Figure 1, this pathway produces inequality by creating new classificatory regimes rather than reproducing old ones.

Platform governance provides a clear illustration of how these two pathways interact. Kojah et al. [18] show that marginalised creators can experience disproportionate suppression not only because of pre-existing inequalities, but also because platform optimisation metrics privilege particular behavioural patterns, aesthetic styles, or temporal rhythms. Duffy and Meisner [19] similarly document how visibility hierarchies emerge from engagement-based ranking systems that generate new standards of credibility and relevance. These hierarchies are not direct extensions of offline social structures; they are produced by platform architectures themselves, as reflected in Figure 1's depiction of optimisation-driven classification.

What the existing scholarship has largely lacked is a framework that accounts for both amplification and novelty within a single institutional model. Research on predictive policing, automated eligibility, and platform moderation often isolates one mechanism while overlooking the others. Figure 1 responds to this gap by situating both pathways within a unified structure of automated judgement, showing how inherited inequality and newly generated exclusion operate simultaneously rather than independently.

This dual process has important implications for how inequality is understood and experienced. When automated systems amplify historical disadvantage, individuals may interpret outcomes as continuations of familiar injustice. When systems generate novel classifications, individuals may instead experience decisions as arbitrary, opaque, or disconnected from recognisable social categories. In both cases, inequality extends beyond material disadvantage to affect how fairness, legitimacy, and institutional responsibility are perceived.

The contribution of this section lies in reframing algorithmic inequality as a two-track process that cannot be reduced to bias alone. Amplification accounts for continuity with past hierarchies, while novelty explains how automated systems create new forms of exclusion that reshape identity, opportunity, and trust. As synthesised in Figure 1, these pathways operate together within contemporary automated systems, helping to explain why algorithmic inequality can appear both familiar and unprecedented at the same time.

3. Methodology: A Multi-Source Sociological Inquiry

Examining how automated systems reshape opportunity, identity, and public trust requires a method that captures both institutional design and social meaning. Many algorithmic systems operate behind organisational and technical opacity, meaning neither public policy documents nor engineering disclosures alone provide a full account of how decisions are produced or how they are experienced. To address this challenge, this study adopts a qualitative, multi-source research design that analyses institutional texts, platform governance materials, technical system descriptions, and composite vignettes derived from publicly documented evidence.

The core empirical foundation of the study consists of policy frameworks, regulatory reports, oversight reviews, platform governance guidelines, transparency disclosures, and publicly available technical documentation relating to automated decision systems in employment, welfare administration, and digital platforms. These materials are treated not as neutral records, but as institutional artefacts that reveal how organisations frame automation, justify algorithmic authority, and embed normative assumptions about risk, merit, compliance, visibility, and fairness. Prior research supports the analytical value of such sources, demonstrating that official documentation often exposes the discursive and moral logics through which automated systems gain legitimacy, even when their internal mechanics remain inaccessible [7,8,18].

However, institutional documents alone cannot explain how automated decisions are interpreted, felt, or socially negotiated. To capture this experiential dimension, the study

employs composite vignettes constructed from recurring patterns documented in regulatory inquiries, investigative journalism, public complaints, audit reports, and the existing scholarly literature. These vignettes do not represent real individuals and do not claim empirical measurement of frequency. Instead, they function as analytical devices designed to illustrate typical modes of encounter with automated judgement—including experiences of exclusion, misrecognition, uncertainty, and behavioural adaptation—without involving direct human participants. This approach reflects findings in digital labour and welfare research showing that people interpret algorithmic outcomes through relational, moral, and emotional frameworks that are not visible in technical records alone [19,25].

The methodological strategy rests on two central commitments. First, institutional texts are analysed as expressions of the value systems, incentives, and governance priorities that shape automated decision infrastructures. Second, vignette-based interpretation is used to surface how these infrastructures are likely to be experienced in everyday social contexts, drawing on documented public evidence rather than primary fieldwork. Together, these sources enable a sociologically grounded analysis of algorithmic authority that traces how automation operates across design, organisational framing, and social interpretation.

The subsections that follow specify how sources were selected, how policy and platform materials were analysed, how composite vignettes were constructed, and how thematic and conceptual insights were derived. This structure ensures transparency about the study's scope, strengthens methodological reproducibility, and situates the findings as interpretive and theory-building rather than statistical or ethnographic claims.

3.1. Rationale for a Multi-Source Design

Automated decision systems cannot be meaningfully understood through a single type of evidence. Technical documentation offers limited insight into how systems operate in practice. Policy texts describe institutional intentions but frequently conceal how automation reshapes power, discretion, and responsibility. Platform governance materials outline formal rules yet rarely reveal how ranking, filtering, or visibility mechanisms actually function in everyday use. Relying on only one method would therefore produce an incomplete and potentially distorted account of algorithmic influence.

This study adopts a multi-source qualitative design in order to capture the technical, institutional, and social dimensions of automated judgement. The approach allows the analysis to examine how automated systems are justified in policy, implemented within organisations, and experienced in real-world social settings.

One component of the dataset consists of institutional and regulatory documents. These include public sector automation strategies, algorithmic accountability frameworks, regulatory oversight reports, audit findings, procurement records, platform transparency disclosures, and governance guidelines relating to automated hiring, welfare administration, and digital content ranking. These materials are analysed because they reveal how institutions define fairness, assign responsibility, frame automation as legitimate, and embed normative assumptions about risk, efficiency, and deservingness. Prior research demonstrates that such documents provide insight into the values and power structures shaping automated systems, even when technical details remain inaccessible [7,8,18].

A second source category includes platform governance materials and publicly available technical documentation. This includes moderation policies, ranking explanations, recommendation system disclosures, engineering publications, patent filings, and official statements describing algorithmic design priorities. These materials are not treated as complete representations of system logic. Instead, they are analysed as institutional narratives that shape public expectations and legitimise algorithmic authority. Existing scholarship

shows that platforms often present a rhetoric of transparency while maintaining significant operational opacity [19,25].

A third source consists of composite vignettes developed from recurring patterns documented in regulatory inquiries, public complaints, investigative journalism, court records, and peer-reviewed research. These vignettes do not represent real individuals, nor do they claim to measure population-level prevalence. Rather, they function as structured interpretive tools that synthesise publicly documented experiences into analytically coherent scenarios. This method enables the study to examine how automated decisions are interpreted and felt in everyday contexts while preserving ethical safeguards and avoiding reliance on unverifiable anecdotal claims.

The study does not present itself as a statistical assessment of frequency or causal impact. Instead, it is a qualitative and interpretive investigation designed to identify recurring mechanisms, institutional logics, and relational consequences of automated decision-making across domains. The multi-source strategy strengthens analytical credibility by enabling triangulation across policy discourse, governance practice, technical framing, and documented public encounters.

By clearly defining the categories of sources used, the institutional contexts examined, and the interpretive role of composite vignettes, this design directly addresses concerns about transparency, traceability, and reproducibility. Readers are therefore able to understand the scope, limits, and evidentiary basis of the study's claims.

3.2. Policy and Regulatory Document Analysis

Policy and regulatory documents provide a critical lens for examining how automated decision systems acquire institutional authority and legitimacy. These texts do not merely describe how systems function; they reveal the normative assumptions, governance priorities, and accountability narratives that shape why and how automation is adopted in public and quasi-public settings. For this reason, policy materials are treated in this study not as neutral background sources but as sociologically meaningful artefacts that encode institutional power, moral judgement, and public justification.

Consistent with earlier research, institutional documentation often offers insight into how organisations frame automated systems, even when the technical logic remains opaque [7]. Larasati et al. demonstrate that welfare policies and administrative manuals frequently embed implicit moral assumptions about risk, compliance, and deservingness, which later become operationalised through automated scoring systems [8]. In a similar vein, platform governance research shows that formal rules and transparency statements shape user expectations while concealing deeper optimisation priorities that determine visibility and ranking outcomes [18].

