Digital Convergence USP 2030 | AI HUB

AI ADOPTION IN SOCIAL PROTECTION

A Global Evidence Review
2020–2025

AI HUB FOR SOCIAL PROTECTION

The AI Hub for Social Protection (AI Hub) is an initiative under the Digital Convergence Initiative (DCI), which helps social protection institutions to harness the benefits of responsible AI for the public good. It supports the development of strategies, frameworks, digital systems and institutional capacities to design, govern and implement AI solutions that are ethical, effective and uphold data sovereignty.

Visit our website at **spdci.org/ai-hub**.

ACKNOWLEDGMENTS

This report was authored by Thomas Byrnes and Avril James (Consultants, MarketImpact Digital Solutions Ltd). We would like to thank the many colleagues who reviewed the various drafts: Juul Pinxten and Phillippe Leite (World Bank), Rob Worthington (Kwantu), Gabriele Erba (UNICEF), Valentina Barca (independent consultant), Franziska Clausius and Nour Barakat (GIZ). We would also like to thank Susan Sellars (copyeditor) for helping to bring the report into its final form.

REQUIRED CITATION

DCI AI Hub for Social Protection, *AI Adoption in Social Protection: A Global Evidence Review (2020-2025)*, Deutsche Gesellschaft für Internationale Zusammenarbeit, Germany, 2026.

AI HUB KNOWLEDGE ARCHITECTURE

The AI Hub is committed to advancing the global dialogue on the use of AI in social protection by creating and sharing knowledge on the responsible adoption of AI solutions. This Global Evidence Review is part of a broader portfolio of AI Hub knowledge products that support policymakers, practitioners, and partners in navigating the opportunities and challenges of AI in social protection.

Artificial intelligence is a fast-moving field, with new developments, evidence, and applications emerging continuously. While this report reflects the best available information at the time of publication, it may not include subsequent advances. Readers are encouraged to consult our AI Hub website for the most recent developments, publications, and resources: **spdci.org/ai-hub**.

CONTENTS

ACRONYMS

ABIS	Automated Biometric Identification System
AHEAD	Anticipatory Humanitarian Action for Displacement
AI	artificial intelligence
AIA	algorithmic impact assessment
BIMS	Biometric Identity Management System
CDR	call detail record
CPGRAMS	Centralised Public Grievance Redress and Monitoring System
DCI	Digital Convergence Initiative
DCRM	Displacement Crisis Response Mechanism
DPIA	data protection impact assessment
DWP	Department for Work and Pensions
GIZ	Deutsche Gesellschaft für Internationale Zusammenarbeit
HIC	high-income country
HITL	human-in-the-loop
HOOTL	human-out-of-the-loop
HOTL	human-on-the-loop
ILO	International Labour Organization
ISSA	International Social Security Association
LIC	low-income country
LLM	large language model
LMIC	low and middle-income country
M&E	monitoring and evaluation
ML	machine learning
NLP	natural language processing
OCHA	United Nations Office for the Coordination of Humanitarian Affairs
OCR	optical character recognition
OECD	Organisation for Economic Co-operation and Development
PMT	proxy means testing

SP	social protection
SyRI	System Risk Indication
UK	United Kingdom
UMIC	upper-middle-income country
UNDP	United Nations Development Programme
UNHCR	United Nations High Commissioner for Refugees
US	United States
USCIS	United States Citizenship and Immigration Services
USD	United States dollar
WFP	World Food Programme

EXECUTIVE SUMMARY: HOW IS AI BEING APPLIED IN SOCIAL PROTECTION?

The use of AI systems in social protection is a reality, not a future prospect. With applications in at least 48 countries as of March 2026, this number is likely to grow. As the consequences of poor AI governance fall on the most vulnerable populations, it is imperative that the responsible use of AI in social protection be documented so that it can inform governments, policymakers and social protection practitioners moving into this space. To chart a way forward that ensures responsible, innovative and sovereign AI adoption we must learn from the past to guide future decisions.

This Global Evidence Review examines what AI in social protection looks like in practice: where it exists, in what form, and with what evidence. Its primary focus is on low-income countries (LICs) and low- and middle-income countries (LMICs), where the evidence gap and policy needs are greatest. Cases from high-income countries (HIC) are also included for the governance lessons and cautionary precedents they provide.

The evidence is clustered thematically around ten use case categories, four of which dominate over all others. *User communication and interaction* accounts for 22 cases, *vulnerability and needs assessment* for 16, *compliance and integrity* for 14, and *identification* for 13. Together these categories contain two thirds of the evidence base. In addition, data quality appears (although as a secondary function) in 12 cases and policy analysis in 3 cases. HICs dominate, with 52 of the 99 cases. This reflects documentation patterns and search asymmetries as much as deployment. Nineteen (19) cases were identified in LMICs and 17 in LICs. Upper-middle-income countries (UMICs) are notably under-represented with 11 cases, despite operating some of the world's largest digitised social protection systems. Classical machine learning accounts for 60 cases and deep learning accounts for 19, concentrated in biometric verification and document processing. Foundation models, including large language models (LLMs), appear in 11 cases, predominantly in chatbots and back-office automation.

It is clear from the review that what matters most is not whether or not you use AI in social protection, but how its use is governed. Across 99 cases in 48 countries, the evidence shows that the same AI technique produces fundamentally different results depending on the institutional arrangements surrounding it. Prediction models map poverty in Togo and detect fraud in the Netherlands, but the outcomes for the people they affect could not be more different. In Togo, AI-assisted poverty mapping helped direct emergency cash transfers to approximately 140,000 beneficiaries during COVID-19. In the Netherlands, SyRI, an algorithmic welfare fraud detection system, was struck down by The Hague District Court for violating the right to privacy, having failed to yield a single proven case of fraud in the targeted deployments that triggered the lawsuit.

