social sciences

MDPI

Review

Building and Eroding the Citizen–State Relationship in the Era of Algorithmic Decision-Making: Towards a New Conceptual Model of Institutional Trust

Jaana Parviainen [1,*], Anne Koski [1], Laura Eilola [2], Hannele Palukka [1], Paula Alanen [3] and Camilla Lindholm [4]

[1] Faculty of Social Sciences, Tampere University, FI-33520 Tampere, Finland; anne.koski@tuni.fi (A.K.); hannele.palukka@tuni.fi (H.P.)
[2] Faculty of Information Technology and Communication Sciences, Tampere University, FI-33520 Tampere, Finland; laura.eilola@tuni.fi
[3] Faculty of Education and Culture, Tampere University, FI-33520 Tampere, Finland; paula.alanen@tuni.fi
[4] Department of Finnish, Finno-Ugrian and Scandinavian Studies, University of Helsinki, 4, FI-00100 Helsinki, Finland; camilla.lindholm@helsinki.fi
* Correspondence: jaana.parviainen@tuni.fi

Abstract: In liberal welfare states, algorithmic decision-making systems are being increasingly deployed, impacting the citizen–state relationship in a multitude of positive and negative ways. This theoretical paper aims to develop a novel conceptual model—the institutional trust model—to analyse how the implementation of automated systems erodes or strengthens institutional trust between policymakers and citizens. In this approach, institutional trust does not simply mean public trust in institutions (though it is an important component of democratic societies); instead, it refers to the responsive interactions between governmental institutions and citizens. Currently, very little is known about policymakers' trust or distrust in automated systems and how their trust or distrust in citizens is reflected in their interest in implementing these systems in public administration. By analysing a sample of recent studies on automated decision-making, we explored the potential of the institutional trust model to identify how the four dimensions of trust can be used to explore the responsive relationship between citizens and the state. This article contributes to the formulation of research questions on automated decision-making in the future, underlining that the impact of automated systems on the socio-economic rights of marginalised citizens in public services and the policymakers' motivations to deploy automated systems have been overlooked.

Keywords: institutional trust; automated decision-making; citizen–state relationship; public administration; welfare service

Academic Editor: Ina Kayser

Received: 14 August 2024
Revised: 22 February 2025
Accepted: 7 March 2025
Published: 17 March 2025

Citation: Parviainen, Jaana, Anne Koski, Laura Eilola, Hannele Palukka, Paula Alanen, and Camilla Lindholm. 2025. Building and Eroding the Citizen–State Relationship in the Era of Algorithmic Decision-Making: Towards a New Conceptual Model of Institutional Trust. *Social Sciences* 14: 178. https://doi.org/10.3390/socsci14030178

1. Introduction

A quiet revolution is unfolding in welfare states: public administrations worldwide are increasingly using algorithmic decision-making to streamline their administrative processes. Automation has transformed street-level administrative work into self-service transactions, with the promise of delivering services more efficiently along with more transparent, value-neutral, unbiased and fair decision-making (Casey 2022; Choi 2016; Cordella and Tempini 2015; Wirtz et al. 2019; Yildiz 2007). For instance, Estonia, which has automated many of its bureaucratic processes, is estimated to save 2% of its GDP each year (Coglianese 2023). Moreover, automation benefits citizens in many ways; for instance,

it facilitates their access to social benefits and helps them file taxes and complete other administrative tasks online, making visiting offices and standing in queue for their turn unnecessary. According to many optimistic visions, societal problems can be solved using innovative algorithms and big data (e.g., Guszcza et al. 2014). However, according to critical scholars, this new algorithmic governance indicates a fundamental public sector reform (Eubanks 2018; C. O'Neill 2016). With algorithmic decision-making, citizens experience the shift to a new phase of neoliberal austerity politics and the increasing hollowness of welfare states, where traditional bureaucracy is amalgamated with citizens' administrative burden and new surveillance systems.

The European Union has been concerned about the impact of automated decision-making systems on institutional trust and how citizens' potential distrust of the fairness and equality of public administration influences the legitimacy of liberal democracies (European Committee for Democracy and Governance 2021; European Parliament 2022). With automated systems becoming more common in the public sector, some citizens have already endured the unfair consequences complex algorithms can have on their lives (see e.g., Hadwick and Lan 2021). In Australia, Canada, Europe, New Zealand and the United States, there have been over 60 instances of cancelling or pausing automated decision-making projects due to their failure to recognise the differences, complexities and rights of individuals and communities (Redden et al. 2022). In the era of a post-democratic dynamic (Behnam Shad 2023; Crouch 2020), feelings of distrust, including frustration with and disillusionment towards the welfare state, are prevalent among Western populations. In this context, scandals related to algorithmic decision-making systems can generate deep distrust towards not only the systems themselves but also the public administration and the institutions of welfare states.

In this paper, we mainly aimed to posit a novel conceptual model of institutional trust to analyse the potential changes and tensions automated decision-making systems cause in the citizen–state relationship. Moreover, based on the literature on institutional trust and past research on automated systems in public administration, we explored how trust and distrust in automated decision-making are related to institutional trust. As the relevance of institutional trust has increased in the social sciences, the concept itself has become a matter of substantial debate (Győrffy 2013). It is commonly conceptualised as a citizen's subjective response—attitude—towards the performance of institutional actors (the government, agencies or other collective actors) (Győrffy 2013; Hetherington 2004; O. O'Neill 2002). In our approach, institutional trust does not simply refer to having trust in institutions (though it is an important component of democratic societies); it also refers to the responsive interactions between governmental institutions and citizens. Institutional trust is understood as a two-way relationship, one in which citizens voluntarily commit to being under the governance of public authorities and authorities are obligated towards citizens—the legitimacy of policymakers depends on the citizens' belief that they wield power in 'restrained and appropriate ways' (Jackson and Gau 2015, p. 60).

Automated decision-making is defined as 'a process where decisions are made by automated means and without human involvement' (European Commission 2023). Automated systems have been intertwined with the transformation of welfare states since the early 1970s, spanning periods of large-scale economic crises and fast technological development (Agar 2003). In terms of algorithmic decision-making in public administration, at least three levels of automation are generally identified based on the three-level scale of human involvement in technology (Anderson and Fort 2022; Peeters 2020). This scale is used to assess human involvement in automated vehicles and weapons: (1) 'humans in the loop'—human oversight of target selection is possible and human follow-up for intervention mechanism is required; (2) 'humans on the loop'—human oversight of target

selection and the default intervention mechanism is possible, but human override of target selection and default intervention mechanism is also possible; and (3) 'humans out of the loop'—human oversight of target or human intervention at the target is impossible or unlikely. In the strict sense, only 'the humans out of the loop' level can be considered automated decision-making. According to Peeters (2020, p. 518), algorithmic decision-making in public administration has plenty of variety in its algorithmic design regarding human involvement: human agents are rarely fully 'out of the loop' in administrative decisions. Still, major control problems are included in the 'humans in/on the loops' of algorithmic decisions. For instance, algorithms can handle the behavioural mechanisms of decision-making and, thereby, limit humans' ability to control the algorithms' processes and outcomes.

The article is organised as follows. We first explore the existing theories on trust to formulate our notion of institutional trust. Then, we discuss trust in the context of algorithmic systems to propose our conceptual model. We expected trust in automation to be essentially related to how people consider the transparency, explainability, fairness and accountability of the decisions made by algorithms. This is followed by a discussion of the recent studies that focused on the interdependence of trust and automated decision-making. Finally, we summarise our key findings with their potential impact on the research field and the significance of this study for future research.

2. Institutional Trust

The erosion of trust in government among individuals and between population groups has been widely documented in liberal countries since the 1990s (Győrffy 2013; Warren 1999). Although this trend is considered bad for democracy, trust has proven to be an ambiguous scientific concept to study. There exists a wide consensus to distinguish between two types of trust: social (intersubjective/generalised) trust and institutional (public/political) trust (Győrffy 2013; Ervasti et al. 2019). Social trust often refers to horizontal trust, while institutional trust refers to vertical trust in institutions (Hardin 1999; Warren 1999). The two types are interrelated (Spadaro et al. 2020), but it remains unclear how strong the link between them is and what their causal relationship looks like (Johansson et al. 2021).

Societal coherence is necessary for a society to function well, and it is maintained by institutional trust (Offe 1999; Rothstein 2011). Institutions are obligated to maintain trust by reinforcing the values of truth-telling, promise-keeping, justice, and solidarity (Offe 1999). Trust is maintained not by a single institution but the collaborative effect of a set of institutions and their administration. Truth-telling and information provision are ensured, for example, by public court proceedings and accounting standards. Promises are enforced by contracts and independent courts. Equality before the law and equal political rights foster a sense of fairness. Moreover, state redistribution to disadvantaged groups of citizens creates a sense of solidarity (Győrffy 2013).