The document corpus analysed in this study includes public automation strategies, regulatory oversight reports, algorithmic accountability frameworks, transparency guidelines, data protection assessments, and governance statements relevant to employment screening, welfare administration, and digital platform moderation. These materials were drawn primarily from governance contexts in the United Kingdom, the European Union, and North America, with publication dates spanning approximately 2018 to 2025. This timeframe captures the period in which automated decision systems transitioned from pilot initiatives into routine institutional infrastructures.

Selection criteria focused on documents that explicitly addressed automated or algorithmic decision-making, institutional accountability, fairness, transparency, risk management, or public legitimacy. Sources were chosen where they offered substantive engagement with governance questions rather than purely technical descriptions or promotional narratives.

The analytical focus was interpretive rather than technical. Documents were examined for how they define decision-making authority, construct fairness and legitimacy, frame the delegation of judgement to automated systems, and describe institutional responsibility when automated outcomes cause harm. Particular attention was paid to how policies articulate categories such as risk, fraud, efficiency, deservingness, neutrality, and public trust, as these concepts frequently shape the social meaning of automated classifications.

This approach aligns with scholarship showing that organisational narratives play a powerful role in legitimising algorithmic authority, often well before systems are publicly deployed [1,7]. It also reflects findings that governance documents frequently present automation as objective or neutral while embedding value-laden assumptions that influence how individuals are classified and treated [8].

In addition to policy texts, oversight and audit reports were analysed to identify tensions between stated institutional commitments and documented real-world effects. This enabled the study to trace where accountability gaps emerge, how institutional responsibility becomes diffused, and how automated decision systems are justified even when their social consequences remain contested.

Rather than treating policy materials as direct evidence of system performance, this study uses them to reconstruct the institutional narratives that surround automation. These narratives shape how automated decisions come to be accepted as authoritative, how responsibility is framed, and how legitimacy is socially produced. By grounding the analysis in clearly defined policy sources while maintaining an interpretive lens, this section strengthens both methodological transparency and analytical credibility.

3.3. Platform Governance and Technical Documentation Review

Digital platforms are among the most visible sites where automated systems shape public voice, credibility, and access to attention. Ranking and recommendation systems determine whose content circulates widely, whose labour becomes economically viable, and whose perspectives remain marginal. To understand these dynamics, this study examines both platform governance documents and publicly available technical materials as sources of institutional meaning and infrastructural power.

Platform governance texts include community standards, content moderation rules, ranking disclosures, and transparency reports published by major social and content platforms. These materials provide insight into how platforms justify automated decision-making, communicate fairness commitments, and shape user expectations about visibility and legitimacy. However, as Kojah et al. show, governance narratives frequently present simplified explanations that conceal the deeper optimisation logics that actually structure visibility outcomes [18]. Duffy and Meisner similarly demonstrate that platforms often portray stability and predictability while routinely adjusting algorithmic parameters in ways that creators cannot anticipate or contest [19]. In this study, such documents are therefore treated not as direct descriptions of algorithmic function but as institutional narratives that reveal platform priorities and power relations.

Publicly accessible technical materials offer a complementary perspective. Although full algorithmic models remain proprietary, platforms and affiliated researchers routinely publish engineering blog posts, research summaries, patent filings, and technical white papers that disclose ranking objectives, feedback signals, and performance metrics. These sources make it possible to infer the behavioural incentives and commercial priorities embedded in platform architectures. Zajko observes that attention-driven ranking systems are structured to maximise engagement rather than user wellbeing or representational equity, thereby privileging particular communication styles and participation patterns [16]. Parviainen et al. further show that technical classification systems shape how individuals

understand their own social positioning, particularly when algorithmic feedback becomes a key measure of relevance or worth [22].

The analytical approach applied to these materials is interpretive rather than computational. Governance and technical texts were examined for how they define relevance, construct quality, classify user behaviour, and justify automated authority. This interpretive stance follows Walter and Farkas, who demonstrate that algorithmic documentation functions as a narrative device that frames automated decisions as reasonable, neutral, or inevitable, even when their social consequences remain opaque [23]. Reading these materials as cultural and institutional artefacts allows the analysis to uncover how platforms construct legitimacy while simultaneously restricting meaningful accountability.

Across the reviewed sources, several consistent patterns emerged. Platforms publicly emphasise transparency while withholding information about the ranking signals that most strongly influence visibility. They promote ideals of fairness and merit while relying on engagement metrics that advantage specific temporal rhythms, linguistic styles, aesthetic conventions, and behavioural norms. They assert neutrality while embedding optimisation priorities that align more closely with commercial performance than with democratic or civic values. These dynamics resonate with Lopezmalo's findings that marginalised creators often experience unexplained suppression that cannot be justified through formal platform rules but can be traced to opaque ranking incentives [10].

Integrating governance narratives with technical disclosures serves two analytical purposes. First, it reveals how platforms frame automated authority in ways that manage public perception and deflect responsibility. Second, it exposes the infrastructural mechanisms through which visibility becomes a form of social power that distributes recognition, opportunity, and symbolic legitimacy. This component of the methodology is therefore essential for understanding how individuals encounter algorithmic judgement in digital spaces and how platform architectures contribute to the reproduction and transformation of inequality.

3.4. Composite Vignettes as Analytical Tools

Automated decision systems are rarely experienced through policy documents or technical interfaces alone. They are encountered through concrete moments in everyday life, such as receiving an unexplained welfare denial, being excluded from a recruitment shortlist, or watching online visibility decline without warning. These encounters often generate confusion, frustration, resignation, or adaptation, yet such experiential dimensions are difficult to capture through institutional texts or technical disclosures alone.

Because this study does not rely on interviews, surveys, or direct observation of individuals, it uses composite vignettes as an interpretive method to examine how automated decisions are commonly experienced and understood. The vignettes are not representations of real individuals. Instead, they are carefully constructed narrative syntheses drawn from recurring patterns documented in public inquiries, regulatory reports, investigative journalism, court cases, and prior peer-reviewed scholarship. Their purpose is not to measure frequency or prevalence, but to illustrate typical ways in which automated authority is encountered and interpreted across institutional settings.

This approach builds on prior research showing that people interpret algorithmic decisions through moral, relational, and identity-based frameworks rather than purely technical reasoning. Studies of welfare automation demonstrate that individuals often evaluate automated outcomes in terms of dignity, recognition, and fairness rather than procedural correctness [25]. Research on digital labour similarly shows that creators interpret algorithmic visibility shifts through narratives of value, legitimacy, and exclusion [19]. Composite vignettes allow these interpretive dynamics to be analysed in a structured way

while avoiding the ethical and procedural requirements associated with collecting primary data from human participants.

To ensure transparency and methodological integrity, each vignette was constructed from publicly documented cases rather than imagined scenarios. Source materials included published regulatory investigations, parliamentary inquiries, civil society reports, platform transparency disclosures, legal challenges, and academic studies. The vignettes therefore do not invent experiences but condense recurring patterns into analytically coherent scenarios that reflect institutional and social realities reported in existing evidence.

The use of vignettes serves two primary analytical functions. First, it allows the study to examine how automated decisions are perceived and interpreted in practice, rather than only how they are designed or justified by institutions. While governance documents and technical materials reveal how automated systems are framed, they cannot fully capture how people experience opacity, misclassification, or denial. Vignettes bridge this gap by showing how institutional decisions are emotionally processed, morally evaluated, and socially negotiated.

Second, vignettes enable a systematic comparison across domains. For example, patterns observed in welfare automation can be analytically compared with experiences documented in platform governance or hiring systems, allowing the study to identify cross-domain similarities in how automated authority affects recognition, opportunity, and trust. This comparative function supports the broader goal of tracing structural dynamics that extend beyond any single institutional setting.

It is important to clarify the limitations of this method. Because the study does not involve direct engagement with participants, the vignettes do not claim to represent lived experience in a statistical or demographic sense. They are interpretive tools rather than empirical measurements. Their value lies in illustrating social mechanisms and meaning making processes, not in quantifying prevalence or predicting individual behaviour. This limitation is acknowledged explicitly to ensure that conclusions drawn from the vignettes are understood as conceptual and analytical rather than empirical generalisations.