The analysis also highlights the inextricability of fairness, equality and human rights principles from AI applications. In five cases, social protection systems using AI were suspended or halted, with three of them due to compliance and integrity concerns. In every instance the need for good governance was reaffirmed to prevent unequal service delivery. An independent analysis of Sweden's Försäkringskassan fraud-scoring model revealed that it flagged women 1.5 times more often, and people with foreign backgrounds 2.5 times more often, than the general population. Ireland suspended facial recognition software used for identity fraud detection in welfare. The Netherlands halted two separate algorithmic fraud systems: SyRI after a court ruling on privacy grounds and Rotterdam's welfare fraud algorithm after an audit confirmed discriminatory profiling. Denmark halted a predictive vulnerability profiling pilot targeted at children before wider rollout, following a data protection and ethics review. These cases are as instructive as cases illustrating the successful use of AI in social protection systems.

Lastly, it is important to note that this review is a desk-research baseline, not a comprehensive census. It provides an analysis of patterns in use-cases, showing the centrality of design choices, oversight arrangements, institutional capacity and legal safeguards in the design of AI systems for social protection. The review documents what we found through structured searches of published and grey literature. Areas where we found little evidence – such as UMIC contexts, labour market programmes, AI-assisted casework, and longitudinal evaluation – may reflect gaps in documentation rather than gaps in deployment. This global evidence review plays a foundational role in mapping the current landscape of AI in social protection, while encouraging further research in contexts and cases that go beyond its scope.

CHAPTER 1. INTRODUCTION

1.1 AI in social protection – The context

As of March 2026, at least 48 countries are running AI systems in their social protection programmes. Most are invisible to the people they affect, yet their impacts are very real. A pensioner in Togo submits a proof-of-life selfie through facial recognition (TGO-002). A farmer in India asks a chatbot whether his cash transfer has been deposited (IND-001; Press Information Bureau, 2023b). A single mother in the Netherlands is flagged for fraud investigation by an algorithm she has never heard of (NLD-002; Lighthouse Reports, 2023c).

As the use of AI in the sector is likely to increase, we need to chart a way forward that ensures responsible, innovative and sovereign AI adoption.[1] To do this, we need to learn from the past to shape future decisions. The policy motivation for this review is, thus, supporting decision making in low-income countries (LICs) and low- and middle-income countries (LMICs). That is where the evidence gap is greatest, where social protection (SP) systems are expanding fastest, and where the consequences of poor AI governance fall on the most vulnerable populations. High-income country (HIC) cases are included because they illuminate risks, governance choices, and transferable lessons. The absence of evidence from LICs and LMICs often reflects documentation scarcity, rather than the absence of deployment.

This review documents what exists, maps the evidence, and identifies patterns. As one of the AI Hub's foundational knowledge products, the review complements and builds upon recent thematic work in this space. The Technology Bytes series (2023–2025) (ISSA, 2026) examined AI applications in specific delivery chain functions, with a focus on social assistance in low-income contexts. The OECD report on AI in social protection (OECD, 2025) surveyed the policy landscape across OECD member states. Our review differs in two respects. First, it applies the DCI AI Hub's *A Taxonomy for AI in Social Protection: Defining Capabilities and Application Areas* (DCI AI Hub, 2026a) as a consistent classification framework across all cases, enabling systematic comparison. Second, it covers a broader geographic and programmatic scope, including social insurance and labour market programmes alongside social assistance.

Following this introductory chapter, we provide an overview of key trends across the evidence base (Chapter 2). Chapter 3 presents the evidence organised into the ten use case categories. Each section is self-standing, so readers can go directly to the category matching their operational interest. Chapter 3 synthesises cross-cutting governance patterns and evidence gaps. Chapter 4 provides a brief conclusion highlighting the evidence gaps. Annex 1 contains a case index with country, programme name and primary citation for each of the 99 cases referenced throughout this report and Annex 2 contains a glossary of key terms.

1 See DCI's AI Hub mission statement: https://spdci.org/ai-hub/.

1.2 Methodology

The evidence for this review was assembled through a multi-stage desk research process conducted between June 2025 and March 2026. Evidence was gathered from peer-reviewed literature, institutional and legal documents, and high-quality grey literature. Although we drew on cases identified by earlier reviews (ISSA, 2025; OECD, 2025) where they met our inclusion criteria, the majority of the 99 cases of AI in social protection were identified through an independent desk research.

The review covers AI applications in social protection, as defined in *A Taxonomy for AI in SP* (DCI AI Hub, 2026a). This includes government-implemented or government-commissioned systems serving the sector: social assistance (non-contributory transfers and services), social insurance (contributory schemes including pensions, unemployment and government health insurance), and labour market programmes (active labour market policies including employment services, training, and job matching). We excluded three types of system: purely private-sector products that do not interface with government social protection programmes; humanitarian-only deployments, unless they directly interface with national social protection systems; and deterministic automation that applies fixed rules without a learning component. We included systems developed by international organisations or the private sector where they were commissioned by or integrated into government programmes.

The initial case identification phase used structured keyword searches across institutional repositories (World Bank Open Knowledge Repository, UNICEF Innocenti, ILO publications, OECD Social Policy Division), academic databases (Google Scholar, SSRN, arXiv), and targeted searches of government digital transformation portals. Search terms combined social protection terminology (social assistance, social insurance, social security, welfare, benefits, pensions, unemployment) with AI and technology terms (artificial intelligence, machine learning, algorithm, predictive analytics, chatbot, biometric, natural language processing, deep learning). We included cases that met three criteria: (i) the system uses AI as defined by the boundaries of *A Taxonomy for AI in SP* (DCI AI Hub, 2026a), meaning a learning or generative component is present, (ii) the system operates within or interfaces with a government social protection programme (i.e. it is not purely humanitarian and parallel), and (iii) sufficient primary or secondary source documentation exists to verify the deployment.