Competing ideas exist in the literature regarding what drives institutional trust. For instance, research has found a causal link between economic growth and institutional trust (Zak and Knack 2001). Welfare service experiences are known to shape citizens' trust in public institutions and their support of the welfare state (Berg and Johansson 2020; Yang and Holzer 2006). However, considering trust as a morally good attitude and distrust as a bad one oversimplifies the discussion on the significance of trust in society. Hardin (1999) claims that doubt and distrust are proper attitudes towards a government that performs poorly. Inglehart and Wenzel (2005) also consider the decline of confidence in public institutions a positive development, since it indicates a critical attitude towards authorities and hierarchies.

Institutional trust has been found to be sensitive to many contextual changes—not only economic crises and elections but also major social and technological reforms. Those who are financially insecure, those with low levels of education, women, young adults and those who belong to a group that is discriminated against—all of them have reported lower levels of trust in government (OECD 2024). When analysing the lack of institutional trust, it is necessary to address structural issues, such as the growing financial insecurity and socioeconomic inequalities. Moreover, social exclusion and marginalisation, low prestige and esteem, and limited mobility across hierarchically arranged social rungs are persistently associated with a sense of existential threat (Adam-Troian et al. 2023; Standing 2011).

In our conceptualisation of institutional trust, we underline the influence of the above-mentioned structural issues. A welfare state and its automated systems depend on citizens' trust and support. It is essential for an automated welfare state to achieve legitimacy, since it organises the relationships between citizens and the state. The need to create and maintain trust in automated systems thus lies at the heart of the debate on the organisation of the welfare state and the efficiency of its services. However, the literature on trust has failed to address how the increasing use of automated decision-making affects how citizens translate their experiences of welfare services into trust in governments and, thus, the ability of governments to build trust in public institutions through welfare services. Inspired by Giddens' argumentation, we suggest that institutional trust is not the same as faith in the reliability of a technical system—instead, "it is what derives from that faith" (Giddens 1991, p. 33).

3. Institutional (Dis)trust in and Through Algorithmic Systems?

In the early state of e-governance, many studies indicated that citizens, especially those with low incomes, were reluctant to use online services because they did not trust public administrations (Pérez-Morote et al. 2020; Reddick 2005). Consequently, the question of how to generate public trust in the administrative use of algorithms has been the focus of several research initiatives (e.g., Lee 2018; Meijer and Grimmelikhuijsen 2020; Robinson 2020). A growing body of literature shows that inaccessible and inexplainable algorithms can erode trust, and this has led to algorithmic transparency becoming key for creating trustworthy algorithms. (Miller 2019; Rudin 2019). Trustworthy algorithms are crucial for vulnerable citizens—including individuals with disabilities, mental health patients, the unemployed and immigrants—especially since they are frequently more dependent on welfare services and benefits and thus at the mercy of algorithmic systems.

Currently, very little is known about policymakers' trust or distrust in automated systems and how their trust or distrust in citizens is reflected in their interest in deploying automated systems in public administration. Policymakers may excessively trust automation while distrusting citizens. The suspicion of welfare fraud could be one of the motivators of government initiatives to develop surveillance systems for social benefits. An emerging body of scholarship has begun to interrogate how the state rehearses, amplifies and circulates discourses that exaggerate the scale and extent of welfare fraud (Gaffney and Millar 2020; Power et al. 2022). The implementation of automated decision-making systems as surveillance systems and their subsequent severe problems, especially for vulnerable citizens, have caused numerous scandals in recent years, such as the 'Toeslagenaffaire' in the Netherlands (Hadwick and Lan 2021). The hidden agendas in governments' digitalisation strategies, according to which some groups are considered 'lazy' and inclined to cheat the system to gain benefits, may reflect the governments' distrust of their citizens. Some researchers and experts have been concerned about the kinds of challenges that algorithmic systems and speeding up public sector automation—with or without machine learning

algorithms—can cause regarding the equal treatment of citizens and the transparency of administrations (Carney 2019; Redden et al. 2022).

The eagerness of policymakers to implement automation in public services—despite many failures and citizens' suspicion of automation—indicates that governments and public authorities trust algorithms' ability to efficiently and equally provide services. Regarding complex systems such as automated decision-making, authorities may benefit from the 'dark side' of institutional trust: confronted with incomprehensible administrative procedures, citizens feel insecure and psychologically compensate for their lack of power and knowledge with excessive and groundless trust in authorities (Shockley and Shepherd 2016). Since legitimacy is relational and won in the interplay between citizens and public administrations (Jackson and Gau 2015), the complex and dynamic network of trust between governments and citizens around algorithmic systems can either build or erode institutional trust.

To account for institutional trust, we identified the four dimensions of the institutional trust model in the state–citizen relationship: (1) policymakers' (dis)trust in automated systems; (2) policymakers' (dis)trust in citizens; (3) citizens' (dis)trust in automated systems; and (4) citizens' (dis)trust in policymakers (including governmental institutions, national parliament, public authorities and politicians).

In this dynamic trust model (see Figure 1), automated decision-making is not only understood as a system that efficiently and neutrally delivers services but as a medium that both demonstrates and affects how trust is built or eroded between the state and citizens. In this model, 'citizens' encompasses multiple groups of citizens, including migrants and refugees. 'State' entails a hierarchically heterogeneous set of actors, including policymakers, politicians, government officers, authorities and civil servants. Within this set of actors, it is worth distinguishing at least those who make and are responsible for political decisions and those who implement them at the street level. In certain welfare-state regimes, particularly in the Nordic social-democratic models, civil servants, situated between the state and its citizens, are considered an important pillar of democracy (Kaun et al. 2024). If civil servants are replaced or constrained by algorithms, their mediating role as well as the state–citizen relationship will change (Bovens and Zouridis 2002; Tammpuu and Masso 2018).

As Elish and Boyd (2018) and Kitchin (2017) argue, automated administration should be considered a sociotechnical phenomenon that goes beyond the merely technical aspects of algorithms. Automation has been introduced in strategic administrative branches where people commonly experience social exclusion, oppression, discrimination and marginalisation, such as social services and benefits, taxation, law and justice. Since many vulnerable groups of citizens are both the targets of automated decision-making systems' implementation and the largest user group that receives welfare benefits, the dynamics of trust primarily concern how mutual trust is built between public authorities and vulnerable citizens. What constitutes a vulnerable person or group is often context-specific, including features of life events (e.g., young or old age or illness), market factors (e.g., information asymmetry or market power), economic factors (e.g., poverty), factors linked to one's identity (e.g., gender, religion or culture) or other factors that leave some individuals in less powerful positions than average/ordinary citizens (Social Protection and Human Rights 2015).

Trust in algorithmic decision-making is essentially related to how people consider the transparency, explainability, fairness and accountability of the decisions made by machines. One key concern is that the problems with algorithmic decisions often regard their opacity—i.e., decisions are processed in 'black boxes' (Pasquale 2015). Being black boxes, algorithms impede human oversight to account for possible validity problems or bias in the input data or decision outcomes (Harcourt 2007; Janssen and Van den Hoven 2015; Kroll et al. 2017). Researchers argue that algorithms may lack transparency for at least

the following three reasons: (1) algorithms are protected by business secrecy (including the public sector) (Kaun 2021); (2) the data volume is too large for humans to process (Burrell 2016); and (3) in the case of AI, algorithms are constantly modified through machine learning (Danaher 2016). The explainability of algorithmic decisions is assumed to sustain transparency in automated decision-making (de Bruijn et al. 2022; Pasquale 2015), but it assumes a certain level of expertise from the public, and many citizens simply lack the knowledge to assess whether an algorithm-based decision is fair or not (Du et al. 2019). While citizens can be taught new procedures (e.g., code-reading skills) to better understand algorithms, this literacy does not increase understanding of, for example, the actual legality of decisions (Belle and Papantonis 2021).