The analytical strategy applied to all sources treats automated decision systems as institutional actors rather than neutral technical instruments. This follows sociotechnical and public administration research showing that automated tools acquire authority through organisational endorsement, legal framing, and cultural legitimacy as much as through computational design [7,8,10]. The analysis therefore reconstructs how institutional priorities, model design choices, governance narratives, and public interpretation converge to produce outcomes experienced as authoritative, legitimate, or exclusionary.

The interpretive process unfolded in three stages. First, institutional and platform materials were examined thematically to identify recurring assumptions about fairness, risk, compliance, relevance, and legitimacy. Second, these institutional logics were connected to documented public responses, reported harms, and observed adaptation strategies through process tracing that mapped how technical design choices translate into lived consequences. Third, insights from all materials were organised using the analytical matrix presented in Table 1, which clarifies how different categories of evidence contribute to understanding automated authority as a social structure rather than a technical output.

This matrix functions as more than a summary device. It serves as the analytical backbone of the study by linking institutional rationales, classification mechanisms, behavioural adaptation, symbolic legitimacy, and inequality outcomes across domains. Through this structure, the study identifies recurring mechanisms such as the normalisation of opacity as a substitute for justification, the transformation of classification into a signal of recognition or exclusion, and the dual production of inequality through both historical reproduction and data-driven novelty.

Table 1. Conceptual–Methodological Matrix for Analysing Algorithmic Power as Institutional Authority Across Evidence Types and Domains. Source: Author's analytical framework and classification of source materials used in this study.

Analytical Dimension	Evidence Base (Explicit Source Types)	Domain Coverage	Analytical Purpose	Core Sociological Mechanism Identified	Concrete Output/Insight	Methodological Contribution
Institutional Rationalities	Public automation strategies, welfare policy frameworks, platform governance reports	Public administration, welfare, platforms	Identify organisational goals shaping automation	Institutional legitimation of automated authority	Reveals how efficiency narratives justify delegation of power	Improves transparency about normative assumptions
Legal and Regulatory Structures	Data protection laws, AI governance guidelines, regulatory audits	Cross-sector	Compare formal legal protections with operational realities	Gap between legal accountability and lived impact	Demonstrates limits of legal safeguards in relational harms	Strengthens reproducibility of document analysis
Organisational Incentives	Internal strategy papers, cost-reduction reports, risk management plans	Public and private institutions	Examine economic and bureaucratic pressures shaping automation	Risk displacement and responsibility diffusion	Explains why institutions prefer opaque automation	Grounds claims in traceable organisational evidence
Technical Model Architectures	Public technical reports, patent filings, engineering disclosures	Platforms, hiring, welfare	Analyse how assumptions enter system design	Encoded normativity and data abstraction	Shows how bias and exclusion become infrastructural	Connects technical artefacts to sociological interpretation
Classification and Ranking Systems	Platform ranking rules, moderation protocols, recommender system descriptions	Digital platforms	Examine how visibility and relevance are constructed	Algorithmic stratification of attention	Reframes visibility as institutional power	Links governance texts to lived outcomes
System Behaviour Under Uncertainty	Error reports, proxy variable disclosures, audit summaries	Cross-sector	Identify unpredictable or unjust system outcomes	Computational arbitrariness	Introduces arbitrariness as a sociological harm	Improves methodological specificity
Delegation of Discretion	Administrative delegation policies, automation workflow manuals	Public administration, welfare	Track movement of decision authority from humans to systems	Displaced accountability	Shows how responsibility becomes structurally obscured	Clarifies institutional chain of responsibility
Publicly Documented Lived Experiences	Ombudsman complaints, investigative journalism, parliamentary hearings	Welfare, platforms, hiring	Extract recurring experiential patterns	Moral injury, misrecognition, relational erosion	Reframes harm as relational rather than technical	Enhances empirical grounding

Table 1. *Cont.*

Analytical Dimension	Evidence Base (Explicit Source Types)	Domain Coverage	Analytical Purpose	Core Sociological Mechanism Identified	Concrete Output/Insight	Methodological Contribution
Composite Vignettes	Synthesised patterns from verified public cases and reports	Cross-domain	Model typical user–system encounters	Identity negotiation under automation	Enables experiential analysis without human subjects	Responds directly to ethics and traceability concerns
Behavioural Adaptation Patterns	Prior platform and welfare studies, reported coping behaviours	Platforms, welfare	Analyse how individuals adapt to opaque systems	Anticipatory compliance, self-surveillance	Explains behavioural reshaping under algorithmic control	Extends sociological interpretation of adaptation
Institutional Counterfactuals	Archived pre-automation procedures, legacy service protocols	Public services	Contrast human-centred vs automated judgement	Loss of empathetic discretion	Identifies social functions automation removes	Strengthens historical grounding
Cross-Domain Mechanism Mapping	Hiring, welfare, finance, platforms, governance	Multi-domain	Identify shared causal logics	Generalised algorithmic governance	Shows automation as systemic institutional regime	Increases analytical generalisability
Symbolic and Cultural Framing	Media narratives, public AI discourse, cultural representations	Societal level	Analyse how automation gains symbolic authority	Objectivity myth, technological legitimacy	Explains why opaque systems still command trust	Adds cultural explanatory depth
Theoretical Synthesis	Sociological theory on legitimacy, inequality, categorisation	Cross-disciplinary	Integrate empirical insights into conceptual contributions	Datafied citizenship, relational legitimacy, institutional AI	Establishes ADM as institutional power	Anchors article's core theoretical contribution

Taken together, this methodological approach strengthens the article's central claim. Automated judgement is not merely a computational procedure. It is a form of institutional power that operates through classification, narrative framing, and social interpretation. Composite vignettes, when grounded in public evidence and analysed transparently, provide a rigorous way to connect institutional design with the everyday meanings people attach to automated decisions.

3.5. Analytical Strategy and Evidence Integration

The analytical strategy adopted in this study is designed to ensure that interpretations remain grounded in documented evidence while allowing for sociological theorisation about how automated authority operates across institutional settings. Rather than treating automated systems as isolated technical artefacts, the analysis reconstructs the institutional, legal, technical, and interpretive processes through which algorithmic decisions become authoritative in everyday life.

All materials were examined through a structured thematic reading aimed at identifying recurring logics of judgement, classification, legitimacy, and institutional responsibility. Policy texts, regulatory reports, welfare manuals, platform governance documents, and technical disclosures were analysed to uncover the assumptions they encode about risk, deservingness, relevance, compliance, and credibility. This step follows prior research showing that legitimacy is shaped not only by outcomes but also by the narratives institutions construct to justify automated decisions [7,8].

To strengthen transparency and analytical traceability, sources were grouped into defined categories, including regulatory and policy documents, platform governance materials, technical system descriptions, legal and investigative reports, and composite vignettes derived from publicly documented cases. Each category was treated according to its epistemic role, meaning that institutional texts were analysed as expressions of organisational intent, technical materials as indicators of design priorities, and vignettes as interpretive syntheses of reported social experience. This layered approach ensures that claims about institutional behaviour, system design, and lived interpretation are not conflated.

The study applies process tracing to connect institutional design choices with their observable social consequences. This involved mapping how organisational objectives, optimisation targets, training data assumptions, and governance narratives translate into classification outcomes that individuals encounter as final or authoritative. Through this approach, it becomes possible to trace how discretion shifts from frontline workers into computational systems, how historical data patterns shape present decision-making, and how opacity influences perceptions of fairness and recognition [4,8,10].

To enhance coherence and cross-domain comparison, insights from all source categories were organised using the analytical matrix presented in Table 1. The matrix does not merely summarise evidence. It structures how institutional rationalities, legal frameworks, model design choices, behavioural responses, and symbolic legitimacy intersect to produce the social effects of automated judgement. By aligning each analytical dimension with a specific category of source material, the matrix clarifies how conclusions are grounded in documented evidence rather than speculative inference.