AI refers to computational systems that use statistical learning, pattern recognition, or generative models to perform tasks that would otherwise require human cognitive effort. *A Taxonomy for AI in SP* (DCI AI Hub, 2026a) identifies eight AI capabilities in two families: five analytical (prediction, classification, clustering, anomaly detection, ranking) and three generative (perception and extraction, large language models, synthetic data generation). Deterministic rule-based systems are thus excluded. A system that applies fixed eligibility thresholds without learning from data is not AI under this definition, even if it is automated. This boundary matters because many social protection 'digitalisation' efforts involve rule-based automation rather than AI.

Each case was coded against the ten AI use-case categories in *A Taxonomy for AI in SP* (DCI AI Hub, 2026a) (see Table 1) and validated against primary sources. Each category describes what the AI system does, whether it predicts, classifies, detects anomalies, generates text, or performs another task applied to a specific social protection function, regardless of the role it plays within the social protection system at the policy, programme design, or programme delivery (i.e. delivery chain) level.[2] These categories thus represent the key roles that AI can play vis-a-vis a broader set of social protection functions. Confidence levels were assigned to the coding: confirmed (verified against multiple independent sources), probable (verified against a single primary source), or unverified (based on secondary reporting only). The full coding schema is available in *A Taxonomy for AI in Social Protection* (DCI AI Hub, 2026a).

2 These generic use cases can subsequently be mapped back to specific social protection functions at each level. Including each stage of the delivery chain, see *A Taxonomy for AI in Social Protection: Defining Capabilities and Application Areas* (DCI AI Hub, 2026a) for details.

A single AI deployment may serve more than one of these functions or roles. A chatbot, for instance, typically performs user communication (its primary function) alongside workflow routing (a secondary function) within the same pipeline. Table 1, therefore, shows two counts for each category: the number of cases where the category is the system's primary classification, and the number where it is a secondary function performed alongside something else. All 99 cases have one primary classification; 79 of them carry at least one secondary classification.

TABLE 1. AI USE-CASE CATEGORIES AND EVIDENCE COUNT (N=99)

USE CASE CATEGORY	PRIMARY	SECONDARY	PRIMARY CAPABILITY	CORE APPLICATIONS (LESS DOCUMENTED IN ITALICS)
Vulnerability, needs and risk assessment	16	16	Prediction	Poverty and vulnerability mapping, welfare scoring (including proxy means testing), predictive profiling, non-take-up detection, outreach and registration segmentation
Trend and shock forecasting	7	0	Prediction	Climate, conflict and food-insecurity early-warning systems
Identification, verification and record linkage	13	0	Perception and extraction	Biometric authentication, deduplication, document verification
Decision support for eligibility and benefits	6	8	Classification	Caseworker advisory tools for eligibility or benefit configuration
Data quality and anomaly detection	0	12	Anomaly detection	Duplicate detection, data validation
User communication and interaction	22	8	Perception and extraction	Chatbots, voice assistants, automated information services
Matching and recommendation	7	5	Classification/ranking	Job matching, skills profiling, programme recommendation
Operational and process automation	11	12	Large language models (LLMs)	Document processing, case summarisation, workflow automation
Compliance and integrity	14	15	Anomaly detection	Fraud detection, risk scoring, conditionality monitoring
Policy analysis, learning and M&E	3	3	Various	Microsimulation, bias auditing, model performance monitoring

***A Taxonomy for AI in SP* (DCI AI Hub, 2026a) also classifies cases along a set of further dimensions that shape how a deployment behaves in practice:** AI capabilities and tasks, lifecycle stages, implementation types, decision criticality, human-oversight arrangements, development process, deployment status, and risk and societal-impact categories. Several of these dimensions – particularly decision criticality, human oversight, and deployment status – are used in the cross-cutting analysis in this review.

We used AI tools in producing this review. Large language models assisted with initial case identification from large document corpora, extraction of structured data from unstructured sources, and drafting of case descriptions. All AI-generated outputs were manually reviewed, fact-checked against primary sources, and corrected where necessary. AI use was documented in accordance with the DCI Style Guide requirements for transparency.

1.3 Limitations

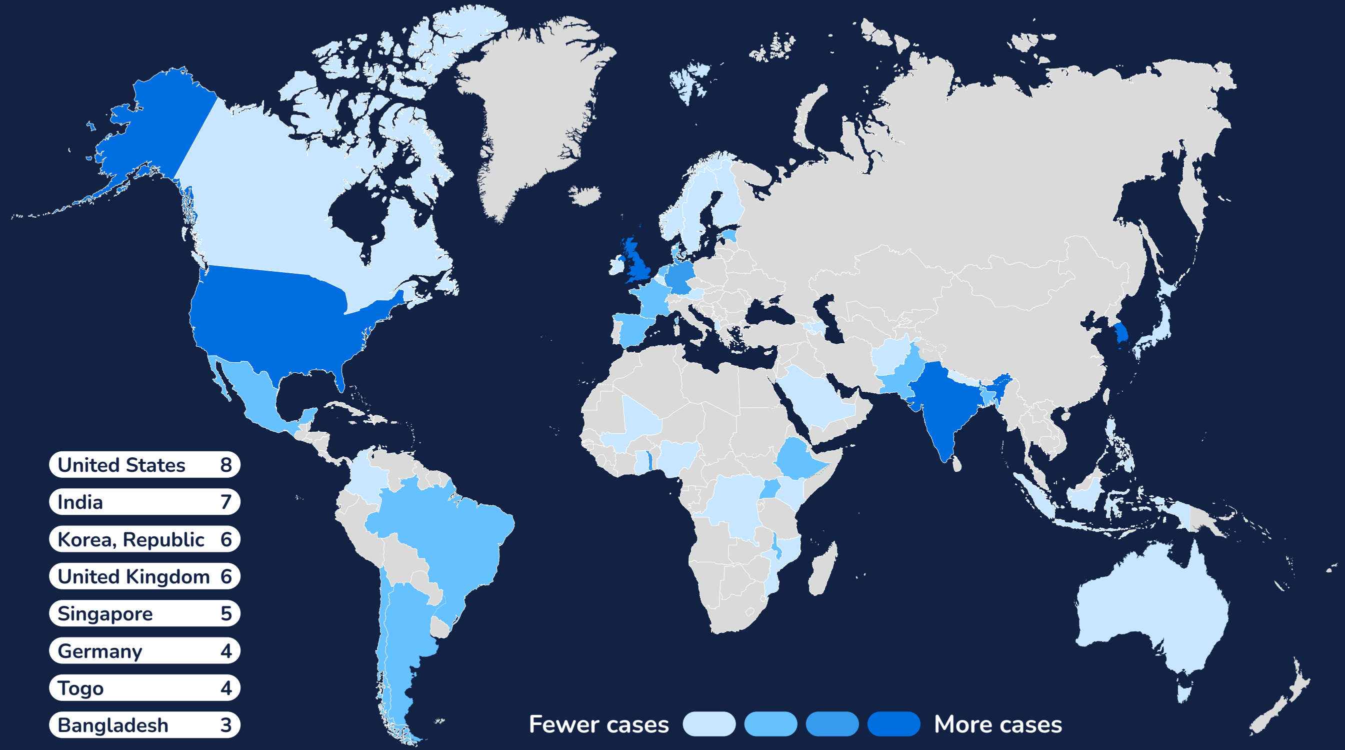
This is a desk-research baseline, not a comprehensive census of AI in social protection. The primary focus of the review was on LIC and LMIC contexts, where the evidence gap is greatest and the policy need most acute. We have included all HIC cases uncovered during the research, but we did not conduct the same deep targeted search for high-income settings. The HIC case count (52) is, therefore, a by-product of the search process rather than a systematic inventory of HIC deployments.

