DOI: 10.5604/01.3001.0055.9068

Artificial Intelligence in Social Insurance: Administrative Promise, Legal Risk, and the Emerging Threat of Deepfake Fraud

Dmitry Erokhin [A,B,D,E,F]

ORCID: 0000-0002-5191-0579

International Institute for Applied Systems Analysis, Austria

ABSTRACT:

Introduction: Artificial intelligence is increasingly used in social protection administration for client service, back-office automation, fraud detection, and resource planning. However, its deployment also creates legal, organisational, and ethical risks, including biased assessments, opaque decision-making, exclusion of people in non-standard situations, and new forms of fraud involving deepfakes and synthetic identities.

Aim: This article aims to evaluate the potential of AI in social insurance and to identify the conditions under which it can enhance institutional capacity without compromising legality, procedural justice, and public trust.

Materials and methods: The study employs a narrative literature review combined with critical socio-legal analysis. It draws on research in public administration, social policy, law, AI ethics, the digital welfare state, fraud detection, deepfakes, and EU regulation, particularly Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), as amended by Regulation (EU) 2026/1744 (Digital Omnibus on AI).

Results: AI can enhance communication, case triage, document management, workload forecasting, and fraud identification. The most significant risks arise when systems influence entitlement, suspension, recovery, or fraud liability without adequate explainability, meaningful human oversight, and effective remedies.

KEYWORDS:

artificial intelligence; social insurance; automated decision-making; fraud detection; deepfakes

ORIGINAL ARTICLE

Received: 23.03.2026
Accepted: 18.05.2026
Published: 11.09.2026

CITE THIS ARTICLE AS:
Erokhin, Dmitry. „Artificial Intelligence in Social Insurance: Administrative Promise, Legal Risk, and the Emerging Threat of Deepfake Fraud". *Ubezpieczenia Społeczne. Teoria i Praktyka* 170, nr 3 (2026): x–xx. https://doi.org/10.5604/01.3001.0055.9068

AUTHORS CONTRIBUTION:
A – Research project
B – Data collection
C – Statistical analysis
D – Data interpretation
E – Manuscript preparation
F – Literature analysis
G – Fundraising
H – Coordination

CORRESPONDING AUTHOR:
International Institute for Applied Systems Analysis
Schlossplatz 1,
2361 Laxenburg, Austria
e-mail: erokhin@iiasa.ac.at

OPEN ACCESS:
Ubezpieczenia Społeczne. Teoria i Praktyka is an open access journal, which means all content is freely available to the user or their institution.

CREATIVE COMMONS CC BY:
This article is licensed under **CC BY-NC 4.0**. You may copy, modify, and distribute it for **non-commercial purposes** only, provided you **attribute the author** and do **not use it for financial gain**.

DOI: 10.5604/01.3001.0055.9068

Sztuczna inteligencja w ubezpieczeniach społecznych: obietnica administracyjna, ryzyko prawne i narastające zagrożenie oszustwami deepfake

ARTYKUŁ ORYGINALNY

Otrzymano: 23.03.2026
Zaakceptowano: 18.05.2026
Opublikowano: 11.09.2026

JAK CYTOWAĆ:
Erokhin, Dmitry. „Artificial Intelligence in Social Insurance: Administrative Promise, Legal Risk, and the Emerging Threat of Deepfake Fraud". *Ubezpieczenia Społeczne. Teoria i Praktyka* 170, nr 3 (2026): x–xx. https://doi.org/10.5604/01.3001.0055.9068

WKŁAD AUTORÓW:
A – Projekt badań
B – Zbieranie danych
C – Analiza statystyczna
D – Interpretacja danych
E – Przygotowanie manuskryptu
F – Analiza literatury
G – Zbieranie funduszy
H – Koordynacja

AUTOR KORESPONDENCYJNY:
Międzynarodowy Instytut Stosowanej Analizy Systemów
Schlossplatz 1,
2361 Laxenburg, Austria
e-mail: erokhin@iiasa.ac.at

OPEN ACCESS:
Ubezpieczenia Społeczne. Teoria i Praktyka jest czasopismem o otwartym dostępie, co oznacza, że wszystkie treści są bezpłatnie dostępne dla użytkownika lub jego instytucji.

LICENCJE:
Artykuł udostępniony na licencji **CC BY-NC 4.0**. Oznacza to, że można go kopiować, modyfikować i rozpowszechniać wyłącznie w celach **niekomercyjnych**, pod warunkiem **uznania autorstwa i nieuzyskiwania korzyści finansowych** z wykorzystania.

ABSTRAKT:

Wstęp: Sztuczna inteligencja wspiera w administracji zabezpieczenia społecznego obsługę klientów, procesy zaplecza, wykrywanie nadużyć i planowanie zasobów. Wdrożenia rodzą ryzyka prawne, organizacyjne i etyczne: stronniczość, nieprzejrzystość, wykluczenie osób w niestandardowych sytuacjach oraz oszustwa oparte na deepfake'ach i syntetycznej tożsamości.
Cel: Celem artykułu jest ocena potencjału AI w ubezpieczeniach społecznych i warunków jej wykorzystania bez naruszenia legalności, sprawiedliwości proceduralnej ani zaufania publicznego.
Materiały i metody: W badaniu zastosowano narracyjny przegląd literatury połączony z krytyczną analizą społeczno-prawną. Analiza obejmuje badania z administracji publicznej, polityki społecznej, prawa, etyki sztucznej inteligencji, cyfrowego państwa opiekuńczego, wykrywania oszustw i technologii deepfake oraz regulacje Unii Europejskiej, w szczególności rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji (akt w sprawie sztucznej inteligencji), zmienione rozporządzeniem (UE) 2026/1744 (cyfrowy omnibus dotyczący AI).
Wyniki: AI może usprawnić komunikację, triage spraw, zarządzanie dokumentami, prognozowanie obciążeń i identyfikację nadużyć. Największe ryzyka pojawiają się, gdy systemy wpływają na prawo do świadczeń, ich zawieszenie, zwrot lub odpowiedzialność za nadużycia bez wyjaśnialności, realnego nadzoru człowieka i skutecznych środków odwoławczych.

SŁOWA KLUCZOWE:

sztuczna inteligencja; ubezpieczenia społeczne; zautomatyzowane podejmowanie decyzji; wykrywanie oszustw; deepfake'i

Introduction

Artificial intelligence has become a focal point for social protection institutions, which face a familiar combination of pressures, including high caseloads, complex rules, fragmented data, rising expectations for digital services, and ongoing demands to prevent errors and fraud without delaying lawful payments. Governments are already using AI in social programmes for client support, back-office automation, and fraud detection, while exploring broader applications such as demand forecasting, early identification of vulnerable individuals, improved outreach, and reduced non-take-up.[1] For instance, in the context of minimum-income benefits and unemployment assistance, AI can enhance timeliness and uptake; however, it may also lead to systematically biased eligibility assessments and exacerbate inequalities if poorly designed or governed.[2]

This duality is structural rather than incidental. The adoption of AI in public administration is driven by the promise of personalised, efficient services, yet it is simultaneously riddled with tensions surrounding fairness, transparency, privacy, and human rights.[3] Automated decision-making (ADM) in public administration consistently creates tensions between efficiency and fairness, transparency and privacy, as well as standardisation and discretion.[4] This is particularly significant in social insurance, where decisions impact income continuity, access to health-related benefits, pension rights, disability claims, family support, and the citizen's daily relationship with the state.

In this context, this study examines the role of artificial intelligence in social insurance from both administrative and regulatory perspectives. Its aim is to assess how AI can enhance the functioning of social insurance institutions, particularly through faster processing, improved service delivery, more effective fraud detection, and better allocation of administrative resources, while also identifying the associated risks, including bias, opacity, exclusion, surveillance, and new forms of technologically enabled fraud such as deepfakes and synthetic identity abuse. By drawing on literature related to public administration, welfare governance, digital government, and AI ethics, the article develops a critical review of the opportunities and dangers of AI in social insurance, arguing that its adoption should be evaluated not only in terms of efficiency but also in light of legality, accountability, fairness, and public trust.

The review is guided by four research questions: (1) In which social-insurance functions can AI plausibly enhance administrative capacity and user experience? (2) What

1 OECD, *AI and the Future of Social Protection in OECD Countries* (OECD Artificial Intelligence Papers, no. 42, OECD Publishing, 2025), 6–8, https://doi.org/10.1787/7b245f7e-en.