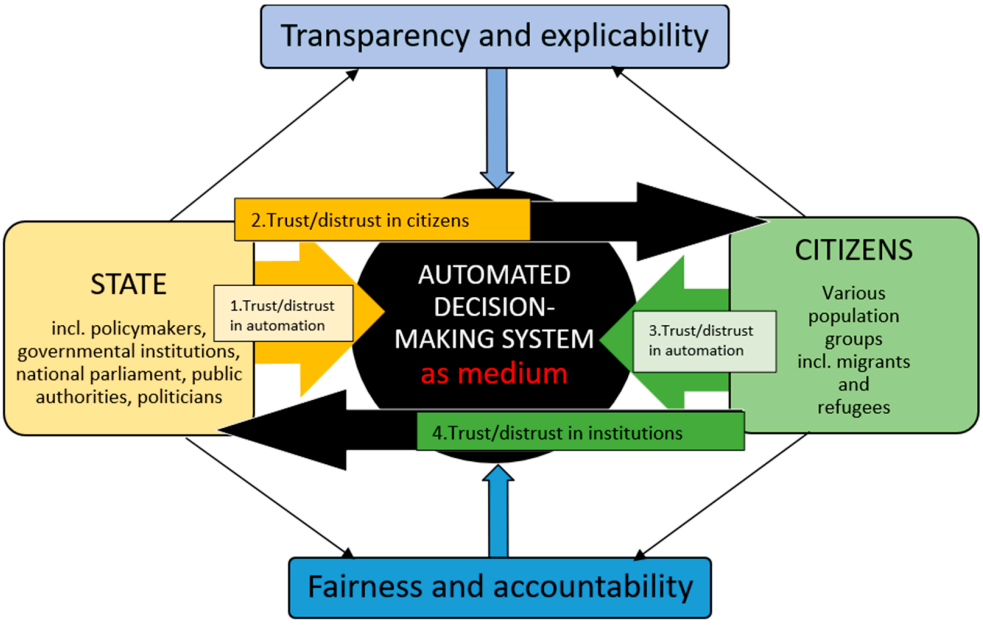

Figure 1. The four dimensions of the institutional trust model in the state–citizen relationship. Trust in automated systems is related to the transparency, explainability, fairness and accountability of decisions made using algorithms.

Another major concern that erodes trust in algorithmic decision-making is related to issues of fairness and accountability—that is, whether automated systems treat different groups of people equally. There exists no clear definition of fairness, and perceptions of fairness may differ between cultures, ethical theories, scientific disciplines and jurisdictions. The integration of algorithmic systems into administration processes has triggered intense debate among academics, policymakers and civil society about the conditions under which automation is acceptable or unacceptable and, above all, the kind of human discretion that is needed in automated processes (AlgorithmWatch 2019; Zouridis et al. 2020). Should human discretion in administrative work override algorithmic decisions when automation causes undesirable results in certain cases? If algorithmic calculation replaces human discretion, there is a risk that administrative systems will hide their official accountability for individual decisions behind digitalisation, blaming algorithmic complexity, system design flaws or lack of access to data (Widlak and Peeters 2020; Zalnieriute et al. 2019). Dencik et al. (2022) addressed this development in algorithmic public services characterised by new forms of professionalism. ICT experts and coders have entered the field of public administration; consequently, administrative work is increasingly delegated to either complex technological systems or ICT helpdesks (Veale and Brass 2019).

Most of the algorithmic decision-making systems used in public services are the so-called rule-based systems; they are not operated by machine learning algorithms (Engstrom

et al. 2020; De Sousa et al. 2019; Sun and Medaglia 2019). However, many administrative tasks are not straightforward and are easily reduced or defined using simple rule-based automation—namely, binary choices. The issues concerning social benefits that might seem 'objective' may be less so on closer inspection—they could instead have highly subjective and political aspects (Hadwick and Lan 2021). Some researchers have addressed several administrative tasks that should not be automated. Lipsky (2010) claimed that 'the nature of service provision calls for human judgment that cannot be programmed and for which machines cannot substitute' (p. 161). Petersen et al. (2020) examined the discretionary practices in the digitalisation of social work and found that automation remains a utopian concept due to the contradictory rules and case complexity in social services. They concluded that discretion in social work is a collaborative effort that cannot be codified and translated into machine-readable programs. As Veale and Brass (2019) put it, 'equitable and effective public services require judgment that cannot be quantified, reduced or encoded in fully automated systems' (p. 5). Interactions with often weak and vulnerable citizens (Minas 2014) demand a substantial level of individual and professional discretion (Lipsky 2010).

4. Literature Material

Next, we analyse the institutional trust model in terms of how the four dimensions of trust manifest themselves in recent literature that focuses on the automated decision-making systems introduced by democratic welfare states. We collected reviews and original studies published between 2017 and 2022 that dealt with automated decision-making and public services. A literature search was conducted on the Andor database using the following four search term clusters: 'automated decision-making', 'public service', 'vulnerable citizens' and 'original article'. The database search included articles in English from between 2017 and 2022 and yielded 366 hits. Altogether, 50 studies were manually included after the title and abstract skimming. Publications were included if they met the following criteria: (1) they covered automated decision-making implemented by public sector services with influence on vulnerable groups of citizens; (2) they were articles in peer-reviewed journals; (3) they were English-language publications; and (4) they were published between 2017 and 2022. The sample of 10 studies (see Table 1) are methodologically different and offer geographically different perspectives. Our analysis was based on the following four questions: (1) how do policymakers trust automated decision-making? (2) How does the potential distrust of policymakers in citizens manifest in their interest in deploying automated systems? (3) How do groups of vulnerable citizens trust automated systems? (4) How does the potential distrust of vulnerable groups of citizens towards automated decision-making systems manifest in their trust towards policymakers and public institutions?

Table 1. Summary of the studies on automated decision-making included in the discussion.

Article	Article Type	Objective	Results	Four Dimensions of Institutional Trust Model
Araujo et al. (2020): In AI we Trust? Perceptions about Automated Decision-Making by Artificial Intelligence	Scenario-based survey experiment	To analyse which personal characteristics can be linked to perceptions of fairness, usefulness and risk in automatic decision-making.	People are concerned about risks and have mixed opinions about the fairness and usefulness of automated at the societal level.	(3) citizens' trust or distrust in automated decision-making

Table 1. *Cont.*

Article	Article Type	Objective	Results	Four Dimensions of Institutional Trust Model
Helberger et al. (2020): Who Is the Fairest of Them All? Public Attitudes and Expectations Regarding Automated Decision-Making.	Original article (empirical survey and interview data)	To examine how the ongoing substitution of human decision-makers with ADM systems raises a question of ADM fairness.	A greater number of respondents considered AI to be a fairer decision-maker.	(3) citizens' trust or distrust in automated decision-making
Goggin and Soldatić (2022): Automated Decision-Making, Digital Inclusion and Intersectional Disabilities.	Discussion article	To gain a critical understanding of automatic decision-making through disability and intersectionality to frame the terms and agenda of digital inclusion for the future.	The study showed that an intersectional understanding of disabilities is not grasped in digital inclusion.	(1) policymakers' trust or distrust in automated decision-making (4) how the implementation of automated decision-making impacts the (dis)trust of (vulnerable) citizens in government
Griffiths (2021): Universal Credit and Automated Decision Making: A Case of the Digital Tail Wagging the Policy Dog?	Discussion article	To discuss digitalisation in welfare and questions of administrative burden and the wider effects and impacts on claimants.	The study showed that increasing digitalisation in public services brings an unnecessary administrative burden and other challenges to citizens.	(1) policymakers' trust or distrust in automated decision-making (4) how the implementation of automated decision-making impacts the (dis)trust of (vulnerable) citizens in government
Grimmelikhuijsen (2022): Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-Making.	Original article (empirical survey and interview data)	To discuss how citizens view algorithmic versus human decision-making.	The study concluded that accessibility is not enough to foster citizens' trust in automated decision-making.	(3) citizens' trust or distrust in automated decision-making
Kaun (2021): Suing the Algorithm: The Mundanization of Automated Decision-Making in Public Services through Litigation	Qualitative research based on in-depth interviews and court rulings	To analyse how different, partly conflicting definitions of what automatic decision-making in social services is and does are negotiated between multiple actors.	The article showed how different sociotechnical imaginaries related to automatic decision-making are established and stabilised.	(1) policymakers' trust or distrust in automated decision-making (4) how the implementation of automated decision-making impacts the (dis)trust of citizens in government