More broadly, it is important to note that the evidence base is shaped by what is documented and accessible in English, French, and Spanish. Deployments that are not publicly documented, that exist only in internal government reports, or that operate in languages we did not search are absent. Some apparent gaps, for instance, the low number of upper-middle-income country (UMIC) cases, may reflect documentation shortfalls rather than a genuine absence of AI deployment. Hence, the evidence base should be treated as a lower bound on global activity, not a complete inventory.

CHAPTER 2. EVIDENCE AT A GLANCE

FIGURE 1. USE CASE DISTRIBUTION AND IMPORTANT FACTS (N=99)

At least 48 countries are running AI systems serving the social protection sector.
This is a trend that is *already* happening, ignoring it is not an option.



OVER 50%
of the 99 cases reviewed come from high-income countries, despite the focus on low- and middle-income countries.

Number of cases across social protection pillars:

- **56** Social assistance
- **10** Labour market programmes
- **33** Social insurance

58% of reviewed systems are already operational and serving real populations.

The most common primary AI capability is **perception & extraction** while generative capabilities are newer and less widespread.

...

Source: DCI AI Hub Tracker (forthcoming), data as of March 2026 (see box below for more)

...

A) Perception and extraction is the most common primary AI capability.

It underpins biometric verification, document processing, and chatbot interfaces. Prediction (26 cases) drives both vulnerability assessment and compliance monitoring, while classification (14 cases) primarily supports eligibility determination and case routing. Generative capabilities are newer and less widespread.

B) High income countries prevail.

HICs account for 52 of the 99 cases reviewed, despite this review's primary focus on LICs and LMICs. This reflects greater documentation capacity and more mature digital infrastructure, not necessarily greater deployment. LMICs contribute 19 cases and LICs 17, driven largely by internationally funded deployments in vulnerability assessment and identification. UMICs are notably under-represented at 11 cases. China, Turkey, Thailand, and South Africa all operate large, digitised social protection systems, but appear nowhere in this evidence base. Whether this reflects genuine under-deployment, or a documentation gap in languages and institutional contexts that our desk research did not cover, is an open question.

C) Most cases are operational.

Thirty-two (32) are in full production and 26 in limited rollout, meaning that 58% of the evidence base represents systems serving real populations. Twelve (12) are scaled and institutionalised, 20 are at the pilot stage, and 4 in at the design stage. Five (5) have been suspended or halted, each for governance failures rather than technical limitations. Table 2 in Chapter 3 lists these cases and the accountability mechanisms that ended each.

This is not a collection of experiments. Most of these systems are making decisions about real people. See the AI tracker for details across all these categories.