2 Annelore Verhagen, *Using AI to Manage Minimum Income Benefits and Unemployment Assistance: Opportunities, Risks and Possible Policy Directions* (OECD Artificial Intelligence Papers, no. 21, OECD Publishing, 2024), 9, https://doi.org/10.1787/718c93a1-en.

3 Rohit Madan and Mona Ashok, „AI Adoption and Diffusion in Public Administration: A Systematic Literature Review and Future Research Agenda," *Government Information Quarterly* 40, no. 1 (2023): 101774, esp. 1, https://doi.org/10.1016/j.giq.2022.101774.

4 Aya Rizk and Ida Lindgren, „Automated Decision-Making in Public Administration: Changing the Decision Space between Public Officials and Citizens," *Government Information Quarterly* 42, no. 3 (2025): 102061, esp. 1, https://doi.org/10.1016/j.giq.2025.102061.

legal, ethical, and administrative harms are associated with AI-enabled welfare and social protection systems? (3) How does generative AI, particularly deepfakes and synthetic identities, alter the fraud and identity assurance landscape for social insurance? (4) What governance safeguards should social insurance institutions adopt to preserve legality, contestability, and public trust? The central research gap addressed here is that the literature on algorithmic welfare has primarily examined eligibility decisions, profiling, fraud scoring, and surveillance, while the interaction between these systems and generative AI-enabled fraud involving identity, evidentiary, and account recovery processes remains comparatively underdeveloped.

Conceptually, this article does not equate artificial intelligence with all ADM. AI is treated as a narrower set of systems that classify, predict, recommend, generate, or detect patterns through statistical, machine learning, symbolic, or generative techniques. ADM is the broader administrative category and includes rule-based workflows, robotic process automation, scoring tools, digital eligibility guidance, and other computational systems that inform or determine public action. Since social insurance agencies often combine these tools in a single workflow, the review analyses AI within the wider ADM ecology, while distinguishing rule-based automation, predictive scoring, and generative AI where their functions and risks differ.

Methodological approach

This article employs a narrative literature review design combined with critical socio-legal analysis.[5] A narrative review is appropriate when a topic is interdisciplinary, conceptually heterogeneous, and policy-facing, and where the purpose is to synthesise bodies of evidence, identify tensions, and develop an interpretive argument rather than to estimate a single effect size.

The review was developed through targeted searches and iterative snowballing across legal, public administration, social policy, and digital government sources. Searches combined terms such as 'artificial intelligence', 'automated decision-making', 'social insurance', 'social security', 'social protection', 'welfare administration', 'benefits', 'fraud detection', 'algorithmic discrimination', 'administrative burden', 'deepfake', 'synthetic identity', 'identity proofing', and 'EU AI Act'. To capture the debates on digital and datafied welfare, the search strategy included 'digital welfare state', 'datafied welfare state', 'data justice', 'human oversight', 'automation bias', 'participatory design', 'democratic oversight', 'Polish public employment services', 'frontline discretion', and 'deprofessionalisation'. The main scholarly sources consulted included Scopus, Web of Science, Google Scholar, ScienceDirect, SpringerLink, Taylor & Francis, Oxford Academic, and HeinOnline, supplemented

5 Trisha Greenhalgh, Sally Thorne, and Kirsti Malterud, „Time to Challenge the Spurious Hierarchy of Systematic over Narrative Reviews?" *European Journal of Clinical Investigation* 48, no. 6 (2018): e12931, esp. 1–3, https://doi.org/10.1111/eci.12931; Hannah Snyder, „Literature Review as a Research Methodology: An Overview and Guidelines," *Journal of Business Research* 104 (2019): 333–339, esp. 333–335, https://doi.org/10.1016/j.jbusres.2019.07.039.

by authoritative policy and legal sources including OECD iLibrary, EUR-Lex, the European Commission AI Act Service Desk, NIST, the U.S. Government Accountability Office, FinCEN, Europol, and GOV.UK algorithmic transparency records.

The core time window was 2019–2026 because the article focuses on contemporary AI deployment, recent welfare administration case studies, the EU AI Act, and the emergence of generative AI fraud. Earlier sources were retained where they were foundational for understanding longer-running debates on public administration automation and digital welfare. Included sources had to meet at least one of four criteria: 1. direct empirical relevance to AI or automation in social protection, welfare, employment services, or social security; 2. legal or regulatory relevance to automated public decision-making; 3. methodological relevance to narrative review design; or 4. technical and policy relevance to deepfakes, identity proofing, or synthetic identity fraud. The article therefore brings together evidence from adjacent but relevant domains to clarify what social insurance institutions can responsibly infer, what they cannot infer, and what kind of governance is required before AI tools are scaled in high-stakes benefit systems.

The reference list is the selected corpus ultimately cited and relied upon in the synthesis. Sources were retained where they materially strengthened the article's research questions, corrected the balance of the review, or clarified the relationship between social insurance, adjacent welfare administration evidence, and broader ADM/AI governance debates.

The term social insurance is used here in a strict institutional sense to refer primarily to schemes connected to insured status, contribution records, qualifying periods, wage histories, employment contingencies, disability or sickness status, pensions, and related statutory entitlements. Some empirical evidence reviewed below comes from social assistance, minimum-income benefits, public employment services, and child benefit automation. Those systems are not identical to contributory social insurance. They are used by analogy only where the administrative mechanisms are comparable, such as intake, verification, eligibility screening, fraud control, risk scoring, digital communication, and appealability. The transfer is strongest where the same institutional features are present, including high-volume processing, vulnerable claimants, complex rules, reliance on administrative data, and asymmetric power between the agency and the citizen.

Where AI can genuinely help social insurance

The strongest case for AI in social insurance is not that it can replace human judgment, but that it can reduce friction in processes that are repetitive, document-heavy, and time-sensitive. In the area of social protection, the current focus is on client support and administrative automation, rather than fully autonomous adjudication.[6] This is the appropriate starting point. In benefits systems, many delays do not stem from legal ambiguity but from incomplete forms, repeated requests for the same documents, poor case routing,

6 OECD, *AI and the Future of Social Protection*, 10.

fragmented records, and insufficient communication with applicants. AI tools can help address these bottlenecks.

Digital guidance tools can assist applicants navigate complex forms, answering questions about required documentation, and supporting the correct completion of applications. Some of these tools are genuinely AI-enabled, such as conversational agents that utilise natural language processing. Others are simpler rule-based or deterministic ADM tools. One illustrative non-AI example is a Canadian simulator whose validated success rate reached 85%, compared with 35% for the pre-existing payment tables interface.[7] This example is evidence of administrative simplification and user guidance. For a social insurance institution, both categories can be valuable, but they should not be conflated. A rule-based simulator, a robotic process automation script, a predictive model, and a generative chatbot perform different tasks and require different controls.

AI can also enhance upstream administrative planning. Predictive analytics is increasingly used to forecast demand and shocks, identify vulnerable clients earlier, and support more targeted outreach to reduce non-take-up.[8] In social insurance, this is particularly attractive for institutions that must anticipate spikes in sickness claims, regional unemployment shocks, demographic changes in pensions, or sudden surges in demand following legislative amendments. When used carefully, predictive models can improve staffing, triage, and service readiness before queues become politically and socially costly.

Another significant advantage is the ability to support consistency in routine case handling. Automation is associated with increased accountability, lower costs, and higher efficiency, especially where institutions aim to enhance citizen-centricity.[9] The point is not that automation is inherently more just than human administration, but that it can reduce certain forms of arbitrariness in repetitive tasks. AI can help improve access to benefits and ensure that the right benefits reach the right people at the right time, provided institutions invest in governance, data quality, and workforce readiness.[10]

In contributory social insurance, these benefits are most defensible when AI supports administration rather than substitutes for legal entitlement judgments, such as matching contribution histories, detecting missing wage records, routing sickness or disability claims to appropriate expertise, identifying incomplete evidence, translating complex rules into accessible guidance, and forecasting caseload pressure. The closer an AI system moves from administrative support to making decisions about entitlement, benefit amounts, suspension, recovery, or fraud liability, the stronger the justification, documentation, human review, and appeal safeguards must become.

These benefits are real. However, the literature clearly indicates that they are conditional. Administrative speed is not synonymous with administrative justice, and consistency does not equate to fairness.