For example, rows addressing institutional rationality and legal framing draw directly on regulatory and policy sources. Rows concerning classification and optimisation reflect insights from technical documentation and platform governance texts. Rows analysing lived interpretation and trust are informed by composite vignettes synthesised from publicly reported cases. This structure improves reproducibility by making explicit which

forms of evidence support each analytical claim, even where proprietary system details remain unavailable.

Across domains including recruitment, welfare administration, and platform governance, the analysis identifies recurring mechanisms. These include the institutional normalisation of opacity as a substitute for justification, the use of classification as a signal of legitimacy or exclusion, the reproduction of inequality through statistical inheritance, and the emergence of new forms of stratification through data driven abstraction. By tracing these mechanisms across multiple evidence sources, the study avoids reliance on single-case interpretations and instead builds a cumulative argument about how automated authority functions socially.

The analytical strategy does not claim to measure prevalence, predict outcomes, or reconstruct proprietary system behaviour. Its aim is interpretive rather than predictive. It seeks to explain how automated systems become socially meaningful, how they acquire legitimacy, and how they reshape recognition, opportunity, and trust. By grounding each interpretive step in traceable source categories and clearly distinguishing empirical documentation from conceptual synthesis, the approach responds directly to concerns about methodological transparency and reproducibility.

In sum, this strategy allows the study to move beyond surface descriptions of automated systems and toward a sociological explanation of how algorithmic judgement becomes embedded in institutional practice and everyday social life.

3.6. Methodological Limitations

This study uses a document-based and vignette-driven qualitative design. It does not measure how frequently the observed patterns occur, nor does it claim statistical generalisability. The composite vignettes are interpretive syntheses based on publicly documented cases rather than firsthand interviews, meaning the analysis captures typical interaction patterns rather than individual life histories.

The findings should therefore be understood as conceptual and sociological explanations of how automated systems shape recognition, opportunity, and trust, rather than empirical claims about prevalence. While this approach limits direct observation of lived experience, it enables systematic examination of institutional logics, governance practices, and public evidence that would be difficult to access through participant-based research alone.

4. Findings: Cross-Domain Patterns in Automated Judgement

This section presents patterns observed across policy documents, regulatory reports, platform governance materials, public complaints, investigative journalism, and composite vignettes constructed from publicly documented cases. Rather than treating automated hiring, welfare systems, and platform algorithms as separate technical domains, the findings highlight recurring institutional dynamics in how automated systems classify individuals, shape access to opportunities, and influence perceptions of legitimacy.

Across sources, a consistent pattern emerges: automated systems are experienced not only as tools that process information but as institutional decision-makers whose outputs carry social meaning. Public records, oversight reports, and platform transparency disclosures repeatedly show that individuals interpret automated outcomes as signals about their credibility, deservingness, employability, or relevance, even when no explanation is provided. These interpretations appear across employment screening, welfare eligibility processes, and digital visibility regimes.

The findings do not claim to measure the frequency of these experiences. Instead, they identify recurrent institutional patterns documented across multiple public sources. Composite vignettes are used as illustrative devices to synthesise these recurring patterns

and to demonstrate how institutional decisions may be encountered in everyday contexts. They do not represent real individuals, nor do they serve as empirical case studies. Rather, they summarise typical modes of interaction reported in public evidence.

Four cross-domain patterns are especially salient in the reviewed materials. First, automated systems frequently replace interpersonal judgement with silent or opaque classification, leaving affected individuals without a clear understanding of how decisions were made. Second, automated decisions tend to function as institutional verdicts rather than provisional recommendations, even when mechanisms for appeal exist in principle. Third, automated classification systems appear to reshape how individuals understand their own status within institutional hierarchies, particularly when outcomes affect employment prospects, access to welfare, or online visibility. Fourth, patterns of inequality arise not only from the reproduction of historical disadvantage but also from the creation of new data-driven categories that lack clear grounding in legal or social norms.

The subsections that follow examine how these patterns manifest in three domains where automated judgement is highly visible: recruitment and hiring, welfare administration, and digital platform governance. Each subsection first outlines documented institutional patterns, then illustrates how these patterns may be encountered through composite vignettes and finally identifies the sociological implications that arise from these observations.

4.1. Automated Recruitment and the Normalisation of Silent Exclusion

Public reports, regulatory discussions, and hiring platform disclosures indicate that automated recruitment systems are increasingly used to screen, rank, and filter job applicants before any human review occurs. These systems typically rely on historical hiring data, CV parsing tools, psychometric indicators, and behavioural proxies to predict candidate suitability or organisational "fit" [15]. Official narratives often frame these tools as efficiency-enhancing and bias-reducing. However, the reviewed materials reveal a different institutional dynamic.

Across documented cases, applicants frequently receive rejection outcomes without explanation, meaningful feedback, or clear routes for appeal. Regulatory complaints and investigative reporting show that candidates are often unable to determine whether they were rejected due to skills mismatch, data error, automated filtering thresholds, or other model-based criteria. This creates a recurring pattern in which exclusion is experienced as procedurally final but informationally empty.

Evidence from algorithmic hiring research suggests that predictive recruitment tools commonly reproduce historical hiring preferences when trained on past organisational data [15]. Even when explicit demographic indicators are removed, proxy variables such as educational background, employment gaps, geographic location, or linguistic markers can indirectly reproduce existing patterns of stratification [16]. As a result, exclusion may occur without any visible discriminatory intent, yet still reflect embedded institutional histories.

The composite vignette synthesises multiple publicly reported cases in which job seekers describe receiving rapid automated rejections without any explanation or human contact. In these narratives, applicants do not merely experience rejection as a routine administrative outcome. Instead, they attempt to interpret the silence itself, speculating whether they were misclassified, deemed unsuitable, or simply filtered out by an unseen process. What emerges from these accounts is not only a concern about fairness, but a broader uncertainty about whether their application was meaningfully evaluated at all.

This pattern highlights a sociological dimension that is not fully captured in dominant hiring-bias debates. Much of the existing literature focuses on whether recruitment algorithms produce accurate or fair predictions. The reviewed materials suggest an additional

institutional effect: automated hiring systems can transform exclusion into an opaque and depersonalised process that weakens applicants' sense of procedural recognition. The absence of explanation does not only limit transparency; it also reshapes how individuals interpret their standing within labour market institutions.

The key finding in this domain is therefore not limited to bias replication. Rather, automated recruitment appears to normalise a mode of institutional silence in which exclusion is delivered without relational engagement or interpretive grounding. This shifts how rejection is understood, from a contestable decision to an automated verdict whose rationale remains inaccessible.

4.2. Eligibility Automation and the Administrative Production of Misrecognition

Public audits, welfare oversight reports, and policy evaluations indicate that automated eligibility systems are increasingly used to assess benefit entitlement, detect risk, and prioritise cases for review [8]. These systems typically rely on administrative records, behavioural indicators, geospatial markers, and predictive scoring to classify claimants according to perceived risk, compliance likelihood, or eligibility status. While institutions often justify automation as a way to improve consistency and reduce administrative burden, documented cases reveal recurring concerns about misclassification and opaque decision-making.

Across reviewed materials, claimants frequently report receiving automated denials or benefit adjustments without clear explanations. Ombudsman complaints, investigative journalism, and regulatory inquiries show that individuals often struggle to understand why their circumstances were categorised as high-risk, non-compliant, or ineligible. In many cases, affected individuals are unable to determine whether decisions were based on outdated records, data errors, rigid thresholds, or predictive inferences.

Existing welfare research suggests that automated scoring systems embed normative assumptions about deservingness, responsibility, and risk [8]. These assumptions are not always explicit in policy texts, yet they shape how individuals are categorised and how their claims are evaluated. Rather than assessing people through direct engagement with their lived conditions, automated systems often translate complex social realities into simplified risk indicators or compliance profiles.