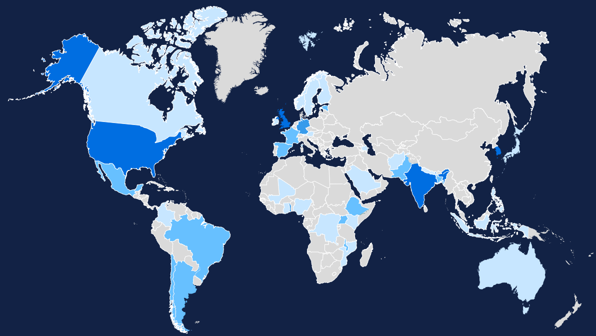


FIGURE 2. TEN MAJOR AI USE CASE CATEGORIES AND THEIR ASSOCIATED GOVERNANCE RISKS

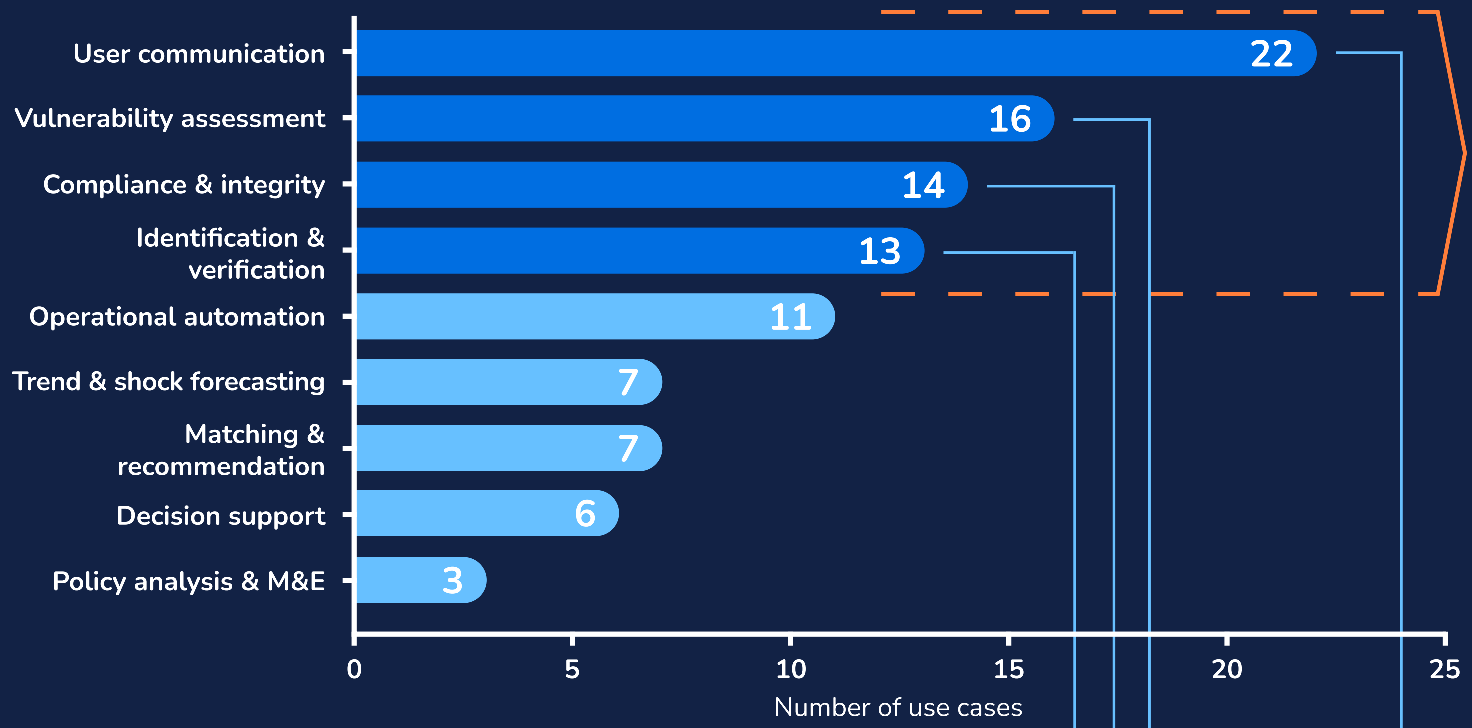
There are 10 major AI use case categories in social protection … these carry different levels of governance risks.

Note: These standard ratings escalate further where outputs of specific use cases feed **individual eligibility**, **sanctions** or **automated adverse action**

#	Category	Description	Risk
1	**Vulnerability, needs and risk assessment**	Identifies for whom and where support is needed most (currently and as future projection) by estimating poverty, vulnerability or risk to guide policy, design and resource allocation	Risk: High
2	**Trend and shock forecasting**	Anticipates future trends, shocks or crises to enable early planning, preparedness and anticipatory action	Risk: Low to moderate
3	**Identification, verification and record linkage**	Confirms identity, validates claims and links records across systems to create a unified view of individuals or households	Risk: High
4	**Decision support (e.g. for determination of benefits and services packages)**	Assesses whether individuals qualify for programmes and what sup port they should receive, through rule-based or model-assisted scoring	Risk: Very high
5	**Data quality and anomaly detection**	Prepares, validates and enhances data, while also detecting unusual or suspicious patterns, prioritising cases for data quality checks or investigation	Risk: High
6	**User communication and interaction**	Facilitates two-way information exchange with beneficiaries or staff through automated or assisted channels	Risk: Low to moderate
7	**Matching and recommendation**	Connects individuals to suitable benefits, services, jobs or training by analysing profiles and external data sources	Risk: Moderate
8	**Operational and process automation**	Streamlines workflows through document processing, generative staff assistance, and forecasting of operational demand	Risk: Low to moderate
9	**Compliance and integrity**	Monitors adherence to programme rules and manages complaints and appeals, including conditionality and behavioural compliance	Risk: High
10	**Policy analysis, learning and M&E**	Generates insights to improve policy and programme design, including bias and fairness evaluation	Risk: Low to moderate

Note: The risk categorisation refers to governance sensitivity. The levels reflected in this visual show the default tier; ratings for specific use cases within these categories escalate where outputs feed individual eligibility, sanctions or automated adverse action.

FIGURE 3. DISTRIBUTION OF USE CASE CATEGORIES (N=99)



4 categories account for 65% of reviewed cases.

This partly reflects where governments have found AI to be most useful.

Common does NOT mean low risk – high volume uses need safeguards.

- Examples include **biometric authentication, deduplication, document verification, and entity resolution** – potentially high risk use cases.
- Examples include **algorithmic risk scoring** for welfare fraud investigation, **anomaly detection and predictive models that prioritise audit targets** – high risk use cases.
- Examples include **geospatial poverty mapping, new approaches to proxy means testing (PMT), and predictive profiling** – high risk use cases.
- **Predominantly chatbots and voice assistants** that handle benefits enquiries, appointment bookings, or programme guidance – lower risk use cases.

Source: DCI AI Hub Tracker (forthcoming), data as of March 2026. For detailed definitions of each category, see *A Taxonomy for AI in Social Protection* (DCI AI Hub, 2026a)[3].