7 Verhagen, *Using AI to Manage Minimum Income Benefits*, 22.

8 OECD, *AI and the Future of Social Protection*, 6–8.

9 Agneta Ranerup and Helle Zinner Henriksen, „Value Positions Viewed Through the Lens of Automated Decision-Making: The Case of Social Services," *Government Information Quarterly* 36, no. 4 (2019): 101377, esp. 1–3, https://doi.org/10.1016/j.giq.2019.05.004.

10 OECD, *Harnessing Artificial Intelligence in Social Security: Use Cases, Governance and Workforce Readiness* (OECD Digital Government Studies, OECD Publishing, 2025), 4, 9–10, https://doi.org/10.1787/b52405c1-en.

Why the same technologies can also be harmful

The critical literature on digital welfare shows that AI-related harms seldom originate in a single, readily identifiable technical failure. They can accumulate through routine administrative choices concerning data, eligibility rules, performance targets, procurement, and system implementation. This matters because a system may meet aggregate accuracy or efficiency targets while distributing errors and administrative burdens unevenly. For example, a fraud-detection model may perform well overall while repeatedly assigning higher risk to claimants with interrupted employment histories, incomplete contribution records, non-standard households, disabilities, migration backgrounds, or limited proficiency in the administrative language. Institutional incentives can reinforce these effects when staff are evaluated primarily on processing speed or fraud detection and have insufficient time, information, or authority to question automated classifications.

Social inequality can enter an automated decision pipeline at three stages: data generation, data analysis, and implementation.[11] During data generation, administrative records may reproduce earlier inequalities in labour-market participation, disability recognition, household classification, language access, and interactions with public authorities. During data analysis, decisions about outcome labels, proxy variables, thresholds, and missing values can transform these historical differences into predictions of eligibility, compliance, or fraud risk. During implementation, interface design, workload pressures, case-routing rules, and appeal procedures determine whether an automated output is critically reviewed or effectively treated as a decision. For social insurance institutions, fairness assessment must therefore extend beyond model accuracy to include data provenance, subgroup error patterns, threshold selection, staff override practices, appeal outcomes, and the distribution of errors across claimant groups.

This point is reinforced by the literature on the digital welfare state and the datafied welfare state. It cautions that harms are not only consequences of model design, poor service delivery, or technical error. They may also reflect political and managerial logics that privilege efficiency, surveillance, fraud control, and fiscal discipline over universalism, equality, social rights, and care.[12] For social insurance, this means that AI/ADM governance cannot be reduced to accuracy testing. It must also consider what policy purpose

11 Frederic Gerdon, Ruben L. Bach, Christoph Kern, and Frauke Kreuter, „Social Impacts of Algorithmic Decision-Making: A Research Agenda for the Social Sciences," *Big Data & Society* 9, no. 1 (2022): 20539517221089305, esp. 2–3, https://doi.org/10.1177/20539517221089305.

12 Philip Alston, „Digital Welfare States and Human Rights: Report of the Special Rapporteur on Extreme Poverty and Human Rights," United Nations General Assembly, A/74/493 (2019), 2–5, accessed August 21, 2026, https://www.ohchr.org/en/documents/thematic-reports/a74493-digital-welfare-states-and-human-rights-report-special-rapporteur; Lina Dencik and Anne Kaun, „Datafication and the Welfare State," *Global Perspectives* 1, no. 1 (2020): 12912, esp. 1–3, https://doi.org/10.1525/gp.2020.12912; Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St. Martin's Press, 2018), 1–14; Jędrzej Niklas, *Social Rights and Data Technologies: Looking for Connections* (Data Justice Lab, Cardiff University, 2022), 3–8, accessed August 21, 2026, https://datajusticeproject.net/wp-content/uploads/2022/01/WP_Social-rights.pdf; Irina Zakharova, Juliane Jarke, and Anne Kaun, „Tensions in Digital Welfare States: Three Perspectives on Care and Control," *Journal of Sociology* 60, no. 3 (2024): 540–559, esp. 542–548, https://doi.org/10.1177/14407833241238312.

the system serves, whose burden is reduced, whose conduct becomes more visible, and whether the technology reinforces a presumption of suspicion toward claimants.

Empirical welfare studies illustrate how these dynamics play out even when systems remain at the pilot, design, or pre-implementation stage. In Austria, the AMS algorithm was developed to profile jobseekers into categories based on predicted labour market prospects, with the stated aim of improving counselling efficiency and the allocation of labour market programmes.[13] The system was piloted but not introduced nationwide after regulatory and judicial challenges. The controversy remains significant because it demonstrated how an ostensibly technical profiling instrument can encode austerity-oriented assumptions and make differentiated access to support appear administratively neutral. Systems introduced in the name of efficient resource allocation may therefore stratify support and entrench prior disadvantage even before they become fully normalised in everyday administration.[14]

The Polish public employment service experience provides an especially relevant example for a Polish social insurance readership. The profiling tool used in Poland between 2014 and 2019 classified unemployed individuals into assistance profiles that shaped access to active labour market measures.[15] Subsequent analysis shows that such tools do not merely execute law but can participate in the 'algorithmic articulation' of law by translating open-textured legal and policy categories into operational variables, thresholds, and backstage technical choices.[16] This example strengthens the argument for social insurance because contribution records, medical evidence, occupational histories, and fraud indicators can likewise be converted into technical categories that silently structure administrative treatment.

The Swedish literature on automated social assistance adds further nuance. A study of 800 applications across four municipalities revealed that the use of robotic process automation correlated with applicants' country of birth, age, and duration of receipt and coincided with less generous decisions that disproportionately affected financially vulnerable groups.[17] This is particularly important for social insurance institutions. Even when the technology itself is rules-based rather than intelligent in a narrow sense, organisational reconfiguration around automation may compromise individualised judgment in favour of

13 Doris Allhutter, Florian Cech, Fabian Fischer, Gabriel Grill, and Astrid Mager, „Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective," *Frontiers in Big Data* 3 (2020): article 5, esp. 2–9, https://doi.org/10.3389/fdata.2020.00005.

14 Doris Allhutter, Anila Alushi, Rafaela Cavalcanti de Alcântara, Maris Männiste, Christian Pentzold, and Sebastian Sosnowski, „Public Value in the Making of Automated and Datafied Welfare Futures," *Internet Policy Review* 13, no. 3 (2024): 1–22, esp. 1–2, https://doi.org/10.14763/2024.3.1803.

15 Jędrzej Niklas, Karolina Sztandar-Sztanderska, and Katarzyna Szymielewicz, *Profiling the Unemployed in Poland: Social and Political Implications of Algorithmic Decision Making* (Fundacja Panoptykon, 2015), 5, 13–16, accessed August 21, 2026, https://panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon_profiling_report_final.pdf; Karolina Sztandar-Sztanderska and Marianna Zieleńska, „What Makes an Ideal Unemployed Person? Values and Norms Encapsulated in a Computerized Profiling Tool," *Social Work & Society* 18, no. 1 (2020): 1–16, esp. 1–5, accessed August 21, 2026, https://ejournals.bib.uni-wuppertal.de/index.php/sws/article/view/638.

16 Karolina Sztandar-Sztanderska and Joanna Mazur, „Politics and Algorithmic Articulation of Law: Tracing the Discrepancies and Backstage Decision-Making in the Development of the Profiling Algorithm in Polish Labour Market Policies," *International Journal of Law in Context* 21, no. 4 (2025): 749–768, esp. 753–762, https://doi.org/10.1017/S1744552325100141.

17 Nora Germundsson and Hugo Stranz, „Automating Social Assistance: Exploring the Use of Robotic Process Automation in the Swedish Personal Social Services," *International Journal of Social Welfare* 33, no. 3 (2024): 647–658, esp. 653–656, https://doi.org/10.1111/ijsw.12633.

efficiency metrics. A case study of two municipalities reinforces this point.[18] It found that ADM reconfigures discretion rather than eliminating it. In practice, discretion migrates into system design, variable selection, threshold setting, and exception handling. This shift can make power less visible and therefore harder to contest. The normative stakes are well captured when automated and datafied welfare are assessed in terms of the public value they create and the risks they distribute across groups of citizens.[19]

The transfer from social assistance to social insurance should be approached cautiously but not dismissed. Social insurance institutions may rely less on discretionary means testing, but they still make consequential judgments about work capacity, contribution gaps, occupational status, dependency, residence, medical evidence, and fraud. In these contexts, algorithmic tools can reproduce disadvantages through historical contribution patterns, fragmented employer reporting, unequal access to medical documentation, language barriers, informal care histories, or disability classifications.