The composite vignette draws on recurring public cases in which parents, carers, or low-income claimants describe receiving benefit denials that conflict with their actual circumstances. In these documented patterns, individuals interpret automated outcomes not merely as bureaucratic decisions but as signals that institutions misunderstand or misjudge them. The sense of harm arises less from the financial decision alone and more from the perception that personal realities, vulnerabilities, or caregiving responsibilities have been reduced to abstract data points.

This pattern reveals a form of institutional misrecognition. Rather than feeling evaluated as persons with specific needs, individuals often experience automated welfare decisions as categorical judgements that misrepresent their identity, intentions, or moral standing. Prior research supports this interpretation, showing that automated welfare infrastructures can transform support systems into surveillance-oriented risk regimes [25].

The findings therefore suggest that harm in automated welfare governance is not limited to technical error or bias. It also stems from the way predictive classification replaces relational assessment. When automated systems substitute lived understanding with risk scoring, they reshape how citizens perceive state institutions, shifting welfare encounters from sites of support toward experiences of procedural distance and moral detachment.

4.3. Platform Visibility and the Algorithmic Rewriting of Social Worth

Digital platforms increasingly rely on ranking, recommendation, and moderation systems to determine which content gains visibility and which remains marginal. Platform transparency reports, governance policies, and prior research indicate that engagement metrics, user interaction patterns, posting frequency, and behavioural signals play a central role in shaping algorithmic visibility [18,19]. These systems therefore influence not only content reach but also whose voices, identities, and perspectives become prominent within digital public spaces.

Documented cases from creators, journalists, and civil society groups reveal recurring experiences of unexplained visibility fluctuations. Users frequently report sudden drops in audience reach, engagement, or monetisation without receiving a clear justification from platform operators. Governance policies typically provide broad explanations centred on quality, safety, or community standards, yet these high-level narratives rarely explain how specific ranking or suppression decisions are made in practice.

Research on platform governance suggests that algorithmic visibility is shaped less by neutral assessments of content value and more by optimisation strategies that prioritise retention, engagement velocity, and commercial performance [18]. Duffy and Meisner [19] further show that creators interpret shifts in visibility not simply as technical outcomes but as reflections of personal credibility, relevance, or social worth. In this sense, platform ranking systems operate as informal gatekeepers of recognition and legitimacy.

The composite vignette synthesises publicly documented patterns in which creators experience visibility loss as arbitrary, punitive, or targeted, even when they cannot identify any violation of platform rules. These patterns align with evidence that ranking systems can privilege particular expressive styles, posting rhythms, language patterns, or audience behaviours, thereby producing uneven exposure across different communities and identity groups [10].

Beyond economic consequences, visibility systems shape how individuals understand their social presence. Walter and Farkas [23] show that algorithmic curation influences perceptions of credibility and authority by determining which voices appear prominent or trustworthy within information environments. When visibility becomes algorithmically mediated, creators and users may begin to interpret reach and engagement as indicators of legitimacy, influence, or personal value.

This produces a distinct form of institutional power. Platforms do not merely distribute attention; they indirectly shape reputational standing, perceived relevance, and symbolic status within digital communities. In this context, algorithmic visibility functions as a form of social classification that affects how individuals are seen by others and how they understand themselves.

The findings therefore suggest that algorithmic ranking systems operate not only as technical filters but as social sorting mechanisms that structure recognition, legitimacy, and participation. Harm in this domain does not arise solely from economic disadvantage or content suppression. It also emerges from the way platform infrastructures transform visibility into a proxy for credibility and worth, reshaping the social meaning of presence in digital life.

4.4. The Cross-Domain Logic of Algorithmic Power

Across recruitment, welfare administration, and digital platforms, automated systems operate in different institutional contexts, yet they generate structurally similar effects on how individuals encounter authority, opportunity, and recognition. Although the technical architectures vary, the underlying pattern remains consistent: decisions are produced at

scale, delivered with limited explanation, and experienced as authoritative despite being difficult to interpret or contest.

In employment screening, automated filtering systems remove applicants from consideration without meaningful feedback, leaving individuals unable to determine the basis of exclusion [15,16]. In welfare contexts, automated eligibility or risk scoring systems produce classifications that claim administrative legitimacy but may conflict with lived circumstances [8,25]. On digital platforms, ranking systems shape visibility and reach without transparent justification, leading users to speculate about hidden rules or implicit penalties [18,19].

Across these domains, individuals tend to interpret automated outcomes not only as technical decisions but as institutional signals. Decisions are read as messages about employability, deservingness, credibility, or social relevance. This pattern appears repeatedly in public complaints, regulatory inquiries, platform governance research, and documented user responses. The shared feature is not simply the presence of automation, but the way automation reshapes how institutional judgement is communicated and understood.

Existing research often examines algorithmic systems within isolated sectors, focusing on bias in hiring, error in welfare automation, or moderation in platform governance. However, the cross-domain material analysed here suggests that automated systems produce a broader institutional transformation. Classification becomes faster, more standardised, and less dialogic. Explanation becomes more limited. Opportunities to contest or contextualise decisions become more constrained.

This shift alters how institutional power is exercised. Instead of judgement being communicated through interaction, deliberation, or narrative reasoning, it is increasingly expressed through categorical outputs such as acceptance or rejection, eligibility or denial, and visibility or suppression. Individuals must then interpret these outputs without access to the reasoning that produced them.

The evidence indicates that this transformation affects not only outcomes but also perceptions of legitimacy. When people cannot understand how a decision was made, they often struggle to determine whether it was fair, mistaken, biased, or indifferent. This uncertainty can weaken confidence in institutional processes, even when formal rules or technical safeguards exist.

Importantly, the cross-domain findings do not suggest that automated systems operate identically across sectors. Rather, they show that automation introduces recurring institutional dynamics: reduced transparency, expanded classification, constrained contestability, and greater distance between decision-makers and decision subjects. These dynamics help explain why automated systems can simultaneously improve administrative efficiency while generating new forms of frustration, mistrust, and perceived exclusion.

This section therefore identifies a unifying institutional pattern rather than a single technical failure. Automated systems increasingly function as intermediaries between individuals and institutions, shaping how decisions are produced, communicated, and interpreted. Understanding their impact requires attention not only to technical performance, but also to how automation transforms recognition, accountability, and the experience of institutional authority.

5. Interpretation: Automated Systems as Emerging Social Institutions

The findings across recruitment, welfare administration, and digital platforms point to a shared institutional transformation. Automated systems are no longer functioning merely as decision-support tools. They increasingly perform roles that were historically associated with institutional judgement, including classification, evaluation, prioritisation, and exclu-

sion. Decisions that once relied on interpersonal reasoning and contextual discretion are now produced through rapid, standardised, and opaque computational processes.

This shift matters because it changes not only how decisions are made, but how authority is exercised and experienced. Prior research shows that automation redistributes discretion from frontline officials to technical infrastructures, altering internal power dynamics within organisations [1]. Comparable dynamics appear in welfare administration [8] and platform governance, where ranking systems perform functions once carried out by editors, caseworkers, or gatekeepers [2,3,9]. The present findings extend this literature by showing that automation also transforms how individuals interpret institutional decisions, how legitimacy is perceived, and how inequality is produced and justified.

Across domains, automated outputs are commonly treated as final, even when their reasoning is inaccessible. This creates a situation in which institutional authority is exercised without explanation, dialogue, or visible accountability. The result is not simply procedural change, but a reconfiguration of legitimacy, recognition, and public trust.

5.1. Automated Systems as Institutional Actors

Institutional theory suggests that authority emerges when outputs are routinely accepted as binding within formal decision processes [29]. Automated systems increasingly meet this condition. In recruitment, welfare, and platform environments, algorithmic decisions are acted upon as authoritative even when affected individuals cannot contest or interpret them.

This reflects what has been described as the cultural construction of automated authority, in which algorithmic outputs are framed as more neutral or objective than human judgement [13]. The evidence in this study indicates that automated systems derive their power not from transparency or technical superiority, but from institutional embedding. When organisations integrate algorithmic outputs into routine governance practices, the system effectively becomes the decision-maker.