Table 1. *Cont.*

Article	Article Type	Objective	Results	Four Dimensions of Institutional Trust Model
Larsson and Haldar (2021): Can Computers Automate Welfare? Norwegian Efforts to Make Welfare Policy More Effective	Discussion: theoretical with an empirical case	To raise questions about the uncritical digitalisation of public services and the ability of welfare organisations to support healthy and inclusive societies.	The study argued that when developing automated digital public services, proactive automation should be precise in its delivery, inclusive of all citizens and still support welfare-orientated policies that are independent of the requirements of the digital system.	(1) policymakers' trust or distrust in automated decision-making (4) how the implementation of automated decision-making impacts the (dis)trust of citizens in government
Mökander et al. (2021): Ethics-Based Auditing of Automated Decision-Making Systems: Nature, Scope and Limitations	Review	To analyse the feasibility and efficacy of ethics-based auditing as a governance mechanism that allows organisations to operationalise their ethical commitments and validate claims made about their ADM systems.	The study concluded that ethics-based auditing should be considered an integral component of multifaced approaches to managing the ethical risks posed by ADM systems.	(4) how the implementation of automated decision-making impacts the (dis)trust of citizens in government
Ranerup and Henriksen (2019): Value Positions Viewed through the Lens of Automated Decision-Making: The Case of Social Services	Discussion: Theoretical with an empirical case	To discuss which instances of value positions and their divergence appear when ADM is used in municipal social assistance.	The study showed that automated systems has partly increased accountability, decreased costs and enhanced efficiency with a focus on citizens.	(1) policymakers' trust or distrust in automated decision-making (2) how the (dis)trust of citizens drives the implementation of automated systems (3) citizens' trust or distrust in automated decision-making
Sleep (2022): From Making Automated Decision Making Visible to Mapping the Unknowable Human: Counter-Mapping Automated Decision Making in Social Services in Australia	Descriptive article	To reflect on the act of counter-mapping ADM in social services in Australia.	The future automatic decision-making mapping needs to focus on making visible those who are subject to the decisions of automated systems but is usually made unknowable by the over-confident calculability of dominant automatic decision-making discourses.	(4) how the implementation of automated decision-making impacts the trust of (vulnerable)citizens in government

5. Discussion in the Sample of Studies on Automated Decision-Making

5.1. Policymakers Trust Automated Systems to Provide Services Efficiently and Equally

As legitimate powerholders, governments and public administrations are entitled to explore different technologies for public governance. Two of the selected articles explicitly addressed the question of authorities' and decision-makers' trust or distrust in automated decision-making (Larsson and Haldar 2021; Ranerup and Henriksen 2019), while three articles indirectly addressed it (Goggin and Soldatić 2022; Griffiths 2021; Kaun 2021).

Ranerup and Henriksen (2019) arrived at contradictory results regarding administrative actors' trust in automated systems. By interviewing politicians and professionals about social service delivery, they found that these systems have, in some respects, increased accountability, reduced the cost of producing services and enhanced efficiency from the perspective of authorities. Some participants, however, were concerned about the lack of professional discretion in the final decisions. A lack of transparency regarding technology's role in decisions was highlighted. Automated systems were considered useful only for playing a secondary and supportive role in the process.

Larsson and Haldar (2021) focused on Norway's child benefit system, which is a proactive form of automated decision-making. According to the data from Norway's national resident register, the system automatically awards child benefits to parents whom the system identifies as eligible. Larsson and Haldar (2021)'s main finding was that proactive automation failed to be inclusive of all citizens in its benefit delivery. Some of the parents who were entitled to benefits, such as single parents and citizens who lived abroad, did not automatically receive child support. According to the authors' analysis, the limitations of the algorithmic system prevented all parents who were entitled to benefits from receiving these benefits automatically. This study showed that the Norwegian government's trust in digital systems led to imprecise automation; consequently, some parents and those in vulnerable situations, such as single parents, had to manually apply for child benefits.

Authorities' overconfidence in and the algorithmic bias of automated systems were indirectly discussed in three other articles (Goggin and Soldatić 2022; Griffiths 2021; Kaun 2021). Goggin and Soldatić (2022) focused on the Australian government's digital programme, widely known by its nickname, 'Robodebt'. Robodebt's unfairness was not intended by the policymakers, but many of those significantly harmed by it were Indigenous Australians with disabilities. This was due to the intersectionality of this diverse group of people and the sanctions of the automated system indirectly targeting them. Robodebt's scandalous failure exposed the Australian government's overconfidence in automation's ability to deliver public benefits efficiently and equally. Robodebt was used for welfare debt recovery in such a way that it did not include human discretion at any stage. Those in receipt of debt recovery letters found it difficult to provide relevant information, ask questions or contest the system's decisions. A threshold issue was that, as part of a drive to digitise all forms of welfare citizen engagement, Robodebt was designed to minimise staff costs and make it difficult to lodge complaints (Goggin and Soldatić 2022).

Goggin and Soldatić (2022) did not explicitly study policymakers or their trust in automated decision-making, but their conclusions indicate that the government had an automation bias regarding the functionality of the system. Their analysis of digital inclusion demonstrated how a lack of administrative knowledge of Indigenous disabled people's capabilities led to catastrophic consequences. The administrative actors were blinded by the idea that everyone had access to and could use digital devices and failed to recognise the special aspects of Indigenous disabled people's lives—that they lacked the resources, such as the financial, social and cultural capital to avail the necessary technology to communicate and engage with the digital governance apparatus. Goggin and Soldatić

(2022) also demonstrated that being both an Indigenous and a disabled citizen—a case where two social hindrances intersect—constituted the danger of being digitally excluded.

As in Goggin and Soldatić (2022), similar features of governmental overconfidence in automated systems can be seen in Griffiths (2021), which focused on Universal Credit (UC) in the UK, although the author did not explicitly discuss the concepts of trust, (over)confidence or algorithmic bias. The UK's Conservative government launched UC based on an automated decision-making system. The governmental actors promised automated payment processing, where applicants would 'go from application to receiving help without any manual intervention' (cited in Griffiths 2021, p. 1). UC was intended to ensure that claimants 'automatically receive everything they are entitled to' (cited in Griffiths 2021, p. 1). However, beset by a series of technical setbacks, the system's target completion date was reset numerous times. Griffiths (2021) highlighted that, against the expectations and objectives of policymakers, the Department for Work and Pensions' spending on ICT increased. Automatically driving down errors, fraud and overpayments turned out to be more difficult than expected due to several reasons (for instance, the ID verification system enabled fraudsters to pose as applicants).

Based on analyses of interviews, documents and news pieces, Kaun (2021)'s qualitative study discussed the use of robotics in a Swedish municipality. Widely known as the 'Trelleborg case', the dispute concerned whether citizens had the right to know what kind of algorithmic code was used to make decisions about their social benefits. In the end, the court ruled in favour of the citizens. We can only speculate as to whether the reluctance of the administrative authorities to reveal the system's algorithms implies that they felt there was no need to reveal the code because they believed the algorithms worked efficiently and fairly. In Kaun (2021)'s interviews, some of the civil servants who had developed the system admitted difficulties in understanding how the algorithm functioned; however, they seemed to trust its ability to make decisions. The group of developers *anthropomorphised* the robot system by naming it 'Ernst' to make it more familiar to themselves (Kaun 2021). Issues of trust are complex and volatile, especially in conditions in which actors feel that an automated system is not sufficiently transparent. The deepest understanding of code often exists in private design companies that provide automated systems for the public sector. As studies of compensatory institutional trust have shown, similar to ordinary citizens confronting complex systems and compensating for their lack of knowledge by excessively trusting the systems, decision-makers and administrative actors can compensate for their lack of knowledge with excessive trust in the systems (Shockley and Shepherd 2016).

5.2. Policymakers' Distrust of Citizens Drives the Implementation of Automated Decision-Making Systems for Administration

Only Ranerup and Henriksen (2019) explicitly addressed policymakers' distrust of citizens as their motivation for developing automated systems. The study identified three dimensions of trust: how the distrust of citizens drives the implementation of automated systems, how the government and authorities trust automation and how citizens trust these systems. Regarding the first dimension, Ranerup and Henriksen (2019) offered an official report highlighting a public authority's mistrust of citizens and justified the benefits of implementing an automated decision-making system:

> The trust in the citizens we serve is too low among public agencies in Sweden. The system is based on the notion that the majority cheat. The control system is designed with that in mind. Our activities are organized for the majority instead of the minority. (p. 7)

This new automated model, as it was associated with specific regulations, was considered to increase the authority's trust in citizens. This contrasted with the previous practices

whereby applicants had to send extensive documentation as an obligatory part of their application process. This automation model shifted the management of cases for social assistance from the social assistance agency to the labour market agency. The control was shifted to a caseworker at the labour market agency who decided whether an applicant was willing to accept employment offers. This evaluation was used in decision-making on social welfare. Officials' confidence in citizens' ability to send applications increased with the automated system; however, at the same time, officials acknowledged the lack of transparency regarding how the algorithm made decisions. Ranerup and Henriksen (2019) thus argued that administrative actors make conflicting value choices: though the actors increased their trust in citizens with an automated system, the system was not transparent in terms of how it made decisions.