Administrative burden, exclusion, and the 'some-stop shop'

A common misconception in digital government debates is the assumption that automation automatically reduces burdens on users. The evidence does not support this assumption. In some cases, automation simplifies access for standard cases while making it more difficult for those whose lives do not fit the data model.

A process-mining study of 50,463 Norwegian child-benefit cases provides a clear illustration.[20] The automatic system generated 34,369 claims (68.1%), while 16,094 cases (31.9%) began with a manual application. Of the full sample, 32,682 cases (64.8%) were completed entirely automatically without requiring work from the recipient. Manual claims took an average of 35.9 days, compared with 29.9 days for automatically generated claims, and manual applicants were nine times as likely to undergo a review. Groups outside the automated pathway included parents or children not registered as residents, separated parents or parents with children from previous relationships, and caregivers who were not the child's natural parent. Across 416 municipalities and 34 urban boroughs, automatic awards were less common in areas with a higher proportion of households living below 60% of the national median income, although the study did not establish a causal relationship. These findings reveal a distributional problem within an otherwise universal benefit. Automation simplified access for recipients whose circumstances matched standard registry categories, while other households (disproportionately those in lower-income areas) had to provide additional information and navigate manual assessment. As government services

18 Agneta Ranerup and Lupita Svensson, „Automated Decision-Making, Discretion and Public Values: A Case Study of Two Municipalities and Their Case Management of Social Assistance," *European Journal of Social Work* 26, no. 5 (2023): 948–962, esp. 952–958, https://doi.org/10.1080/13691457.2023.2185875.

19 Allhutter et al., „Public Value," 17–18.

20 Karl Kristian Larsson, „Digitization or Equality: When Government Automation Covers Some, but Not All Citizens," *Government Information Quarterly* 38, no. 1 (2021): 101547, esp. 6–8, https://doi.org/10.1016/j.giq.2020.101547.

move towards a 'no-stop shop' for standard cases, citizens with less typical circumstances may encounter a 'some-stop shop' characterized by additional work, review, and delay.

The term 'some-stop shop' is used here to emphasise that automation may create a seamless pathway for standard, data-complete cases while leaving non-standard cases to navigate fragmented, manual, slower, or less intelligible procedures. In social insurance, such cases may include cross-border individuals, workers with interrupted careers, people with multiple employers, those with partial disabilities or fluctuating conditions, carers with incomplete contribution histories, migrants, and individuals whose evidence is held outside interoperable administrative databases.

The same issue appears in the UK literature on universal credit.[21] The classic framework of administrative burden remains highly relevant in automated welfare but must be expanded to include a technical dimension. Automation can intensify learning costs, compliance costs, and emotional costs, particularly for people already facing economic precarity. In other words, a fast system can still be a punishing system if its logic is obscure, its interfaces brittle, or its exceptions poorly designed.

This is one reason public trust cannot be inferred from institutional self-confidence. OECD survey evidence shows that across 27 countries, only 40% of respondents agreed that government use of AI to help process and approve social programme applications is beneficial for users.[22] A significant share were uncertain, and a quarter disagreed. This public hesitation is not irrational; it reflects an intuitive understanding that automated systems can lower transaction costs for institutions while shifting them onto citizens.

Opacity, legal certainty, and accountability

If social insurance decisions affect livelihoods, then citizens must be able to understand, question, and challenge them. This is where opacity becomes a legal and democratic issue, not merely a technical inconvenience. The lack of transparency and explainability can prevent individuals from understanding why benefits were denied or adjusted, from exercising their rights, and from obtaining effective redress.[23]

Transparent and explainable AI is important not as an end in itself, but because it supports fairness and legitimacy in public decision-making.[24] However, 'blanket' transparency is not a cure-all. Institutions still need to determine what kinds of explanations are owed, to whom, and for what purpose. This is highly relevant to social insurance, where useful explanations must often operate at several levels simultaneously. The claimant must understand the immediate reason for an outcome, reviewers must be able to audit the model's logic, and courts or oversight bodies must be able to reconstruct the decision pathway.

21 Hayley Bennett, Morgan Currie, and Lena Podoletz, „Universal Credit: Administrative Burdens of Automated Welfare," *Journal of Social Policy* 55, no. 1 (2026): 24–42, esp. 29–38, https://doi.org/10.1017/S0047279424000175.

22 OECD, *AI and the Future of Social Protection*, 8.

23 Verhagen, *Using AI to Manage Minimum Income Benefits*, 36.

24 Kate Vredenburgh, „Transparency and Explainability for Public Policy," *LSE Public Policy Review* 3, no. 3 (2024): article 4, 1–11, esp. 2–6, https://doi.org/10.31389/lseppr.111.

The rule-of-law stakes are not hypothetical. The analysis of the Dutch SyRI litigation shows that digital welfare initiatives can fail to meet basic requirements of legality and protection against arbitrariness.[25] The Hague District Court held that the legislation and use of the welfare fraud system did not satisfy Article 8 ECHR requirements of necessity and proportionality. More recent Dutch experiences have reinforced this warning. The Rotterdam welfare-fraud algorithm controversy and the Dutch childcare benefits scandal (*toeslagenaffaire*) exposed the harm that can arise when discriminatory indicators and opaque algorithms are embedded in administrative enforcement. In Rotterdam, Dutch-language proficiency became a proxy for fraud risk; in the childcare-benefits case, nationality was used in fraud determinations, contributing to structural discrimination and serious financial and social consequences for affected families.[26] Citizens do not grant public bodies a free pass simply because harm was produced by an algorithm rather than a human. Organisational responsibility remains.

There is also a subtler risk that accountability deficits can arise when algorithms are outsourced.[27] Citizens assign more responsibility when the algorithm is developed in-house than when it comes from a private vendor, suggesting that procurement itself may blur responsibility. For social insurance institutions, this is a significant warning. Core adjudicative or fraud-related functions should not be outsourced in a manner that leaves the institution unable to explain, monitor, or correct the system it deploys.

The accountability problem also predates contemporary machine learning. Classic public administration scholarship already warned that ICT can shift discretion from street-level officials to system designers and system-level bureaucracies.[28] Contemporary AI accountability literature makes this point more urgent. Algorithmic accountability requires institutional capacity to justify the system, monitor outcomes, assign responsibility, and create enforceable routes for challenge rather than relying solely on abstract transparency.[29]

Human oversight is necessary, but not sufficient

The standard policy response to these concerns is to assert that there will always be human oversight. However, the literature suggests that this answer is incomplete. When studying ADM and good administration, six values are at stake: carefulness, respect for individual

25 Adamantia Rachovitsa and Niclas Johann, „The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case," *Human Rights Law Review* 22, no. 2 (2022): ngac010, esp. 1, 10–15, https://doi.org/10.1093/hrlr/ngac010.

26 Saar Alon-Barkat, Madalina Busuioc, Kayla Schwoerer, and Kristina S. Weißmüller, „Algorithmic Discrimination in Public Service Provision: Understanding Citizens' Attribution of Responsibility for Human versus Algorithmic Discriminatory Outcomes," *Journal of Public Administration Research and Theory* 35, no. 4 (2025): 469–488, esp. 470–472, https://doi.org/10.1093/jopart/muaf024.

27 *ibidem*, 480–484.

28 Mark Bovens and Stavros Zouridis, „From Street-Level to System-Level Bureaucracies: How Information and Communication Technology Is Transforming Administrative Discretion and Constitutional Control," *Public Administration Review* 62, no. 2 (2002): 174–184, esp. 175–181, https://doi.org/10.1111/0033-3352.00168.

29 Madalina Busuioc, „Accountable Artificial Intelligence: Holding Algorithms to Account," *Public Administration Review* 81, no. 5 (2021): 825–836, esp. 828–833, https://doi.org/10.1111/puar.13293.

rights, professionalism, trustworthiness, responsiveness, and empowerment.[30] Automation can either support or undermine these values depending on institutional design and managerial attention. Human presence alone does not guarantee that these values will be protected.

The common opposition between fully automated and human decisions is misleading, as individuals who are formally in the loop may still defer to algorithmic outputs.[31] Human oversight can therefore become ritualised rather than effective. In a social insurance context, this means a caseworker may nominally retain authority while, in practice, giving decisive weight to a risk score, fraud flag, or predicted eligibility outcome that is difficult to interrogate under time pressure.