Individuals tend to interpret automated outcomes as institutional verdicts rather than technical suggestions. Rejections, risk scores, and visibility shifts are experienced as judgments about competence, deservingness, or credibility. In this sense, automated systems do not simply serve institutional goals. They increasingly participate in defining what counts as merit, risk, relevance, or legitimacy. This represents a substantive expansion of algorithmic power beyond technical assistance into institutional governance.

5.2. The Emergence of Datafied Citizenship

Automated classification systems also reshape how individuals understand their relationship with institutions. Prior research suggests that algorithmic governance alters how citizens perceive what the state values and recognises [11]. The findings here indicate that this process extends beyond governance into everyday identity formation.

Across the analysed domains, individuals interpret automated classifications as signals about their social worth. In recruitment, exclusion becomes a marker of institutional non-recognition. In welfare contexts, risk scoring is read as a moral assessment of deservingness. On platforms, visibility becomes a proxy for credibility and social relevance. These interpretations emerge not from direct interaction with decision-makers, but from how algorithmic systems categorise and display individuals.

This aligns with arguments that inequality in automated systems operates through data-driven stratification rather than explicit discrimination [16]. Individuals increasingly encounter institutions through predictive profiles and behavioural inferences rather than interpersonal engagement. The concept of datafied citizenship captures this transformation:

people come to understand their institutional standing through algorithmic categories that shape access to opportunities, recognition, and participation.

5.3. Trust as a Sociotechnical Relationship

Trust in institutional decision-making has traditionally depended on both procedural fairness and perceived recognition. Experimental studies indicate that people may initially accept automated decisions as credible even without understanding them [7]. However, evidence from welfare systems and digital platforms shows that this credibility weakens when automated outcomes conflict with lived experience or appear indifferent to personal context [2,3,18].

The findings suggest that trust in automated systems is not determined solely by technical accuracy or formal transparency. Instead, trust depends on whether individuals feel seen, understood, and acknowledged by the institution behind the system. Where automated processes eliminate relational cues, opportunities for explanation, or meaningful avenues for contestation, individuals interpret decisions as impersonal, arbitrary, or unjust.

This supports the view that legitimacy in automated governance depends on engagement rather than technical disclosure alone [24]. In the cases examined, distrust arises not simply because systems are opaque, but because they replace interpersonal accountability with impersonal classification. Trust therefore emerges as a sociotechnical relationship shaped by recognition, responsiveness, and institutional care, not merely by computational performance.

5.4. A Dual Logic of Inequality in Automated Societies

Debates on algorithmic inequality often focus on whether automated systems reproduce historical bias. While this reproduction mechanism is well established, the findings indicate that automated inequality operates through two interacting pathways: amplification of past disadvantage and production of new forms of exclusion.

Amplification occurs when predictive systems embed historical inequalities into ongoing decision processes. In employment screening, hiring models have been shown to replicate gendered and racialised patterns even when protected characteristics are excluded [15]. Proxy variables can reintroduce social hierarchies into ostensibly neutral systems [16,20]. In these cases, automation stabilises long-standing inequalities by encoding historical patterns into present classifications.

However, automated systems also generate novel forms of inequality that do not derive directly from past social structures. Welfare automation illustrates this process by creating new administrative categories such as predicted risk or anticipated non-compliance that have no equivalent in legal or moral discourse [8]. These categories reshape how individuals are evaluated, often without clear justification [18]. Inequality here is produced through data abstraction rather than historical continuity.

Platform governance reinforces this pattern. Visibility and relevance are shaped by optimisation metrics that privilege particular behavioural rhythms, aesthetic styles, or engagement patterns [2,3]. These systems create new hierarchies of credibility and reach that did not exist prior to algorithmic mediation [9]. Inequality in this context is not merely inherited; it is actively constructed through platform architecture.

As Table 2 demonstrates, this dual logic appears across multiple institutional domains, including employment, welfare, finance, healthcare, education, public safety, digital platforms, urban administration, and cross-domain identity profiling. In each case, inequality emerges both through reinforcement of existing disadvantage and through the creation of new classificatory identities that detach individuals from familiar legal, social, or moral categories.

Table 2. Cross-Domain Matrix of Automated Inequality Across Sociotechnical Arenas, Showing Amplification and Novel Inequality Mechanisms. Source: Author's synthesis and cross-domain analytical mapping based on the reviewed literature and documented institutional cases.

Domain	Institutional Function of Automation	Core Data & Infrastructures	Mechanism of Amplification (Historical Reproduction)	Mechanism of Novel Inequality (Algorithmic Production)	Consequences for Recognition, Opportunity, and Legitimacy	Key Sources
Employment Screening	Predict applicant suitability and streamline hiring	Historical hiring records, CV parsers, behavioural scoring tools	Past gendered and racial hiring patterns reproduced as signals of "fit"	New behavioural and personality risk categories generated without grounding in labour law or workplace norms	Applicants experience rejection as opaque, depersonalised, and institutionally final	Fabris et al. [15]; Zajko [16]; Noble [20]
Welfare & Social Benefits	Allocate resources, detect fraud, manage eligibility	Claims histories, payment records, risk-flag databases	Longstanding associations between poverty, non-compliance, and risk encoded into scoring models	Emergence of algorithmic identities such as "high-risk claimant" or "likely violator," absent from statutory categories	Claimants experience misrecognition, moral judgement, and loss of institutional trust	Larasati et al. [8]; Sleep & Reddington [18]; Hillo & Vento [25]
Platform Visibility & Cultural Production	Rank content, allocate attention, maximise engagement	Engagement metrics, follower behaviour, interaction velocity	Visibility advantages accrue to groups historically dominant in attention economies	New expressive norms privileging specific styles, timing patterns, and linguistic formats	Creators interpret visibility shifts as personal devaluation or targeted suppression	Kojah et al. [2]; Duffy & Meisner [3]; Walter & Farkas [9]
Education Technologies	Predict learning outcomes and personalise instruction	Test scores, attendance logs, platform interaction data	Students from historically under-resourced schools flagged as low-ability	New categories such as "learning risk" or "attention deficit" created without pedagogical grounding	Students internalise automated labels as fixed judgements of intellectual worth	Zajko [16]

Table 2. *Cont.*

Domain	Institutional Function of Automation	Core Data & Infrastructures	Mechanism of Amplification (Historical Reproduction)	Mechanism of Novel Inequality (Algorithmic Production)	Consequences for Recognition, Opportunity, and Legitimacy	Key Sources
Predictive Policing & Public Safety	Forecast crime risk and allocate enforcement resources	Crime statistics, geographic data, patrol records	Over-policing of historically marginalised communities reinforced through predictive targeting	Individuals assigned algorithmic suspicion profiles not recognised in criminal law	Community trust erodes as risk scoring is perceived as arbitrary and discriminatory	Noble [20]; O'Neil [22]
Credit & Financial Access	Estimate repayment probability and automate lending decisions	Credit histories, transactional data, online behaviour traces	Historically excluded groups continue to receive lower creditworthiness scores	Creation of opaque financial personas determining access beyond actual financial behaviour	Individuals experience financial exclusion without meaningful procedural explanation	Noble [20]; Grimmelikhuijsen [13]
Healthcare Triage & Benefit Prioritisation	Allocate care resources and predict clinical risk	Electronic health records, insurance claims, diagnostic histories	Groups historically under-diagnosed assigned lower medical priority	New algorithmic risk labels detached from clinically meaningful categories	Patients feel algorithmically devalued or unseen by medical institutions	Atoum et al. [5]; Hillo et al. [30]
Urban Governance & Public Administration	Automate service allocation and administrative triage	Municipal records, complaints systems, sensor data	Historical disparities in neighbourhood investment encoded into predictive service allocation	New bureaucratic identities such as "undeserving service user" or "low-priority citizen"	Citizens perceive decisions as procedurally closed and politically unresponsive	Goldsmith & Yang [1]; Yuan & Chen [6]
Cross-Domain Identity Construction	Link profiles across employment, welfare, finance, and platforms	Integrated identity datasets, behavioural traces, profiling tools	Disadvantage follows individuals across institutional domains through linked records	Entirely new algorithmic identities such as "persistent risk subject" or "low-value user" produced	Individuals lose control over how they are known, evaluated, and categorised across systems	Parviainen et al. [11]; Taylor & Calvert [28]; Hillo et al. [30]

This dual mechanism explains why automated inequality can feel simultaneously familiar and unfamiliar. Amplification explains continuity with historical injustice. Novel production explains the emergence of new exclusionary categories that are harder to recognise, contest, or regulate. Together, they transform not only how inequality is distributed, but how it is experienced, interpreted, and legitimised.