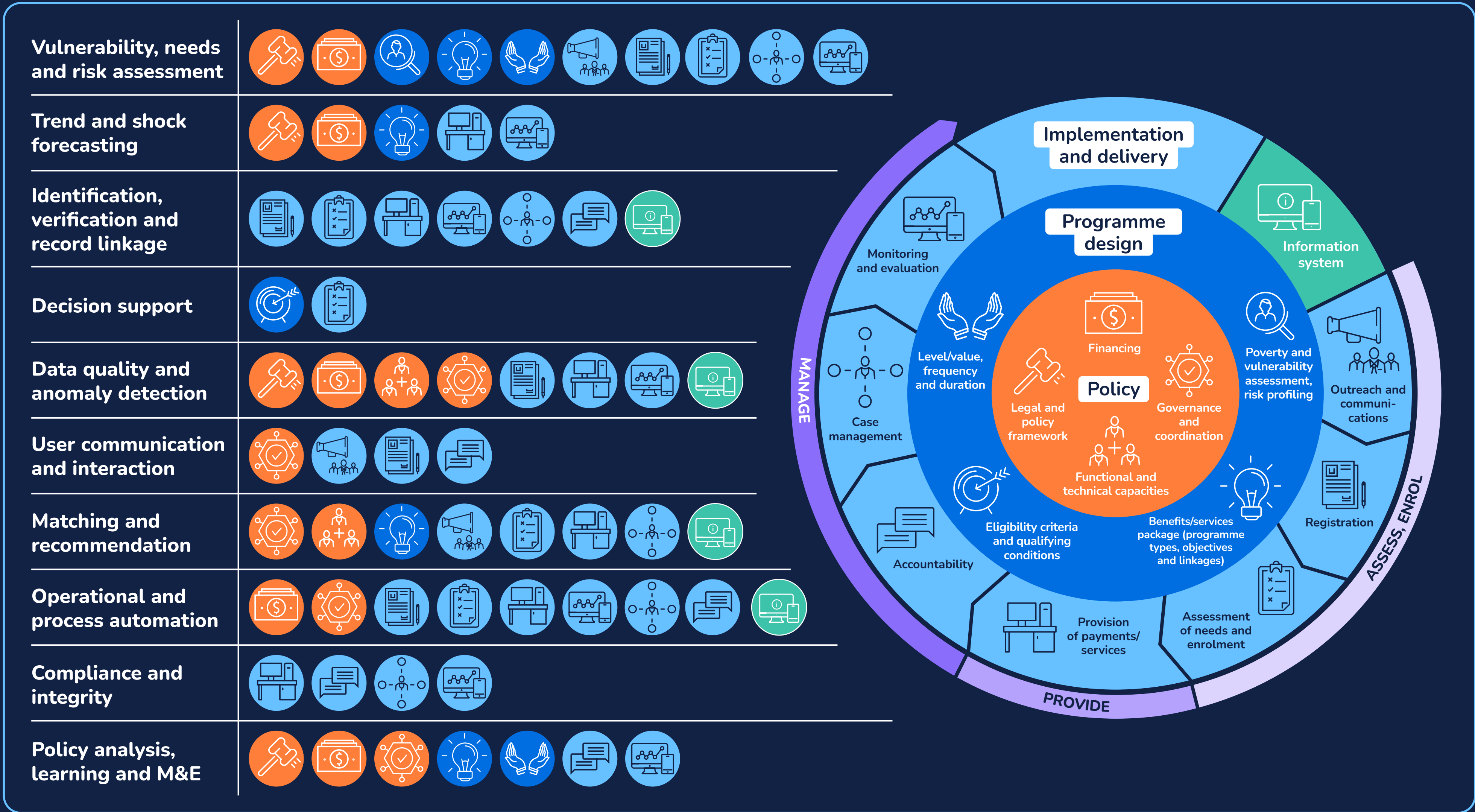
3 Note: This distribution partly reflects where governments have found AI to be most applicable, but it also reflects what is most documented (internal analytical tools and embedded data quality processes are less likely to appear in the literature). In fact, several of the less common cases feature as secondary classifications, as Table 1 showed.

CHAPTER 3. EVIDENCE BY USE CASE CATEGORY

This chapter presents the evidence by use case, with three patterns worth highlighting.

- **First**, identification and forecasting never appear as secondary classifications. Systems built to verify identity or forecast shocks are built as standalone systems, rather than added as components to systems with another primary function.
- **Second**, data quality and anomaly detection show the opposite; AI-driven data-quality work is ubiquitous in the evidence base, but always embedded inside a vulnerability, compliance or operations system. It is never deployed as a standalone primary function.
- **Third**, vulnerability assessment and compliance monitoring appear as secondary classifications almost as often as they do as primary. Both categories serve as headline applications and as supporting components of larger pipelines.

FIGURE 4. THE TOP 10 AI USE CASE CATEGORIES SERVE DIFFERENT SOCIAL PROTECTION FUNCTIONS AT THE POLICY, DESIGN AND IMPLEMENTATION LEVELS



Source: *A Taxonomy for AI in Social Protection: Defining Capabilities and Application Areas* (DCI AI Hub, 2026a)

Note: for the details of how each use case maps to each specific social protection system function, read the relevant section in the Taxonomy.

3.1 Vulnerability, needs and risk assessment

Note: See Figure 4 above for the full mapping against social protection system functions and refer to *A Taxonomy for AI in Social Protection* for more details (DCI AI Hub, 2026a).

Social protection functions

This category covers AI systems that help estimate for whom and where support is needed most (and may be needed in the future) by estimating poverty, vulnerability, and deprivation or risk of adverse outcomes, as well as segmenting populations to guide policy, design and operational decisions (e.g. programmatic adjustments) and resource allocation. Typical applications found in the literature include geospatial poverty mapping from satellite imagery, PMT using administrative or phone metadata, and predictive profiling that flags individuals at risk of homelessness, child welfare concerns, or benefit dependency.

What the evidence shows

The evidence base contains 16 vulnerability assessment cases spread across 14 countries. The dominant AI capabilities are prediction (13 cases) and classification (1 case).