Ranerup and Henriksen (2019) explicitly focused on the fact that the policymakers' trust and distrust of citizens motivated them to develop an automated system. This was not discussed in other studies. We can only speculate how far the development of the Australian government's Robodebt (Goggin and Soldatić 2022), UC in the UK (Griffiths 2021) and Trelleborg's robotic system in Sweden (Kaun 2021), among others, were motivated by a distrust of citizens. From the perspective of institutional trust, citizens' trust in public administration can be highly calibrated if they do not realise that the administration deserves, deliberately or not, to be trusted (Neal et al. 2016). In automated decision-making systems, without special expertise, it is extremely difficult for citizens to understand the features enabling surveillance and control that potentially target them.

5.3. The Perspectives of Vulnerable Citizens Are Missing in the Discussions on Trust in Automated Systems

As mentioned above, the legitimate use of automated decision-making in public administration partly depends on citizens' trust in automated systems. In the selected studies, the researchers were mostly interested in the question of how and on what basis the public trusts the administrative use of algorithms, although none of the studies dealt with vulnerable groups' explicitly perceived trust in automation (Araujo et al. 2020; Grimmelikhuijsen 2022; Helberger et al. 2020).

Unlike previous research in which human experts were more trusted than technologies (Madhavan and Wiegmann 2007), the findings of Araujo et al. (2020) and Helberger et al. (2020) show that respondents trust automated decision-making more than human decision-makers. In Helberger et al.'s (2020) survey-based study, the majority of the representative sample of the Dutch adult population considered the decisions made by AI-driven systems to be fairer than those made by human discretion. One of the major reasons for this is that automated decision-making was generally found/believed to be more impartial and immune from manipulation. Araujo et al.'s (2020) scenario-based survey revealed that the decisions taken automatically by AI were often better evaluated than those taken by human experts. The researchers argue that one of the reasons for the respondents' trust in automated decision-making could be their lack of trust in human decision-making more generally.

Both studies offer mixed views about the perceptions of trust in automated decision-making. In Helberger et al. (2020), the ability to consider the broader context of human discretion was considered fairer than computer systems in complex cases. Many respondents thought that there was a limit to generalisability and the modelling of reality. However, the role of emotions was found to be more controversial: some participants considered it fairer for emotions to be part of decision-making, while others considered the decisions made without emotional influence fairer. Araujo et al. (2020) showed that the respondents were by and large concerned about the risks of algorithms and had mixed opinions about the fairness and usefulness of automated decision-making at the societal level, with their

general attitudes towards automation influenced by their individual characteristics, such as age and education level.

Neither Araujo et al. (2020) nor Helberger et al. (2020) specified whether there were differences in terms of trust if the respondents belonged to vulnerable groups or were well off. However, both studies revealed that the individual features of citizens—gender, age and level of education—influence their perceptions of automated systems' fairness and usefulness. The younger and more educated the respondents were, the more likely they were to consider automated decision-making fairer than human decision-making (Helberger et al. 2020). Some of the respondents in Helberger et al. (2020) also considered these systems' wider societal context and long-term societal implications. One respondent was pessimistic about the increased use of automation in administration, stating that it may lead to the hardening of attitudes in society, such as encouraging the treatment of humans as pure statistics and dividing people into 'winners and losers' (Helberger et al. 2020). Although this was only an individual opinion, it showed that citizens did not see automation as a neutral tool of administration but that it had political and societal consequences.

Grimmelikhuijsen (2022)'s survey experimented with hypothetical scenarios concerning two core elements of algorithmic transparency—accessibility and explainability—that were defined as crucial for strengthening the perceived trustworthiness of street-level decision-making. Contrary to Araujo et al. (2020) and Helberger et al. (2020), Grimmelikhuijsen (2022) found that human decision-makers (public officials) were trusted more than algorithms, although the latter were initially hailed as improving the efficiency and fairness of administrative decision-making. Grimmelikhuijsen (2022) argues that algorithmic transparency should not be merely understood as 'access to code'. Although source code is important for accountability, it is unlikely to increase people's understanding or the perceived trustworthiness of algorithmic decision-making. Moreover, the author showed that the explainability of algorithms positively affected citizens' trust in algorithms. Accessibility was not enough to create trust, and public authorities needed to address the explainability of algorithmic decision-making to earn citizens' institutional trust in algorithms and in the public officials who handle these algorithms.

5.4. Do Discriminating Systems Erode Vulnerable Citizens' Trust in Administrative Operations?

In the reviewed literature, very few studies have addressed the question of how the implementation of automated decision-making potentially impacts the trust of citizens, not to mention the vulnerable citizen groups in society, in public administration. Larsson and Haldar (2021)'s study of Norway's child benefits discussed the social impact that occurs when vulnerable citizens are excluded from a system; all those who are exceptions to the 'norm' (risking a false benefit decision—for instance, in the case of single parenthood) had to manually apply for the benefit. The authors suggested that if public authorities trust digital systems more than citizens, they could lose the vulnerable citizens' trust in them. If the authorities extend the scope to all possible recipients, the administration will increase trust and responsibility among citizens. The current technical system, however, does not allow for such changes.

In their study of the ethics-based auditing of automated systems, Mökander et al. (2021) argued that automated decision-making in general may undermine the role of trust in and the legitimacy of liberal democracy. They found that ethics-based auditing systems linked to automation can increase public trust in technology and improve user satisfaction by enhancing operational consistency and procedural transparency. The authors thus underlined the need to implement automated decision-making in administration as a natural part of the discussion on ethics and as a tool to strengthen institutional

trust. Moreover, the authors highlighted that if ethical challenges are not sufficiently addressed, a lack of institutional trust in automation may hamper the adoption of such systems. Following Cookson (2018), the authors argued that public distrust leads to significant costs through the underuse of available and well-designed technologies. That said, by promoting transparency, automated decision-making could facilitate morally good actions and strengthen trust between different stakeholders. Mökander et al. (2021), though they did not focus on vulnerable groups, highlighted that, while automation may improve the overall accuracy of decisions, it could discriminate against specific subgroups in the population.

Trust as a concept did not appear in Griffiths (2021), not to mention citizens' trust in the government's capabilities to implement automated decision-making systems at the administrative level. However, aspects related to citizens' trust in the government emerged. This case study revealed how people with irregular earnings—a typical status for vulnerable people—found that UC payments could fluctuate unpredictably, causing these people to have more budgeting difficulties than those with more regular incomes. The claims of vulnerable citizens could thus face more rejections and fault statements than others. Self-managing systems have transformed the administrative burden from public officials to individual citizens. Indirectly, the electronic systems seemed unfair and discriminating, especially for women who uploaded data to the UC system on behalf of their family members. Some working women even reduced their work hours so that they could self-manage the system for their families. We can conclude that the difficulties faced by vulnerable groups when using the UC system could have decreased their confidence in the capability of public authorities to design these systems.

From the angle of institutional trust, an even more crucial point is the observation by Griffiths (2021) that exerting greater compliance and administrative demands from citizens, who are claimants, through automated systems is potentially overrunning the traditional notion of accountability, according to which administration is expected to be accountable for citizens and not the other way around. If citizens, as users of automated systems, feel that they are placed in a defensive position by the administration, this will probably decrease their institutional trust in the long run. Goggin and Soldatić (2022), Helberger et al. (2020), Kaun (2021) and Sleep (2022) shared the same observation that citizens' ability to defend their rights and object to automated decisions is highly restricted—doing so is only possible in the automated processes, where administrative procedures can be argued to have been illegal. Helberger et al. (2020) questioned whether there should be additional grounds that entitle people to object to automated decisions—for example, arguments related to dignity, a lack of trust in such systems rendering a balanced decision or moral objections to the very idea of being subject to automation.

Goggin and Soldatić's (2022) analysis did not explicitly include discussions of trust in how the Robodebt scandal potentially erased the trust of Indigenous people in administrative operations (which seemed to be low in the first place). However, the researchers stressed that the affected citizens—some of the poorest and most precarious in Australia—often found it difficult to 'speak back' to the authorities and have their responses acknowledged by them, as the digital and bureaucratic systems were not designed for good administrative or political 'listening' (Goggin and Soldatić 2022; McNamara 2018). This Australian case is helpful because it clearly shows the unintended and intended impacts of automated decision-making systems on vulnerable populations' digital inclusion.