5.5. Contributions to Sociological Understanding of Algorithmic Power

This section advances three central contributions.

First, it reframes algorithmic inequality by showing that automated systems do not merely replicate biased data but actively construct new administrative and social categories that shape opportunity and recognition.

Second, it demonstrates that inequality under automation operates across institutional domains rather than within isolated sectors. Individuals may encounter aligned classificatory logics in employment, welfare, finance, and digital life, even when these systems are formally independent.

Third, it provides a conceptual foundation for governance reform by clarifying that fairness interventions must address both inherited bias and newly produced forms of exclusion. Addressing only one pathway leaves significant dimensions of automated harm intact.

These insights ground the transition to the next section, which examines how intelligibility, accountability, inclusion, and public trust can be strengthened in institutional environments increasingly shaped by automated judgement.

6. Governance and Social Responsibility: Institutional Reform in an Age of Automated Authority

The evidence presented in this article shows that automated systems have moved beyond their role as technical aids. They now exercise forms of authority that shape opportunity, recognition, and institutional legitimacy across multiple social domains. Automated judgement increasingly performs work once carried out by human bureaucracies, including evaluation, categorisation, prioritisation, and exclusion. Governance frameworks must therefore respond to automated systems not as neutral tools, but as institutional actors with real social consequences. Existing regulatory approaches remain fragmented and largely focused on technical compliance, leaving the relational, ethical, and legitimacy dimensions of automation insufficiently addressed. What is required is a governance orientation that restores intelligibility, strengthens accountability, expands inclusion, and rebuilds public trust.

6.1. Intelligibility as a Condition of Institutional Legitimacy

Across the material analysed, a recurring problem is that individuals encounter decisions they cannot meaningfully interpret. Trust breaks down not only because explanations are missing, but because people cannot determine whether their circumstances were understood or considered. Fang et al. [7] show that perceptions of legitimacy improve when individuals believe a system has engaged with their specific case, even when full technical transparency is absent. By contrast, Sleep and Reddington [18] demonstrate that unexplained welfare decisions are often experienced as moral disregard rather than administrative error.

These findings indicate that intelligibility should be treated as a civic requirement rather than a technical add-on. Legitimate automated systems must provide forms of explanation that people can understand in social and practical terms, not only in computational language. This requires more than publishing model documentation or technical

summaries. Institutions must develop communication practices that translate automated reasoning into explanations that align with how people interpret fairness, recognition, and responsibility. Reforms should therefore prioritise enforceable explanatory rights, user-centred interpretability tools, and communication standards designed around human meaning rather than technical disclosure.

6.2. Accountability in Distributed Decision-Making Systems

Automation complicates accountability because responsibility is no longer located in a single identifiable decision-maker. Instead, judgement becomes dispersed across organisational policies, data pipelines, model architectures, and institutional routines. Skeja and Sadiku Dushi [4] show that organisations often treat automation as a neutral delegation of responsibility, creating ethical blind spots. Herrera Poyatos et al. [14] further demonstrate that design choices and modelling assumptions acquire institutional authority long before systems are deployed or scrutinised.

An effective accountability framework must therefore treat automated systems as institutional processes rather than isolated technical products. Accountability depends on whether organisations can justify how and why automated classifications were produced, and whether affected individuals can contest those classifications in meaningful ways. This requires independent audit bodies with the authority to examine systems as governance mechanisms, public oversight structures that allow external scrutiny of high-impact deployments, and internal institutional processes that document data selection, model design decisions, and optimisation objectives. These measures reconnect automated outcomes to institutional responsibility rather than allowing accountability to dissolve into technical abstraction.

6.3. Inclusion Beyond Statistical Fairness

The findings in Section 5 demonstrate that automated inequality arises both from historical bias and from newly created algorithmic categories that detach individuals from established social and legal identities. Noble, Zajko, and Larasati et al. show that marginalised populations are disproportionately affected because their experiences are underrepresented, misinterpreted, or structurally disadvantaged within training data and optimisation systems [8,16,20]. Addressing these harms requires more than improving demographic parity metrics or refining fairness constraints.

Inclusion in automated governance must be institutional and participatory, not merely statistical. This means involving affected communities in system design, evaluation, and oversight, especially in high-impact contexts such as welfare, employment, credit, and policing. Mechanisms such as citizen advisory panels, public interest impact assessments, and structured consultation processes can help ensure that automated systems reflect lived social realities rather than abstract technical assumptions. These practices shift inclusion from numerical representation toward institutional recognition, reinforcing the legitimacy of automated decision-making systems in diverse social contexts.

6.4. Rebuilding Public Trust Through Recognition and Responsiveness

Public trust depends not only on whether decisions are accurate, but on whether institutions appear attentive, respectful, and willing to justify their actions. Parviainen et al. [11] show that automated systems weaken trust by removing the interpersonal cues that once signalled institutional care. Goldsmith and Yang [1] similarly argue that trust in automated governance cannot be restored through technical transparency alone, because the deeper challenge is relational distance between institutions and citizens.

Rebuilding trust therefore requires embedding automated systems within administrative processes that preserve opportunities for dialogue, explanation, and contestation.

Appeal mechanisms must allow individuals to challenge automated classifications. Public communication must openly acknowledge uncertainty, limitations, and the potential for error in algorithmic judgement. Institutions must demonstrate a willingness to revise or override automated outputs when they conflict with evidence, fairness, or lived experience.

Trust is not restored through better optimisation alone. It is restored when people feel seen, heard, and taken seriously by the institutions that govern them. When automated systems operate within frameworks that prioritise recognition, responsiveness, and accountability, they can become tools of public value rather than sources of alienation or distrust.

7. Conclusions: Artificial Intelligence and the Reshaping of Social Order

This article has shown that automated systems can no longer be understood as neutral technical tools. They increasingly function as institutional actors that classify people, shape access to opportunity, and influence how legitimacy and trust are constructed in everyday life. What appears to be routine automation in hiring, welfare, and digital platforms is in fact a deeper transformation in how institutions recognise individuals and exercise authority.

The framework developed in this study identifies three mechanisms through which automated systems gain institutional power. Delegated authority turns algorithmic outputs into binding judgements. Classification structures who becomes visible, credible, or marginal within institutional processes. Procedural opacity disrupts the relational foundations through which fairness and accountability have traditionally been assessed. Together, these mechanisms allow automated systems not only to support decision-making, but to redefine the social conditions under which decisions are experienced as legitimate.

The empirical analysis further demonstrates that people encounter automated decisions as both authoritative and difficult to contest. Across employment, welfare, and digital environments, individuals face outcomes that carry institutional weight while offering limited explanation or opportunity for challenge. Their responses range from confusion and resignation to adaptation and withdrawal. These reactions provide insight into how algorithmic power is lived, interpreted, and negotiated, revealing that automation reshapes not only outcomes, but also how people understand recognition, fairness, and institutional belonging.

A central contribution of this article is its account of inequality under automation. Automated systems both reinforce existing social hierarchies and generate new forms of exclusion through data-driven categorisation and optimisation. This dual dynamic explains why algorithmic inequality can feel simultaneously familiar and novel. Individuals remain tied to inherited disadvantage while also being sorted into new classifications that do not align with established legal, moral, or social categories. Inequality under automation is therefore not only reproduced, but transformed in structure and expression.