Experimental work on public sector decision-making confirms that human reviewers may over-rely on algorithmic advice or selectively adhere to it when it aligns with stereotypes.[32] Human oversight requirements may provide a false sense of security if they legitimise controversial systems without ensuring that humans can actually perform the expected oversight functions.[33]

Recent welfare administration studies further show that oversight is shaped by organisational context, workload, legal design, interface design, and managerial expectations rather than by individual vigilance alone.[34] This is significant for social insurance because caseworkers can become 'moral crumple zones'. They are formally responsible for validating AI-supported outputs while lacking sufficient transparency, time, authority, or organisational support to contest them.[35] A related risk is deprofessionalisation, where profiling tools and managerial narratives narrow the practical space for professional judgment even when frontline staff remain formally present.[36]

This is why trustworthy AI requires more than a human signature at the end of the workflow. Studies identify up to eighteen trustworthiness requirements for AI-enabled ADM in the public sector and emphasise context sensitivity and multiple forms of accountability.[37]

30 Ulrik B. U. Røhl, „Automated Decision-Making and Good Administration: Views from Inside the Government Machinery," *Government Information Quarterly* 40, no. 4 (2023): 101864, esp. 4–8, https://doi.org/10.1016/j.giq.2023.101864.

31 Hannah Ruschemeier and Lukas J. Hondrich, „Automation Bias in Public Administration—An Interdisciplinary Perspective from Law and Psychology," *Government Information Quarterly* 41, no. 3 (2024): 101953, esp. 2–7, https://doi.org/10.1016/j.giq.2024.101953.

32 Saar Alon-Barkat and Madalina Busuioc, „Human–AI Interactions in Public Sector Decision Making: 'Automation Bias' and 'Selective Adherence' to Algorithmic Advice," *Journal of Public Administration Research and Theory* 33, no. 1 (2023): 153–169, esp. 161–165, https://doi.org/10.1093/jopart/muac007.

33 Ben Green, „The Flaws of Policies Requiring Human Oversight of Government Algorithms," *Computer Law & Security Review* 45 (2022): 105681, esp. 4–8, https://doi.org/10.1016/j.clsr.2022.105681.

34 Karolina Sztandar-Sztanderska, „How Context Matters: Human Oversight of Automated Decision-Making Systems in Welfare Administration," *Journal of European Social Policy* 36, no. 3 (2026): 255–269, esp. 257–265, https://doi.org/10.1177/09589287251358069.

35 Marta Choroszewicz, „Positioning Public Sector Practitioners as 'Moral Crumple Zones': Mechanisms in the Early Use of Generative AI Work Support Tools," *Government Information Quarterly* 43, no. 2 (2026): 102128, esp. 5–9, https://doi.org/10.1016/j.giq.2026.102128.

36 Alizée Delpierre, Didier Demazière, and Hajar El Fatihi, „The Stealth Legitimization of a Controversial Policy Tool: Statistical Profiling in French Public Employment Service," *Regulation & Governance* 18, no. 2 (2024): 499–512, esp. 506–509, https://doi.org/10.1111/rego.12541.

37 Olusegun Agbabiaka, Adegboyega Ojo, and Niall Connolly, „Requirements for Trustworthy AI-Enabled Automated Decision-Making in the Public Sector: A Systematic Review," *Technological Forecasting and Social Change* 215 (2025): 124076, esp. 8–11, https://doi.org/10.1016/j.techfore.2025.124076.

The OECD's social security report reaches the compatible conclusion that data quality, governance, and workforce readiness are not auxiliary issues but core preconditions for legitimate deployment.[38]

Deepfakes and the new fraud frontier

The literature on AI in social insurance has focused heavily on eligibility, profiling, and welfare surveillance. It has paid less attention to a newer but increasingly important threat of generative AI fraud. This gap is now significant. Deepfakes are not only a misinformation problem; they also pose challenges related to identity, authentication, and evidence.

A 2025 integrative review describes deepfakes as a serious threat to public trust, cybersecurity, and the reliability of image, audio, and video evidence.[39] The technology can create convincing false personas and has direct relevance for document fraud and broader criminal misuse.[40] In an alert issued in November 2024, FinCEN reported that suspicious-activity reports submitted by financial institutions describing suspected deepfake-enabled fraud had increased beginning in 2023 and continuing into 2024. The reported schemes included altered or fabricated identity documents used to circumvent identity-verification and authentication controls.[41]

For social insurance institutions, the implications are immediate. Many systems are transitioning towards remote onboarding, digital identity proofing, self-service portals, document uploads, account recovery, and virtual customer support. In this context, a deepfake is not merely a fake video; it can be a synthetic face used to pass biometric checks, a cloned voice used to impersonate a beneficiary or representative, an AI-generated identity document, or manipulated supporting evidence used to sustain a claim or redirect payments. The technical literature on biometric spoofing already indicates that deepfakes raise impersonation, identity theft, and scam risks, and that many detection models still lack strong generalisability across different deepfake generation methods.[42]

The implications for social insurance are therefore twofold. AI may assist institutions detect fraud, but it also arms fraudsters with better tools. This is not speculative. For example, the Social Security Administration in the United States has already had to invest

38 OECD, *Harnessing Artificial Intelligence in Social Security*, 9–10.

39 Sonam Singh and Amol Dhumane, „Unmasking Digital Deceptions: An Integrative Review of Deepfake Detection, Multimedia Forensics, and Cybersecurity Challenges," *MethodsX* 15 (2025): 103632, esp. 1–3, https://doi.org/10.1016/j.mex.2025.103632.

40 Europol, *Facing Reality? Law Enforcement and the Challenge of Deepfakes: An Observatory Report from the Europol Innovation Lab* (Publications Office of the European Union, 2022), 9–13, 17, https://doi.org/10.2813/08370.

41 U.S. Department of the Treasury, Financial Crimes Enforcement Network, „FinCEN Alert on Fraud Schemes Involving Deepfake Media Targeting Financial Institutions," FIN-2024-Alert004, November 13, 2024, 1–2, accessed August 21, 2026, https://www.fincen.gov/system/files/shared/FinCEN-Alert-DeepFakes-Alert508FINAL.pdf.

42 Hilary Zen, Rohan Wagh, Miguel Wanderley, Gustavo Bicalho, Rachel Park, Megan Sun, Rafael Palacios, Lucas Carvalho, Guilherme Rinaldo, and Amar Gupta, „Ensemble-Based Biometric Verification: Defending Against Multi-Strategy Deepfake Image Generation," *Computers* 14, no. 6 (2025): 225, esp. 2, 16–17, https://doi.org/10.3390/computers14060225.

in mechanisms designed to counter synthetic identity fraud.[43] In practical terms, this means the AI debate in social insurance cannot stop at adjudication models; it must also encompass identity infrastructure.

Responsible deployment under the EU AI Act and social insurance governance

A responsible model for AI in social insurance should begin with legal classification and institutional humility. Under Annex III, point 5(a), of the EU AI Act, AI systems used by or on behalf of public authorities to evaluate eligibility for essential public assistance benefits and services, including healthcare services, or to grant, reduce, revoke, or reclaim such benefits and services, are classified as high-risk. The Act entered into force on 1 August 2024 and became generally applicable on 2 August 2026. Following Regulation (EU) 2026/1744, however, Sections 1–3 of Chapter III apply to high-risk systems under Article 6(2) and Annex III from 2 December 2027, and to high-risk systems under Article 6(1) and Annex I from 2 August 2028.[44] Even where a particular social insurance use case falls outside the strict wording of that category, the regulatory logic is instructive. Systems that shape livelihoods, access to benefits, or effective remedies require stronger safeguards than ordinary productivity tools.

However, the qualification of a social insurance AI system under the EU AI Act will require case-specific legal analysis. A pension information chatbot, a staff-only document routing tool, a fraud risk model that triggers investigation, and a model used to recommend benefit suspension do not carry the same legal significance. Some tools may fall outside Annex III, some may be high-risk because they materially affect benefit access or recovery, and some may trigger additional obligations because they rely on biometric verification or sensitive data. The practical point is that institutions should classify use cases by function, legal effect, data type, autonomy, and impact on rights before procurement or deployment, rather than treating AI in social insurance as a single regulatory category. Responsible deployment should also include participatory and democratic oversight. Decisions about datafication should not be left solely to system developers, procurement teams,

43 U.S. Government Accountability Office, *Social Security Administration: Actions Needed to Help Ensure Success of Electronic Verification Service* (GAO-24–106770, September 10, 2024), 9–15, accessed August 21, 2026, https://www.gao.gov/products/gao-24–106770.