Drawing on the case of Australian social services, Sleep (2022) reflected on how counter-mapping can provide a resistance method for moving beyond the dominant discourses of efficiency and cost-cutting to allow the voices of vulnerable groups to be heard when developing automated systems. By counter-mapping, she means grassroots-level

activism to resist the dominant discourses. While Sleep (2022) confessed that her mapping exercise was based on privileged expertise—rather than amplifying the voices of less powerful people—this tool can make visible those who are subject to the decisions of automated systems but who are usually made unknowable by the overconfident calculability of dominant discourses. She argues that the users of AI-driven services were unfairly treated by the administration and that more research is needed to focus on matching their attitudes towards procedures of public administration.

Kaun (2021) showed how the work of investigative journalists and unions for professionals significantly contributed to the algorithm code becoming publicly accessible. It was only after extensive publicity and lawsuits that the transparency of public service decisions also came to include algorithms. However, if national media and other significant institutions face such difficulties, this raises the question of how citizens in fragile positions can have their voices heard.

6. Conclusions

In this theoretical study, we aimed to develop a novel conceptual model capable of identifying how the deployment of automated decision-making systems erodes or strengthens the institutional trust between policymakers and citizens. We did not define institutional trust as trust in institutions; we understood it as a two-way relationship between governmental institutions and citizens. Using a sample of recent studies on automated decision-making, we examined the potential of the institutional trust model to identify how the four dimensions of trust can be used to enhance the citizen–state relationship.

Administrative actors' overconfidence in automated systems. First, by exploring policymakers' trust in automation, we perceived that most of the studies dealt with the topic only indirectly. However, it is reasonable to argue that administrative actors are overconfident about automated systems, even though these systems often fail to organise public services equally or fairly for everyone. This technological idolisation is a result of automation in some respects having increased accountability, decreased costs of service provision and enhanced efficiency from the authorities' perspective. Having excessive faith and overconfidence in the scientific validity, neutrality and rationality of algorithmic procedures and results often leads to algorithmic bias (see Peeters 2020). However, administrative actors' trust in automated systems varies with respect to their professional discretion and the systems' lack of transparency. The lack of transparency can lead to misplaced trust in automated systems if the officials or decision-makers compensate for their lack of technical expertise with trust. More than policymakers, street-level administrative actors are concerned about the lack of professional discretion in the systems' final decisions.

Policymakers' distrust of citizens. We then focused on how policymakers' potential distrust of citizens manifests in their motivation to build automated systems. Only one article clearly highlighted that the implementation of automated decision-making systems was strongly influenced by the assumption that most citizens cheat to get benefits to which they are not entitled. However, in many articles, the authors indirectly indicated that a strong driver of the development of automated decision-making systems was the administrative actors' distrust of citizens. Consequently, these systems, since they were associated with specific regulations, were seen as increasing the authorities' trust in citizens. Moreover, distrust of public administration can lead to increased control of citizens who are dependent on social benefits. For example, self-reporting, which is typical of automated systems, exerts greater compliance and administrative demands on citizens. As Coglianese (2023) suggests, the implementation of automated systems results in the administration losing its empathetic relationship with its citizens. If citizens feel that the automated systems somehow place them in a more defensive position than before in the face of public

administration, they would be left discontented, which can eventually affect their trust in public administration.

The voice of vulnerable groups is missing in research. We then explored vulnerable citizens' trust in automated systems. While several articles focused on how and on what basis the public trusts the administrative use of algorithms, none of them considered vulnerable groups. Nevertheless, our findings support the existing notion that citizens hold mixed views about the reliability of automated systems. Individuals trusted automated systems when they perceived or believed automated decision-making as more impartial and immune to manipulation than human decision-makers. However, in complex cases, the ability to consider the broader context of human discretion is considered fairer than automated systems. We conclude that the socioeconomic characteristics of citizens—their level of education, gender and age—influence their degree of trust in automated systems. The younger and more educated an individual is, the more likely they are to consider automated decision-making fairer than human decision-making. Moreover, citizens do not view algorithms as neutral administration tools; instead, they acknowledge algorithms' political and societal consequences. Citizens' former experiences in dealing with public officials also influence their institutional trust or distrust.

The lack of a feedback loop in automated systems weakens citizens' trust in the administration. Finally, we explored how the vulnerable citizens' potential distrust in automated decision-making systems affects their trust in policymakers and public institutions. This issue was only indirectly addressed in our research sample. Based on their findings, some of the authors in our sample speculated that authorities lose citizens' trust in the administration if the automated systems significantly harm the citizens or treat some groups unequally or unfairly. In many cases, citizens' ability to defend their rights and object to automated decisions was shown to be restricted. It is thus reasonable to state that automated systems can erode citizens' trust in an administration, as automated systems lack a feedback mechanism for policymakers. Moreover, Shockley and Shepherd (2016)'s discussion of institutional trust shows that the citizens who belong to the most deserving and least powerful group probably have to compensate for feeling insecure by trusting—instead of resisting—the automated systems. Since these systems are not designed to contribute 'administrative listening', the potentially growing distrust of citizens towards the legitimacy of liberal democracy remains a kind of 'black box' for policymakers. As Carney (2023) states, when once broken, restoring trust in the function of welfare systems can be challenging for policymakers. If future automation is to remain faithful to the values of transparency and fairness, governments should engage with marginalised communities to develop new innovative solutions.

This article contributes to the literature on automated decision-making in the field of social science in three major ways. First, we argue that, instead of being a neutral bureaucratic tool in the public sector, an automated decision-making system can be understood as an affective medium that links the state and its citizens, potentially causing tensions and conflicts. Second, our institutional trust model extends the notion of trust to include public authorities' trust in technology and citizens—an aspect that has been almost entirely absent from the current discussion on automated systems. Consequently, we suggest that institutional trust should be approached from the viewpoint of policymakers, since governmental institutions are obligated towards citizens—not just for building legitimacy but for sustaining 'an empathic state'. Our research thus contributes to the discussions on the impact of automated decision-making on institutional trust and how this trust influences the legitimacy of liberal democracies. Finally, by highlighting vulnerable citizens as the key targets of automated intervention and as the key users of automated systems, we wish to highlight the issue of trust regarding this group. Our discussion revealed a lack

of empirical studies on vulnerable groups' perceived trust in automated decision-making systems. Only a few empirical studies have systematically evaluated the effects of these systems on these groups, which involve individuals with disabilities, mental health patients, the unemployed and immigrants. As Carney (2023) remarked, the conversation in social services about automated systems' implications for the socioeconomic rights of marginalised citizens has barely begun. Our theoretical paper is not exhaustive, which means that extensive empirical and theoretical work must be conducted to understand how deploying automated decision-making in public services influences the institutional trust between the state and citizens.

Author Contributions: Conceptualization, J.P. and A.K.; methodology, L.E. and J.P.; software, L.E.; validation, P.A. and C.L.; formal analysis, H.P.; investigation, J.P., L.E. and H.P.; resources, J.P.; data curation, L.E.; writing—original draft preparation, J.P. and A.K.; writing—review and editing, All authors; visualization, J.P.; supervision, J.P.; project administration, J.P.; funding acquisition, J.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Kone Foundation, grant number 202102433.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analysed in this study.

Acknowledgments: We would like to thank the journal's anonymous reviewers for their valuable and insightful comments.

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

References

Adam-Troian, Jais, Maria Chayinska, Maria Paola Paladino, Ödzen Melis Uluğ, Jeroen Vaes, and Pascal Wagner-Egger. 2023. Of Precarity and Conspiracy: Introducing a Socio-functional Model of Conspiracy Beliefs. *British Journal of Social Psychology* 62: 136–59. [CrossRef] [PubMed]

Agar, Jon. 2003. *The Government Machine. A Revolutionary History of the Computer*. Cambridge: MIT Press.

AlgorithmWatch. 2019. *Automating Society Taking Stock of Automated Decision-Making in the EU. A Report by AlgorithmWatch in cooperation with Bertelsmann Stiftung, Supported by the Open Society Foundations*, 1st ed. Berlin: AW AlgorithmWatch GmbH. Available online: www.algorithmwatch.org/automating-society (accessed on 6 March 2025).

Anderson, Mark, and Karën Fort. 2022. Human Where? A New Scale Defining Human Involvement in Technology Communities from an Ethical Standpoint. *International Review of Information Ethics* 31: hal-03762035. [CrossRef]

Araujo, Theo, Natali Helberger, Sanne Kruikemeier, and Claes H. De Vreese. 2020. In AI We Trust? Perceptions about Automated Decision-making by Artificial Intelligence. *AI & Society* 35: 611–23.