Geospatial poverty estimation is the most established cluster within this category. In **Pakistan**, ensemble deep learning applied to satellite imagery produced 1 km² resolution poverty maps for Sindh province, achieving 71% recall and 67% precision on hold-out tests (PAK-001; Agyemang et al., 2023). In **Malawi** (MWI-001; Gualavisi & Newhouse, 2025), World Bank researchers achieved rank correlations of 0.75 between predicted and actual village-level welfare. In **Mozambique** (MOZ-002; World Food Programme, 2019), the World Food Programme's (WFP's) DEEP system used drone imagery and machine learning to assess post-cyclone building damage, achieving roughly 85% accuracy and shortening assessment timelines from weeks to days (also see **Togo** case in Box 1).

BOX 1. TOGO NOVISSI, AI-ASSISTED POVERTY MAPPING AND BENEFICIARY TARGETING

In the early months of the COVID-19 pandemic, the government of Togo faced a problem common across sub-Saharan Africa: how to get emergency cash to the poorest households when no registry existed to identify them. The answer was Novissi, a system that combined satellite imagery with mobile phone metadata to classify 10,119 grid cells and rank all 397 cantons by poverty level, selecting the poorest for programme coverage (TGO-001; Aiken et al., 2022). A second Novissi model (TGO-005; Aiken et al., 2022) then used phone-based poverty prediction to help prioritise roughly 57,000 individual beneficiaries within those areas, without requiring fresh household surveys. Together, the two models helped deliver cash transfers to approximately 140,000 beneficiaries in the rural expansion phase.

Predictive vulnerability profiling takes a different approach, using administrative data to identify individuals at risk.

In **Australia**, researchers developed a machine learning system to predict which recipients of income support would need recertification review (AUS-001). At the local government level, Maidstone Borough Council in **England** piloted a homelessness prediction system (GBR-001; Maidstone Borough Council, 2019) that generated over 650 alerts and reported a 40% reduction in the local homelessness rate. **Chile**'s Sistema de Alerta Niñez (CHL-002; Diario y Radio Universidad Chile, 2019) applies prediction to child welfare: machine learning identifies children at elevated risk of rights violations, reaching 2,262 children across 12 pilot comunas. In the **United States**, the Allegheny Family Screening Tool (USA-002; Vaithianathan et al., 2017, see Box 2) has been in continuous operation since 2016, scoring referrals to child protective services. In Los Angeles County (USA-001; California Policy Lab, 2024), predictive analytics power the Homelessness Prevention Unit, which uses model-generated risk scores to identify individuals at high risk of homelessness and offer them intensive case management and financial support upstream. This reduced shelter need and entry among participants by 71% compared to the comparison group.

BOX 2. THE ALLEGHENY FAMILY SCREENING TOOL – ETHICAL CONSIDERATIONS

As it has a long history, this case is among the most-studied cases in algorithmic accountability literature. Eubanks (2019) argued that predictive risk models in child welfare disproportionately surveil low-income families and communities of colour, and that by drawing predominantly on public-system data such as birth records, Medicaid, and juvenile probation records, the model effectively encodes existing patterns of state surveillance of poor families rather than neutral risk. Subsequent independent audits by the ACLU with the Human Rights Data Analysis Group, by the Associated Press, and an ongoing United States (US) Department of Justice civil-rights inquiry have examined racial and disability disparities in the tool's outputs and operation. The model has itself been revised several times since 2016, most recently in 2019 with a move to a LASSO (least absolute shrinkage and selection operator) specification, although the county has not publicly linked specific revisions to the critical-audit findings. The case illustrates a pattern seen elsewhere in this review. Technical performance and equity performance are distinct questions, and evaluating one does not settle the other.

Further social protection applications identified in *A Taxonomy for AI in SP* (DCI AI Hub, 2026a), but less documented in the literature include predictive vulnerability and risk scoring to target prevention efforts, non-take-up detection (identifying eligible individuals who have not claimed entitlements), and outreach/registration segmentation and optimisation (using predictive models to target outreach efforts toward underserved populations). These have potential, but remain under-documented. For a full list, see Section 1 of *A Taxonomy for AI in SP* (DCI AI Hub, 2026a).

Important considerations

This category skews towards low-income contexts: 6 of 16 cases are in LIC or LMIC settings, driven largely by international organisation partnerships. HIC cases tend to be research prototypes or subnational pilots. This also raises questions as to whether such models would be acceptable to HIC constituencies, especially where these are used to directly inform targeting of individuals and households, and what this implies for LIC and LMIC contexts.

Decision criticality is rated high in most of these cases, especially where outputs link to targeting decisions that determine who receives assistance. The oversight picture is mixed. In geospatial estimation, the AI output is typically one input among several, and errors affect geographic allocation rather than individual rights. In predictive profiling, in contrast, the model score may directly shape outcomes for named individuals and families, with correspondingly higher stakes.

Several risks warrant attention.

- Feedback effects: Where predictive profiling draws on administrative data generated by previous contact with state systems, model scores can entrench existing patterns of inclusion and exclusion rather than reveal new need (see in Box 2 on the Allegheny case).
- Opacity: The 'black box' nature of many predictive models lowers transparency and complicates accountability, particularly where affected individuals cannot see the factors that led to their classification.
- Consent and purpose limitation: Several cases in the evidence base use administrative data collected for one purpose, such as tax records or phone metadata, to infer vulnerability for another, raising questions about informed consent and data reuse.
- Exclusion of the already-excluded: People without administrative footprints, undocumented migrants, recent arrivals, those outside the formal economy, are systematically invisible to models trained on registered populations, and may be missed by the very systems designed to identify need.
- Gender bias. For example, where the data assumes a male-headed household structure or uses assets (vehicles, formal employment) that correlate with male economic activity, model output carries a latent gender bias, even when no protected attribute is an input.