Taken together, these findings point to a broader reconfiguration of social authority. Automated systems are changing how institutions know people, how decisions are justified, and how trust is sustained. If societies are to integrate these technologies responsibly, governance approaches must address more than technical accuracy or efficiency. Intelligibility, accountability, inclusion, and recognition must be treated as foundational conditions for legitimate institutional power.

Future research should examine the long-term effects of automated judgement on institutional practices, explore how individuals reshape their identities in response to persistent algorithmic classification, and compare governance models across national and regulatory contexts. There is also a growing need to connect studies of administrative automation with broader analyses of platform power and political economy.

https://doi.org/10.3390/soc16020059

This article concludes with a straightforward claim. Artificial intelligence is no longer external to social institutions. It is becoming part of the infrastructure through which social order is organised and authority is exercised. Understanding its impact requires attention not only to technical design, but to the social relationships through which fairness, legitimacy, and belonging are constructed. The challenge ahead is to ensure that automated systems strengthen, rather than erode, the conditions for a just and recognitional social world.

Author Contributions: Conceptualization, O.B.A.; Methodology, O.B.A. and O.T.O.; Investigation, O.B.A.; Resources, O.B.A., I.M.-K. and O.M.O.; Data curation, O.B.A., I.M.-K. and O.M.O.; Writing—original draft, O.B.A., I.M.-K., O.T.O. and O.M.O.; Writing—review and editing, O.B.A., I.M.-K., O.T.O. and O.M.O.; Visualisation, O.B.A.; Supervision, O.B.A. and O.T.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Goldsmith, S.; Yang, J. AI and the Transformation of Accountability and Discretion in Urban Governance. *arXiv* **2025**, arXiv:2502.13101. [CrossRef]
2. Kojah, S.; Zhang, B.; Are, C.; Delmonaco, D.; Haimson, O. Dialing it Back: Shadowbanning, Invisible Digital Labour, and the Mitigation of Opaque Platform Governance. *Proc. ACM Hum. Comput. Interact.* **2025**, *9*, 1–22.
3. Duffy, B.; Meisner, C. Platform Governance at the Margins: Social Media Creators' Experiences with Algorithmic Visibility. *Media Cult. Soc.* **2023**, *45*, 285–304. [CrossRef]
4. Skeja, A.; Sadiku Dushi, N. Toward Sustainable AI Leadership: Ethical Blind Spots, Accountability Gaps and the CARE Governance Framework. *Leadersh. Organ. Dev. J.* **2025**, *ahead-of-print*. [CrossRef]
5. Atoum, I. Holistic AI Governance: Addressing Bias and Ensuring Accountability. *Int. J. Adv. Comput. Sci.* **2025**, *16*, 830–839.
6. Yuan, Q.; Chen, T. Holding AI Based Systems Accountable in the Public Sector: A Systematic Review. *Public Perform. Manag. Rev.* **2025**, *48*, 1389–1422. [CrossRef]
7. Fang, X.; Zhou, H.; Chen, S. Explaining Why the Computer Says No: Algorithmic Transparency and Perceived Trustworthiness. *Public Adm.* **2025**, *83*, 241–262.
8. Larasati, Z.; Yuda, T.; Sunaryo, A. Digital Welfare State and Emerging Problems: An Exploration and Research Agenda. *Int. J. Sociol. Soc. Policy* **2023**, *43*, 537–549. [CrossRef]
9. Walter, C.; Farkas, J. Behind the Screens: How Algorithmic Imaginaries Shape Science Content. *J. Sci. Commun.* **2025**, *24*, A02. [CrossRef]
10. Lopezmalo, A. *Digital Suppression: How Algorithmic Bias Silences Marginalised Communities*; Loyola University Research Reports; Center for Engaged Learning, Teaching, & Scholarship: Chicago, IL, USA, 2025.
11. Parviainen, J.; Koski, A.; Eilola, L. Building and Eroding the Citizen State Relationship in the Era of Algorithmic Decision Making. *Soc. Sci.* **2025**, *14*, 178. [CrossRef]
12. Hastuti, H.; Maulana, H.; Lawelai, H. Algorithmic Influence and Media Legitimacy: Systematic Review of Social Media's Impact on News Production. *Front. Commun.* **2025**, *10*, 1667471. [CrossRef]
13. Grimmelikhuijsen, S. Constructing Automated Societies: Sociocultural Determinants of Automated Decision Making in Public Services. *Public Adm. Rev.* **2025**.
14. Herrera Poyatos, A.; Del Ser, J.; de Prado, M.L.; Wang, F.-Y.; Herrera-Viedma, E.; Herrera, F. Responsible Artificial Intelligence Systems: Auditability, Accountability and Governance. *arXiv* **2025**, arXiv:2503.04739.
15. Fabris, A.; Baranowska, N.; Dennis, M.J.; Graus, D.; Hacker, P.; Saldivar, J.; Borgesius, F.Z.; Biega, A.J. Fairness and Bias in Algorithmic Hiring: A Multidisciplinary Survey. *ACM Trans. Intell. Syst. Technol.* **2025**, *16*, 1–54. [CrossRef]

16. Zajko, M. Artificial Intelligence, Algorithms and Social Inequality: Sociological Contributions. *Sociol. Compass* **2022**, *16*, e12962. [CrossRef]
17. Bennani-Taylor, S. Infrastructuring AI: The stabilization of 'artificial intelligence' in and beyond national AI strategies. *First Monday* **2024**, *29*. [CrossRef]
18. Sleep, L.; Redden, J. Reimagining Failed Automation: From Neoliberal Punitive Automated Welfare Towards a Politics of Care. CQUniversity. Chapter. 2024. Available online: https://hdl.handle.net/10779/cqu.28433015.v1 (accessed on 6 January 2026).
19. Fayezi, S.; Klassen, R. Reducing Modern Slavery Risks Through Sociotechnical Systems. *Prod. Oper. Manag.* **2025**, *34*, 1028–1037. [CrossRef]
20. Noble, S.U. *Algorithms of Oppression*; NYU Press: New York, NY, USA, 2018.
21. Fine, A.; Berthelot, E.; Molinaro, S. Public Perceptions of Judges Use of AI Tools in Courtroom Decision Making. *Behav. Sci.* **2025**, *15*, 476. [CrossRef]
22. O'Neil, C. *Weapons of Math Destruction*; Penguin: London, UK, 2016.
23. Mathe, L. *The Algorithmic Power and Language Policy in African Digital News*; Springer: Berlin/Heidelberg, Germany, 2025.
24. Decker, M.; Wegner, L.; Larson, C. Procedural Fairness in Algorithmic Decision Making: Public Engagement and Legitimacy. *Ethics Inf. Technol.* **2025**, *27*, 1–16. [CrossRef]
25. Van Toorn, G.; Henman, P. Introduction to the Digital Welfare State. *J. Sociol.* **2024**, *60*, 507–522. [CrossRef]
26. Wong, P.; Singh, A.; Brumby, D. Climate Community for People of Colour on Digital Platforms. *Proc. ACM Hum. Comput. Interact.* **2025**, *9*, CSCW088. [CrossRef]
27. Taylor, S.H.; Choi, M. An Initial Conceptualization of Algorithm Responsiveness: Comparing Perceptions of Algorithms Across Social Media Platforms. *Soc. Media + Soc.* **2022**, *8*, 20563051221144322. [CrossRef]
28. Rudko, I.; Bashirpour Bonab, A.; Fedele, M.; Formisano, A. New Institutional Theory and AI in Organisations. *J. Manag. Hist.* **2025**, *31*, 261–284.
29. Qian, Y.; Siau, K.L.; Nah, F.F. Societal impacts of artificial intelligence: Ethical, legal, and governance issues. *Soc. Impacts* **2024**, *3*, 100040. [CrossRef]
30. Hillo, J.; Vento, I.; Erkkilä, T. Algorithmic Governance: Experimental Evidence on Citizens' and Public Administrators' Legitimacy Perceptions of Automated Decision-Making. *Public Adm.* **2025**, 1–16. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.