Behnam Shad, Klaus. 2023. Artificial Intelligence-related Anomies and Predictive Policing: Normative (Dis)orders in Liberal Democracies. *AI & Society*, 1–12. [CrossRef]

Belle, Vaishak, and Ioannis Papantonis. 2021. Principles and Practice of Explainable Machine Learning. *Frontiers in Big Data* 4: 688969. [CrossRef]

Berg, Monika, and Tobias Johansson. 2020. Building Institutional Trust Through Service Experiences—Private Versus Public Provision. *Journal of Public Administration Research and Theory* 30: 290–306. [CrossRef]

Bovens, Mark, and Stavros Zouridis. 2002. From Street-level to System-level Bureaucracies: How Information and Communication Technology is Transforming Administrative Discretion and Constitutional Control. *Public Administration Review* 6: 174–84. [CrossRef]

Burrell, Jenna. 2016. How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data & Society* 3: 1–12.

Carney, Terry. 2019. Robo-debt Illegality: The Seven Veils of Failed Guarantees of the Rule of Law? *Alternative Law Journal* 44: 4–10. [CrossRef]

Carney, Terry. 2023. The Automated Welfare State: Challenges for Socioeconomic Rights of the Marginalised. In *Money, Power, and AI: Automated Banks and Automated States*. Edited by Zophia Bednarz and Monika Zalnieriute. Cambridge: Cambridge University Press, pp. 95–115.

Casey, Simone J. 2022. Towards Digital Dole Parole: A Review of Digital Self-service Initiatives in Australian Employment Services. *Australian Journal of Social Issues* 57: 111–24. [CrossRef]

Choi, Intae. 2016. *Digital Era Governance: IT Corporations, the State, and e-Government*. Abingdon: Taylor & Francis.

Coglianese, Cary. 2023. Law and Empathy in the Automated State. In *Money, Power, and AI: Automated Banks and Automated States*. Edited by Zophia Bednarz and Monika Zalnieriute. Cambridge: Cambridge University Press, pp. 173–88.

Cookson, Clive. 2018. Artificial Intelligence Faces Public Backlash, Warns Scientist. *Financial Times*. June 9. Available online: https://www.ft.com/content/0b301152-b0f8-11e8-99ca-68cf89602132 (accessed on 6 March 2025).

Cordella, Antonio, and Niccolò Tempini. 2015. E-government and Organizational Change: Reappraising the Role of ICT and Bureaucracy in Public Service Delivery. *Government Information Quarterly* 32: 279–86. [CrossRef]

Crouch, Colin. 2020. *Post-Democracy After the Crises*. Cambridge: Polity Press.

Danaher, John. 2016. The Threat of Algocracy: Reality, Resistance and Accommodation. *Philosophy & Technology* 29: 245–68.

de Bruijn, Hans, Martijn Warnier, and Martijn Janssen. 2022. The Perils and Pitfalls of Explainable AI: Strategies for Explaining Algorithmic Decision-Making. *Government Information Quarterly* 39: 1–8. [CrossRef]

De Sousa, Weslei G., Elis R. P. De Melo, Paulo H. De Souza Bermejo, Rafael A. Sousa Farias, and Adalmir O. Gomes. 2019. How and Where is Artificial Intelligence in the Public Sector Going? A Literature Review and Research Agenda. *Government Information Quarterly* 36: 101392. [CrossRef]

Dencik, Lina, Arne Hintz, Joanna Redden, and Emiliano Treré. 2022. *Data Justice*. Thousand Oaks: Sage Publications.

Du, Mengnan, Ninghao Liu, and Xia Hu. 2019. Techniques for Interpretable Machine Learning. *Communications of the ACM* 63: 68–77. [CrossRef]

Elish, Madeleine C., and Danah Boyd. 2018. Situating Methods in the Magic of Big Data and AI. *Communication Monographs* 85: 57–80. [CrossRef]

Engstrom, David F., Daniel E. Ho, Catherine M. Sharkey, and Mariano-Florentino Cuéllar. 2020. *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*. New York: NYU School of Law, pp. 20–54. [CrossRef]

Ervasti, Heikki, Antti Kouvo, and Takis Venetoklis. 2019. Social and Institutional Trust in Times of Crisis: Greece, 2002–2011. *Social Indicators Research* 141: 1207–31. [CrossRef]

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

European Commission. 2023. Automated Decision-Making Impacting Society. The Pervasiveness of Digital Technologies Is Leading to an Increase in Automated Decision-Making. Available online: https://knowledge4policy.ec.europa.eu/foresight/automated-decision-making-impacting-society_en (accessed on 16 February 2025).

European Committee for Democracy and Governance. 2021. Study on the Impact of Digital Transformation on Democracy and Good Governance. Available online: https://rm.coe.int/study-on-the-impact-of-digital-transformation-on-democracy-and-good-go/1680a3b9f9 (accessed on 6 March 2025).

European Parliament. 2022. Digitalisation and Administrative Law. European Added Value Assessment. Available online: https://www.europarl.europa.eu (accessed on 6 March 2025).

Gaffney, Stephen, and Michelle Millar. 2020. Rational Skivers or Desperate Strivers? The Problematisation of Fraud in the Irish Social Protection System. *Critical Social Policy* 40: 69–88. [CrossRef]

Giddens, Anthony. 1991. *The Consequences of Modernity*. Cambridge: Polity Press.

Goggin, Gerard, and Karen Soldatić. 2022. Automated Decision-making, Digital Inclusion and Intersectional Disabilities. *New Media & Society* 24: 384–400.

Griffiths, Rita. 2021. Universal Credit and Automated Decision Making: A Case of the Digital Tail Wagging the Policy Dog? *Social Policy and Society: A Journal of the Social Policy Association* 23: 1–18. [CrossRef]

Grimmelikhuijsen, Stephan. 2022. Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-making. *Public Administration Review* 23: 1–18. [CrossRef]

Guszcza, James, David Schweidel, and Shantanu Dutta. 2014. The Personalized and the Personal: Socially Responsible Innovation through Big Data. *Deloitte Review* 14: 95–109.

Győrffy, Dóra. 2013. *Institutional Trust and Economic Policy Lessons from the History of the Euro: Lessons from the History of the Euro*. Amsterdam: Amsterdam University Press.

Hadwick, David, and Shimeng Lan. 2021. Lessons to Be Learned from the Dutch Childcare Allowance Scandal: A Comparative Review of Algorithmic Governance by Tax Administrations in the Netherlands, France and Germany. *World Tax Journal* 13: 1–53. [CrossRef]

Harcourt, Bernard E. 2007. *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*. Chicago: Chicago University Press.

Hardin, Russel. 1999. Do We Want to Trust the Government? In *Democracy and Trust*. Edited by Mark E. Warren. Cambridge: Cambridge University Press, pp. 22–41.

Helberger, Natali, Theo Araujo, and Claes H. De Vreese. 2020. Who is the Fairest of Them All? Public Attitudes and Expectations regarding Automated Decision-making. *Computer Law & Security Review* 39: 105456.

Hetherington, Marc J. 2004. *Why Trust Matters: Declining Political Trust and the Demise of American Liberalism*. Princeton: Princeton University Press.

Inglehart, Ronald, and Christian Wenzel. 2005. *Modernization, Cultural Change, and Democracy: The Human Development Sequence*. Cambridge: Cambridge University Press.

Jackson, Jonathan, and Jacinta M. Gau. 2015. Carving up Concepts? Differentiating between Trust and Legitimacy in Public Attitudes towards Legal Authority. In *Interdisciplinary Perspectives on Trust: Towards Theoretical and Methodological Integration*. Edited by Ellie Shockley, Tess M. S. Neal, Lisa M. PytlikZillig and Brian H. Bornstein. Berlin and Heidelberg: Springer International Publishing AG, pp. 49–69.

Janssen, Marijn, and Jaroen Van den Hoven. 2015. Big and Open Linked Data (BOLD) in Government: A Challenge to Transparency and Privacy? *Government Information Quarterly* 32: 363–68. [CrossRef]

Johansson, Bengt, Jacob Sohlberg, Peter Esaiasson, and Marina Ghersetti. 2021. Why Swedes Don't Wear Face Masks during the Pandemic—A Consequence of Blindly Trusting the Government. *Journal of International Crisis and Risk Communication Research* 4: 335–58. [CrossRef]

Kaun, Anne. 2021. Suing the Algorithm: The Mundanization of Automated Decision- Making In Public Services through Litigation. *Information, Communication & Society* 25: 2046–62.