44 European Parliament and Council of the European Union, *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No. 300/2008, (EU) No. 167/2013, (EU) No. 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*, *Official Journal of the European Union* L, 2024/1689 (July 12, 2024), Annex III, point 5(a), and Article 113, first to third paragraphs, point (c), as amended, accessed August 21, 2026, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX-%3A02024R1689–20260727; European Parliament and Council of the European Union, *Regulation (EU) 2026/1744 of the European Parliament and of the Council of 8 July 2026 amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI)*, *Official Journal of the European Union* L, 2026/1744 (July 24, 2026), accessed August 21, 2026, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AL_202601744.

expert auditors, or legal compliance units. Affected citizens, civil society organisations, claimant representatives, and frontline staff are often best placed to identify foreseeable harms, administrative burdens, stigma, and exclusion that may not be visible in technical validation.[45] For social insurance, this means that impact assessment should be a public governance process as well as a technical exercise.

The EU AI Act obligations summarised in Table 1 show that responsible AI governance in social insurance must extend across the entire system lifecycle, from pre-deployment risk assessment and data governance to human oversight, redress, cybersecurity, and post-deployment monitoring.

The first governance principle is functional restraint. High-stakes uses should remain decision-support tools unless institutions can demonstrate that automation improves, rather than undermines, legality, equality, and appealability. The second principle is data governance. High-quality, relevant, interoperable data is a prerequisite for trustworthy AI.[46] The third principle is workforce design. Social security institutions need recruitment, training, and organisational development strategies that align AI deployment with institutional goals rather than treating technology as a standalone procurement exercise.

The fourth principle is meaningful contestability. Claimants must be informed when AI materially shapes a process, what the main reasons for the outcome were, how to challenge it, and how a human reviewer can deviate from the system recommendation. This is necessary not only for legal certainty but also for institutional trust.[47] The fifth principle is continuous auditing. Institutions should monitor false positives, false negatives, error concentration across protected and vulnerable groups, and the real-world consequences of system outputs rather than merely model accuracy in the abstract. Fairness must be evaluated in relation to concrete social outcomes and contexts, not solely computational metrics.[48]

The sixth principle concerns procurement and accountability. Public institutions should retain the ability to inspect system logic, test outputs, demand documentation, and halt deployment when harms arise. Otherwise, outsourcing can generate accountability deficits even when formal responsibility remains public.[49] The seventh principle is security against generative fraud. Official guidance now advocates for layered identity proofing and liveness controls rather than reliance on a single verification signal. For example, GOV.UK's One Login record describes the use of facial matching and liveness detection to verify that a remote user both matches the image on a trusted ID document and is genuinely present in real time.[50] NIST's current digital identity guidance likewise frames

45 Allhutter et al., „Public Value," 17–18; Joanna Redden, Jessica Brand, Ina Sander, and Harry Warne, *Automating Public Services: Learning from Cancelled Systems* (Carnegie UK and Data Justice Lab, 2022), 13–16, accessed August 21, 2026, https://carnegieuk.org/publication/automating-public-services-learning-from-cancelled-systems/; Kristen M. Scott et al., „Algorithmic Tools in Public Employment Services: Towards a Jobseeker-Centric Perspective," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery, 2022), 2138–2148, esp. 2144–2147, https://doi.org/10.1145/3531146.3534631.

46 OECD, *Harnessing Artificial Intelligence in Social Security*, 37.

47 Verhagen, *Using AI to Manage Minimum Income Benefits*, 40.

48 Gerdon et al., „Social Impacts of Algorithmic Decision-Making," 5–9.

49 Alon-Barkat et al., „Algorithmic Discrimination in Public Service Provision," 480–484.

50 Cabinet Office, Department for Science, Innovation and Technology, and Government Digital Service, „DSIT: GOV.UK One Login—Liveness & Likeness Checks," *GOV.UK*, published December 16, 2025, accessed August 21, 2026, https://www.gov.uk/algorithmic-transparency-records/dsit-gov-dot-uk-one-login-liveness-and-likeness-checks.

identity proofing as a multi-step process of resolution, validation, and verification, explicitly linking fraud mitigation to preventing access to benefits and services through fraudulent identities.[51]

Table 1. Mapping key EU AI Act obligations to the governance principles proposed in this article

EU AI Act obligation	Social insurance relevance	Governance implication
Risk management (Article 9)	Benefit systems can produce foreseeable risks to fundamental rights when eligibility, suspension, recovery, or fraud flags affect livelihoods.	Conduct pre-deployment risk analysis, define acceptable residual risk, test against intended and foreseeable misuse, and update controls throughout the lifecycle.
Data and data governance (Article 10)	Historical administrative records may encode labour market inequality, disability-recognition gaps, migration status, and uneven access to documentation.	Assess data provenance, relevance, representativeness, bias, gaps, and proxy variables before model training or rule automation.
Technical documentation and logging (Articles 11–12)	Claimants, internal reviewers, auditors, and courts need to reconstruct why a system generated a recommendation or flag.	Maintain auditable documentation, decision logs, version histories, and traceability for every high-stakes system output.
Transparency and instructions for use (Article 13)	Caseworkers must understand system capabilities, limits, intended purpose, and risks before relying on outputs.	Translate technical documentation into operational guidance that explains when outputs may be used, questioned, overridden, or escalated.
Human oversight (Article 14)	Formal human-in-the-loop arrangements can fail if staff defer to scores under workload pressure.	Design oversight roles with authority, time, training, override powers, and monitoring for automation bias.
Accuracy, robustness, and cybersecurity (Article 15)	Incorrect or vulnerable systems can replicate errors at scale and expose identity infrastructure to manipulation.	Set performance thresholds by use case, test robustness across groups and operational contexts, and integrate fraud-resistant security controls.
Deployer obligations and fundamental rights impact assessment (Articles 26–27)	Public bodies deploying high-risk AI must understand affected groups, foreseeable harms, oversight arrangements, and complaint mechanisms.	Adopt a documented pre-deployment review that combines legal, technical, operational, and rights-based assessments.
Post-market monitoring and serious incident reporting (Articles 72–73)	A model that performs acceptably in pilots may fail after legal, economic, demographic, or fraud-pattern changes.	Monitor false positives, false negatives, group-level impacts, appeals, complaints, staff overrides, and incidents after deployment.
Complaint and explanation rights (Articles 85–86)	Affected persons need meaningful redress when AI outputs contribute to decisions with legal or similarly significant effects.	Provide clear notices, reasons, challenge routes, human review, and explanations of the AI system's role in the decision path.

51 David Temoshok, Christine Abruzzi, Yee-Yin Choong, James L. Fenton, Ryan Galluzzo, Connie LaSalle, Naomi Lefkovitz, Andrew Regenscheid, and Maria Vachino, *Digital Identity Guidelines: Identity Proofing and Enrollment* (NIST Special Publication 800–63A-4, National Institute of Standards and Technology, 2025), 1–3, 11–16, https://doi.org/10.6028/NIST.SP.800–63A-4.

Limitations

This review uses a narrative and critical design. This approach supports the integration of evidence from several disciplines and permits close attention to legal and institutional context, but it does not employ the exhaustive search, selection, and quality-appraisal procedures associated with a systematic review. Relevant publications may therefore have been omitted, while the relative weight assigned to individual sources depends partly on the author's analytical judgement. Future research could address this limitation through a systematic or scoping review based on a preregistered protocol, transparent eligibility criteria, formal quality assessment, and reproducible search procedures.

A second limitation concerns the available empirical evidence. Direct studies of AI applications in contributory social insurance remain less extensive than research on social assistance, minimum-income benefits, unemployment services, and broader welfare administration. Some conclusions consequently rely on evidence transferred from institutionally comparable settings. Differences in financing arrangements, entitlement rules, administrative data, claimant populations, and appeal procedures may limit the applicability of these findings to contributory schemes. Future studies should examine AI applications within pension, sickness, disability, and unemployment insurance using administrative records, case files, interviews with claimants and practitioners, and comparative research across countries and institutional settings.

The evidence concerning deepfakes and synthetic identity fraud presents a further limitation. The analysis draws partly on financial-sector fraud alerts and law-enforcement assessments because documented incidents within social insurance agencies are not yet supported by a substantial public evidence base. These sources identify plausible fraud mechanisms, but they cannot establish the prevalence, scale, or consequences of deepfake-enabled fraud in social insurance. Future research should combine agency case studies with standardised incident reporting, identity-proofing audits, controlled security assessments, and comparative analysis of verification failures across benefit systems.

Conclusion

The central lesson from the reviewed literature is straightforward. AI can enhance social insurance administration, but it can also render it harsher, less intelligible, and more exclusionary. The same tools that accelerate case handling can also compress discretion. The same models that identify anomalous claims can also misclassify vulnerable citizens. The same digital channels that expand access can also intensify burdens for those who do not conform to the system's assumptions.

For this reason, social insurance institutions should pursue institutionally disciplined augmentation. The term refers here to the use of AI to extend administrative capacity

within clearly defined legal, procedural, and organisational boundaries. AI may support analysis and routine processing, while public officials retain responsibility for consequential decisions, explanations, review, and correction. In a well-governed social insurance system, AI should reduce avoidable delays, improve communication, support staff in routine processing, strengthen forecasting, and help identify genuine fraud. It should not become an unreviewable mechanism for stratifying claimants, obscuring reasons, or normalising suspicion. The OECD's recent work on social security appropriately frames the goal as trustworthy use that improves access while safeguarding equity, transparency, and public trust.

The research questions can therefore be answered in practical terms. AI is most defensible firstly as an administrative support tool for communication, document management, triage, forecasting, and quality control. It is riskiest when it materially affects entitlement, recovery, suspension, or fraud liability without meaningful explanation and challenge. Its main governance challenge is not merely technical accuracy but the preservation of legality, individualised justice, and public trust in high-volume benefit administration. The deepfake and synthetic identity threat extends this challenge from decision-making into the identity infrastructure on which digital social insurance increasingly depends.

A social insurance institution considering AI should proceed in a clear sequence. First, it should classify each proposed use case by legal effect, claimant impact, data sensitivity, autonomy, and reversibility. Second, it should conduct a pre-deployment rights and risk assessment that tests data quality, bias, accuracy, explainability, identity security implications, and administrative burden effects. Third, it should deploy only with meaningful human oversight, operational guidance, claimant notification, reasons, appeal routes, and authority to override or halt the system. Fourth, it should monitor the system after launch through audits of error patterns, appeals, group-level impacts, fraud adaptation, staff behaviour, and serious incidents. Finally, it should treat procurement as an accountability decision. No system should be acquired or retained if the institution cannot inspect, explain, audit, suspend, and correct it.

This roadmap should be applied according to technology type. Rule-based ADM, robotic process automation, predictive scoring, biometric verification, and generative AI do not raise identical risks. A rule-based eligibility checker may primarily require legal correctness and exception handling. A predictive fraud model requires bias monitoring and contestability. A biometric or liveness system requires identity security safeguards and proportionality review, while a generative AI work support tool requires controls against hallucination, opacity, and misplaced responsibility on frontline staff. Treating these systems as a single category would obscure precisely the differences that responsible social insurance governance must manage.

The deepfake problem makes this even more urgent. Social insurance institutions are entering a period in which both administration and fraud are becoming AI-enabled. The institutions that cope best will not be those that automate the fastest, but those that combine technical innovation with legal discipline, procedural justice, robust identity assurance, and credible human redress. In a domain as socially consequential as social insurance, that is the only defensible standard.

REFERENCES

Agbabiaka, Olusegun, Adegboyega Ojo, and Niall Connolly. „Requirements for Trustworthy AI-Enabled Automated Decision-Making in the Public Sector: A Systematic Review." *Technological Forecasting and Social Change* 215 (2025): 124076. https://doi.org/10.1016/j.techfore.2025.124076.

Allhutter, Doris, Anila Alushi, Rafaela Cavalcanti de Alcântara, Maris Männiste, Christian Pentzold, and Sebastian Sosnowski. „Public Value in the Making of Automated and Datafied Welfare Futures." *Internet Policy Review* 13, no. 3 (2024): 1–22. https://doi.org/10.14763/2024.3.1803.

Allhutter, Doris, Florian Cech, Fabian Fischer, Gabriel Grill, and Astrid Mager. „Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective." *Frontiers in Big Data* 3 (2020): article 5. https://doi.org/10.3389/fdata.2020.00005.

Alon-Barkat, Saar, and Madalina Busuioc. „Human–AI Interactions in Public Sector Decision Making: 'Automation Bias' and 'Selective Adherence' to Algorithmic Advice." *Journal of Public Administration Research and Theory* 33, no. 1 (2023): 153–169. https://doi.org/10.1093/jopart/muac007.

Alon-Barkat, Saar, Madalina Busuioc, Kayla Schwoerer, and Kristina S. Weißmüller. „Algorithmic Discrimination in Public Service Provision: Understanding Citizens' Attribution of Responsibility for Human versus Algorithmic Discriminatory Outcomes." *Journal of Public Administration Research and Theory* 35, no. 4 (2025): 469–488. https://doi.org/10.1093/jopart/muaf024.

Alston, Philip. „Digital Welfare States and Human Rights: Report of the Special Rapporteur on Extreme Poverty and Human Rights." United Nations General Assembly, A/74/493, 2019. Accessed August 21, 2026. https://www.ohchr.org/en/documents/thematic-reports/a74493-digital-welfare-states-and-human-rights-report-special-rapporteur.

Bennett, Hayley, Morgan Currie, and Lena Podoletz. „Universal Credit: Administrative Burdens of Automated Welfare." *Journal of Social Policy* 55, no. 1 (2026): 24–42. https://doi.org/10.1017/S0047279424000175.

Bovens, Mark, and Stavros Zouridis. „From Street-Level to System-Level Bureaucracies: How Information and Communication Technology Is Transforming Administrative Discretion and Constitutional Control." *Public Administration Review* 62, no. 2 (2002): 174–184. https://doi.org/10.1111/0033–3352.00168.

Busuioc, Madalina. „Accountable Artificial Intelligence: Holding Algorithms to Account." *Public Administration Review* 81, no. 5 (2021): 825–836. https://doi.org/10.1111/puar.13293.

Cabinet Office, Department for Science, Innovation and Technology, and Government Digital Service. „DSIT: GOV.UK One Login—Liveness & Likeness Checks." *GOV.UK*. Published December 16, 2025. Accessed August 21, 2026. https://www.gov.uk/algorithmic-transparency-records/dsit-gov-dot-uk-one-login-liveness-and-likeness-checks.

Choroszewicz, Marta. „Positioning Public Sector Practitioners as 'Moral Crumple Zones': Mechanisms in the Early Use of Generative AI Work Support Tools." *Government Information Quarterly* 43, no. 2 (2026): 102128. https://doi.org/10.1016/j.giq.2026.102128.

Delpierre, Alizée, Didier Demazière, and Hajar El Fatihi. „The Stealth Legitimization of a Controversial Policy Tool: Statistical Profiling in French Public Employment Service." *Regulation & Governance* 18, no. 2 (2024): 499–512. https://doi.org/10.1111/rego.12541.

Dencik, Lina, and Anne Kaun. „Datafication and the Welfare State." *Global Perspectives* 1, no. 1 (2020): 12912. https://doi.org/10.1525/gp.2020.12912.

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

European Parliament and Council of the European Union. *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No. 300/2008, (EU) No. 167/2013, (EU) No. 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Official Journal of the European Union* L, 2024/1689, July 12, 2024. Accessed August 21, 2026. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX%3A02024R1689–20260727.

European Parliament and Council of the European Union. *Regulation (EU) 2026/1744 of the European Parliament and of the Council of 8 July 2026 amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI). Official Journal of the European Union* L, 2026/1744, July 24, 2026. Accessed August 21, 2026. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AL_202601744.

Europol. *Facing Reality? Law Enforcement and the Challenge of Deepfakes: An Observatory Report from the Europol Innovation Lab*. Publications Office of the European Union, 2022. https://doi.org/10.2813/08370.