Kaun, Anne, Anders O. Larsson, and Anu Masso. 2024. Automating Public Administration: Citizens' Attitudes towards Automated Decision-making across Estonia, Sweden, and Germany. *Information, Communication & Society* 27: 314–32. [CrossRef]

Kitchin, Rob. 2017. Thinking Critically About and Researching Algorithms. *Information, Communication & Society* 20: 14–29.

Kroll, Joshua Joanna Huey A., Salon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu. 2017. Accountable Algorithms. *University of Pennsylvania Law Review* 165: 633–705.

Larsson, Karl K., and Marit Haldar. 2021. Can Computers Automate Welfare? Norwegian Efforts to Make Welfare Policy More Effective. *Journal of Extreme Anthropology* 5: 56–77. [CrossRef]

Lee, Min K. 2018. Understanding Perception of Algorithmic Decisions: Fairness, Trust, And Emotion in Response to Algorithmic Management. *Big Data & Society* 5: 2053951718756684.

Lipsky, Michael. 2010. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation.

Madhavan, Poornima, and Douglas A. Wiegmann. 2007. Effects of Information Source, Pedigree, and Reliability on Operator Interaction with Decision Support Systems. *Human Factors* 49: 773–85. [CrossRef]

McNamara, Jim. 2018. *Organizational Listening: The Missing Essential Ingredient in Public Communication*. Lausanne: Peter Lang.

Meijer, Albert, and Stephan Grimmelikhuijsen. 2020. Responsible and accountable algorithmization. How to generate citizen trust in governmental usage of algorithms. In *The Algorithmic Society: Technology, Power, and Knowledge*. Edited by Marc Schuilenburg and Rik Peeters. London: Routledge.

Miller, Tim. 2019. Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artificial Intelligence* 267: 1–38. [CrossRef]

Minas, Renate. 2014. One-stop Shops: Increasing Employability and Overcoming Welfare State Fragmentation? *International Journal of Social Welfare* 23: S40–S53. [CrossRef]

Mökander, Jakob, Jessica Morley, Mariarosaria Taddeo, and Luciano Floridi. 2021. Ethics-Based Auditing of Automated Decision-Making Systems: Nature, Scope, and Limitations. *Science and Engineering Ethics* 27: 44–44. [CrossRef]

Neal, Tess M. S., Ellie Shockley, and Oliver Schilke. 2016. The "Dark Side" of Institutional Trust. In *Interdisciplinary Perspectives on Trust*. Edited by Ellie Shockley, Tess Neal, Lisa PytlikZillig and Brian Bornstein. Cham: Springer. [CrossRef]

OECD. 2024. *OECD Survey on Drivers of Trust in Public Institutions—2024 Results: Building Trust in a Complex Policy Environment*. Paris: OECD Publishing. [CrossRef]

Offe, Claus. 1999. How Can We Trust Our Fellow Citizens? In *Democracy and Trust*. Edited by Mark E. Warren. Cambridge: Cambridge University Press, pp. 42–87.

O'Neill, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing.

O'Neill, Onora. 2002. *A Question of Trust*. Cambridge: Cambridge University Press.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard: Harvard University Press.

Peeters, Rik. 2020. The Agency of Algorithms: Understanding Human-Algorithm Interaction in Administrative Decision-Making. *Information Polity* 25: 507–22. [CrossRef]

Petersen, Anette, Lars R. Christensen, and Thomas T. Hildebrandt. 2020. The Role of Discretion in the Age of Automation. *Computer Supported Cooperative Work* 29: 303–33. [CrossRef]

Pérez-Morote, Rosario, Carolina Pontones-Rosa, and Montserrat Núñez-Chicharro. 2020. The Effects of E-Government Evaluation, Trust and the Digital Divide in the Levels of E-Government Use in European Countries. *Technological Forecasting and Social Change* 154: 119973. [CrossRef]

Power, Martin, Eoin Devereux, and Majka Ryan. 2022. Framing and Shaming: The 2017 Welfare Cheats, Cheat Us All Campaign. *Social Policy and Society* 21: 646–56. [CrossRef]

Ranerup, Agneta, and Helle Z. Henriksen. 2019. Value Positions Viewed Through the Lens of Automated Decision-making: The Case of Social Services. *Government Information Quarterly* 36: 101377. [CrossRef]

Redden, Joanna, Jessica Brand, Ina Sander, and Harry Warne. 2022. *Automating Public Services: Learning from Cancelled Systems*. Carnegie: Data Justice Lab.

Reddick, Christopher G. 2005. Citizen Interaction with E-Government: From the Streets to Servers? *Government Information Quarterly* 22: 38–57. [CrossRef]

Robinson, Stephen C. 2020. Trust, Transparency, and Openness: How Inclusion of Cultural Values Shapes Nordic National Public Policy Strategies for Artificial Intelligence (AI). *Technology in Society* 63: 101421. [CrossRef]

Rothstein, Bo. 2011. *The Quality of Government: Corruption, Social Trust and Inequality in International Perspective*. Chicago: Chicago University Press.

Rudin, Cynthia. 2019. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence* 1: 206–15. [CrossRef]

Shockley, Ellie, and Steven Shepherd. 2016. Compensatory Institutional Trust: A "Dark Side" of Trust. In *Interdisciplinary Perspectives on Trust*. Edited by Ellie Shockley, Tess Neal, Lisa PytlikZillig and Brian Bornstein. Cham: Springer. [CrossRef]

Sleep, Lyndal N. 2022. From Making Automated Decision Making Visible to Mapping the Unknowable Human: Counter-mapping Automated Decision Making in Social Services in Australia. *Qualitative Inquiry* 28: 848–58. [CrossRef]

Social Protection and Human Rights. 2015. Disadvantaged and Vulnerable Groups. Available online: https://socialprotection-humanrights.org/key-issues/disadvantaged-and-vulnerable-groups/ (accessed on 6 March 2025).

Spadaro, Giuliana, Katharina Gangl, Jan-Willem Van Prooijen, Paul A. M. Van Lange, and Christina O. Mosso. 2020. Enhancing Feelings of Security: How Institutional Trust Promotes Interpersonal Trust. *PLoS ONE* 15: e0237934. [CrossRef]

Standing, Guy. 2011. *The Precariat: The New Dangerous Class*. London: Bloomsbury.

Sun, Tara Q., and Rony Medaglia. 2019. Mapping The Challenges of Artificial Intelligence in the Public Sector: Evidence From Public Healthcare. *Government Information Quarterly* 36: 368–83. [CrossRef]

Tammpuu, Piia, and Anu Masso. 2018. 'Welcome to the Virtual State': Estonian e-residency and the Digitalised State as a Commodity. *European Journal of Cultural Studies* 21: 543–60. [CrossRef]

Veale, Michael, and Irina Brass. 2019. Administration by Algorithm? Public Management Meets Public Sector Machine Learning. In *Algorithmic Regulation*. Edited by Karen Yeung and Martin Lodge. Oxford: Oxford University Press. Available online: https://ssrn.com/abstract=3375391 (accessed on 6 March 2025).

Warren, Mark E. 1999. Introduction. In *Democracy and Trust*. Edited by Mark E. Warren. Cambridge: Cambridge University Press, pp. 1–21.

Widlak, Arjan, and Rik Peeters. 2020. Administrative Errors and the Burden of Correction and Consequence: How Information Technology Exacerbates the Consequences of Bureaucratic Mistakes for Citizens. *International Journal of Electronic Governance* 12: 40–56. [CrossRef]

Wirtz, Bernd, Jan C. Weyerer, and Carolin Geyer. 2019. Artificial Intelligence and the Public Sector—Applications and Challenges. *International Journal of Public Administration* 42: 596–615. [CrossRef]

Yang, Kaifeng, and Marc Holzer. 2006. The performance–trust link: Implications for performance measurement. *Public Administration Review* 66: 114–26. [CrossRef]

Yildiz, Mete. 2007. E-Government Research: Reviewing the Literature, Limitations, and Ways Forward. *Government Information Quarterly* 24: 646–65. [CrossRef]

Zak, Paul J., and Stephen Knack. 2001. Trust and Growth. *The Economic Journal* 111: 295–321. [CrossRef]

Zalnieriute, Monika, Lyria Bennet Moses, and George Williams. 2019. The Rule of Law and Automation of Government Decision-Making. *The Modern Law Review* 82: 425–55. [CrossRef]

Zouridis, Stavros, Marlies Van Eck, and Mark Bovens. 2020. Automated Discretion. In *Discretion and The Quest for Controlled Freedom*. Edited by Tony Evans and Peter Hupe. London: Palgrave Macmillan.