Gerdon, Frederic, Ruben L. Bach, Christoph Kern, and Frauke Kreuter. „Social Impacts of Algorithmic Decision-Making: A Research Agenda for the Social Sciences." *Big Data & Society* 9, no. 1 (2022): 20539517221089305. https://doi.org/10.1177/20539517221089305.

Germundsson, Nora, and Hugo Stranz. „Automating Social Assistance: Exploring the Use of Robotic Process Automation in the Swedish Personal Social Services." *International Journal of Social Welfare* 33, no. 3 (2024): 647–658. https://doi.org/10.1111/ijsw.12633.

Green, Ben. „The Flaws of Policies Requiring Human Oversight of Government Algorithms." *Computer Law & Security Review* 45 (2022): 105681. https://doi.org/10.1016/j.clsr.2022.105681.

Greenhalgh, Trisha, Sally Thorne, and Kirsti Malterud. „Time to Challenge the Spurious Hierarchy of Systematic over Narrative Reviews?" *European Journal of Clinical Investigation* 48, no. 6 (2018): e12931. https://doi.org/10.1111/eci.12931.

Larsson, Karl Kristian. „Digitization or Equality: When Government Automation Covers Some, but Not All Citizens." *Government Information Quarterly* 38, no. 1 (2021): 101547. https://doi.org/10.1016/j.giq.2020.101547.

Madan, Rohit, and Mona Ashok. „AI Adoption and Diffusion in Public Administration: A Systematic Literature Review and Future Research Agenda." *Government Information Quarterly* 40, no. 1 (2023): 101774. https://doi.org/10.1016/j.giq.2022.101774.

Niklas, Jędrzej. *Social Rights and Data Technologies: Looking for Connections*. Working Paper. Data Justice Lab, Cardiff University, 2022. Accessed August 21, 2026. https://datajusticeproject.net/wp-content/uploads/2022/01/WP_Social-rights.pdf.

Niklas, Jędrzej, Karolina Sztandar-Sztanderska, and Katarzyna Szymielewicz. *Profiling the Unemployed in Poland: Social and Political Implications of Algorithmic Decision Making*. Fundacja Panoptykon, 2015. Accessed August 21, 2026. https://panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon_profiling_report_final.pdf.

OECD. *AI and the Future of Social Protection in OECD Countries*. OECD Artificial Intelligence Papers, no. 42. OECD Publishing, 2025. https://doi.org/10.1787/7b245f7e-en.

OECD. *Harnessing Artificial Intelligence in Social Security: Use Cases, Governance and Workforce Readiness*. OECD Digital Government Studies. OECD Publishing, 2025. https://doi.org/10.1787/b52405c1-en.

Rachovitsa, Adamantia, and Niclas Johann. „The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case." *Human Rights Law Review* 22, no. 2 (2022): ngac010. https://doi.org/10.1093/hrlr/ngac010.

Ranerup, Agneta, and Helle Zinner Henriksen. „Value Positions Viewed Through the Lens of Automated Decision-Making: The Case of Social Services." *Government Information Quarterly* 36, no. 4 (2019): 101377. https://doi.org/10.1016/j.giq.2019.05.004.

Ranerup, Agneta, and Lupita Svensson. „Automated Decision-Making, Discretion and Public Values: A Case Study of Two Municipalities and Their Case Management of Social Assistance." *European Journal of Social Work* 26, no. 5 (2023): 948–962. https://doi.org/10.1080/13691457.2023.2185875.

Redden, Joanna, Jessica Brand, Ina Sander, and Harry Warne. *Automating Public Services: Learning from Cancelled Systems*. Carnegie UK and Data Justice Lab, 2022. Accessed August 21, 2026. https://carnegieuk.org/publication/automating-public-services-learning-from-cancelled-systems/.

Rizk, Aya, and Ida Lindgren. „Automated Decision-Making in Public Administration: Changing the Decision Space between Public Officials and Citizens." *Government Information Quarterly* 42, no. 3 (2025): 102061. https://doi.org/10.1016/j.giq.2025.102061.

Røhl, Ulrik B. U. „Automated Decision-Making and Good Administration: Views from Inside the Government Machinery." *Government Information Quarterly* 40, no. 4 (2023): 101864. https://doi.org/10.1016/j.giq.2023.101864.

Ruschemeier, Hannah, and Lukas J. Hondrich. „Automation Bias in Public Administration—An Interdisciplinary Perspective from Law and Psychology." *Government Information Quarterly* 41, no. 3 (2024): 101953. https://doi.org/10.1016/j.giq.2024.101953.

Scott, Kristen M., Sonja Mei Wang, Milagros Miceli, Pieter Delobelle, Karolina Sztandar-Sztanderska, and Bettina Berendt. „Algorithmic Tools in Public Employment Services: Towards a Jobseeker-Centric Perspective." In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2138–2148. Association for Computing Machinery, 2022. https://doi.org/10.1145/3531146.3534631.

Singh, Sonam, and Amol Dhumane. „Unmasking Digital Deceptions: An Integrative Review of Deepfake Detection, Multimedia Forensics, and Cybersecurity Challenges." *MethodsX* 15 (2025): 103632. https://doi.org/10.1016/j.mex.2025.103632.

Snyder, Hannah. „Literature Review as a Research Methodology: An Overview and Guidelines." *Journal of Business Research* 104 (2019): 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039.

Sztandar-Sztanderska, Karolina. „How Context Matters: Human Oversight of Automated Decision-Making Systems in Welfare Administration." *Journal of European Social Policy* 36, no. 3 (2026): 255–269. https://doi.org/10.1177/09589287251358069.

Sztandar-Sztanderska, Karolina, and Joanna Mazur. „Politics and Algorithmic Articulation of Law: Tracing the Discrepancies and Backstage Decision-Making in the Development of the Profiling Algorithm in Polish Labour Market Policies." *International Journal of Law in Context* 21, no. 4 (2025): 749–768. https://doi.org/10.1017/S1744552325100141.

Sztandar-Sztanderska, Karolina, and Marianna Zieleńska. „What Makes an Ideal Unemployed Person? Values and Norms Encapsulated in a Computerized Profiling Tool." *Social Work & Society* 18, no. 1 (2020): 1–16. Accessed August 21, 2026. https://ejournals.bib.uni-wuppertal.de/index.php/sws/article/view/638.

Temoshok, David, Christine Abruzzi, Yee-Yin Choong, James L. Fenton, Ryan Galluzzo, Connie LaSalle, Naomi Lefkovitz, Andrew Regenscheid, and Maria Vachino. *Digital Identity Guidelines: Identity Proofing and Enrollment*. NIST Special Publication 800–63A-4. National Institute of Standards and Technology, 2025. https://doi.org/10.6028/NIST.SP.800–63A-4.

U.S. Department of the Treasury, Financial Crimes Enforcement Network. „FinCEN Alert on Fraud Schemes Involving Deepfake Media Targeting Financial Institutions." FIN-2024-Alert004. November 13, 2024. Accessed August 21, 2026. https://www.fincen.gov/system/files/shared/FinCEN-Alert-DeepFakes-Alert508FINAL.pdf.

U.S. Government Accountability Office. *Social Security Administration: Actions Needed to Help Ensure Success of Electronic Verification Service*. GAO-24–106770. September 10, 2024. Accessed August 21, 2026. https://www.gao.gov/products/gao-24–106770.

Verhagen, Annelore. *Using AI to Manage Minimum Income Benefits and Unemployment Assistance: Opportunities, Risks and Possible Policy Directions*. OECD Artificial Intelligence Papers, no. 21. OECD Publishing, 2024. https://doi.org/10.1787/718c93a1-en.

Vredenburgh, Kate. „Transparency and Explainability for Public Policy." *LSE Public Policy Review* 3, no. 3 (2024): article 4, 1–11. https://doi.org/10.31389/lseppr.111.

Zakharova, Irina, Juliane Jarke, and Anne Kaun. „Tensions in Digital Welfare States: Three Perspectives on Care and Control." *Journal of Sociology* 60, no. 3 (2024): 540–559. https://doi.org/10.1177/14407833241238312.

Zen, Hilary, Rohan Wagh, Miguel Wanderley, Gustavo Bicalho, Rachel Park, Megan Sun, Rafael Palacios, Lucas Carvalho, Guilherme Rinaldo, and Amar Gupta. „Ensemble-Based Biometric Verification: Defending Against Multi-Strategy Deepfake Image Generation." *Computers* 14, no. 6 (2025): 225. https://doi.org/10.3390/computers14060225.