The critical unanswered question is whether AI-assisted vulnerability, needs and risk assessment actually reaches people who would be missed by simpler methods, and at what cost. Independent evaluation is rare. Most evidence comes from implementer reports. The comparative cost-effectiveness of AI-assisted approaches relative to others is to this date almost entirely undocumented, especially given its 'black box' nature, which lowers transparency and complicates accountability.

> The critical unanswered question is whether AI-assisted vulnerability, needs and risk assessment actually reaches people who would be missed by simpler methods, and at what cost.

A useful distinction to address the risks deriving from this use case can be drawn between aggregate and individual-level applications. AI models that shape aggregate decisions (e.g. geographic areas for prioritising registration efforts or highlight coverage gaps like non-take-up, mapping of unmet needs to guide legal framework and policy design) raise less concerns than those used to identify and target individuals and households. Aggregate uses affect resource allocation; individual level uses affect rights and entitlements. The safeguards required should scale accordingly.

3.2 Trend and shock forecasting

Legal and policy framework

Financing

Benefits/ services package

Provision of payments/ services

Monitoring and evaluation

Note: See Figure 4 above for the full mapping against social protection system functions and refer to *A Taxonomy for AI in Social Protection* for more details (DCI AI Hub, 2026a).

Social protection functions

This category covers AI systems that anticipate future trends, shocks or crises to enable early planning and preventive or adaptive measures, including anticipatory action. In practice, they forecast events such as droughts, floods, conflict or displacement, as well as economic, demographic, health or other trends, and can be designed to link these forecasts to pre-agreed response protocols. Unlike vulnerability assessment, which estimates current need, forecasting anticipates future need so that social protection systems can act before situations degenerate. The output is often a trigger: when a forecast crosses a defined threshold, pre-committed funding (based on actuarial forecasting) is released or programme coverage is expanded.

What the evidence shows

Seven forecasting cases were documented across six countries, all using prediction as their primary capability. Despite the small case count, several of these systems are more operationally mature than those in larger categories, because they have pre-agreed trigger protocols that link AI forecasts directly to disbursement.

Climate and anticipatory action systems form the largest cluster. In **Ethiopia**, SEWAA uses AI-enhanced rainfall forecasting to trigger anticipatory action (ETH-002; World Food Programme, 2023). In September 2024, a SEWAA-informed activation deployed USD 6.6 million across 15 woredas (districts), providing cash transfers of USD 228 per household. In the **Philippines** (PHL-001; Centre for Humanitarian Data, 2022), the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) and the United Nations Central Emergency Response Fund (CERF) piloted AI-enabled typhoon impact forecasting with pre-allocated funding of up to USD 7.5 million, generating impact maps every 6–12 hours and triggering pre-emptive cash transfers 72 hours before landfall. In southern **Malawi** (MWI-002), a machine learning food security forecasting model achieved an F1 score of approximately 81% in predicting household food insecurity outcomes. In **Mali**, Ignitia provides AI-enabled SMS weather advisories to smallholder farmers (MLI-003), with WFP reporting roughly 10% yield improvement among the 5,481 farmers who received forecasts. **Uganda** is also developing a SEWAA-linked climate forecasting system for its national early warning infrastructure (UGA-002), although this remains in design phase.

A second cluster focuses on displacement. In **Uganda**, an AI model forecasts refugee arrivals to trigger the Displacement Crisis Response Mechanism (UGA-001; Financial Protection Forum, 2023). The first DCRM activation in 2021 disbursed USD 1.2 million to two host districts. A second activation in 2023 disbursed USD 3.3 million across all 13 eligible districts, prioritising water infrastructure during drought. The model demonstrated over 80% accuracy on unseen data. In May 2024, the Anticipatory Humanitarian Action for Displacement (AHEAD) displacement foresight model predicted a surge in displacement in Akobo County, **South Sudan**. An early response was triggered that prevented an estimated 2,800 people from being displaced, at a reported cost-benefit ratio of EUR 6.60 saved for every EUR 1 spent (SSD-001; ACLED, 2024). The model can predict next year's displacement with an average margin of error down to 6% in its best-performing countries.

No examples were found focusing on AI-enabled forecasting of economic, demographic, health, or other non-crisis trends, such as the micro-simulation and actuarial forecasting common in non-AI social protection planning tools. This gap is notable: AI-enhanced demographic or economic forecasting could inform programme design parameters (e.g. projected caseload growth, contribution adequacy), but no documented deployment applies AI to these functions.

Further social protection applications **less documented in the literature include forecasting to stress-test the adequacy of legal frameworks and anticipate the need for expanded statutory entitlements or new programme categories (e.g. ahead of emerging trends and shocks), or to inform adaptive and risk-informed design parameters. For a full list see Section 1 of *A Taxonomy for AI in SP* (DCI AI Hub, 2026a).**

Important considerations

First, this category is primarily driven by non-government organisation (NGO) and humanitarian organisations (e.g. WFP, OCHA, United Nations High Commissioner for Refugees [UNHCR], etc.), experimenting with forecasting tools developed alongside, rather than within, national government systems. Of seven cases, six are led or co-led by international organisations. This raises capacity and sustainability questions: whether national institutions can maintain, retrain, and govern these models after international funding cycles end; whether the technical expertise built during pilot phases transfers to government staff; and whether forecast-based financing mechanisms can be sustained within national fiscal frameworks rather than donor-funded contingency pools.

Second, what makes this category particularly useful is its pairing with early action triggers and institutional embedding. When a forecast crosses a threshold, pre-agreed protocols trigger automatic disbursement. This removes the delay inherent in discretionary humanitarian response.

Third, this category poses relatively low governance risks where outputs support planning, preparedness and early warning rather than individual-level decisions. Governance sensitivity escalates when forecasts directly trigger anticipatory transfers or automatic benefit adjustments.

Other risks also need to be accounted for: degradation in model accuracy as patterns shift (e.g. due to climate), the consequences of false positives and false negatives, and dependency on data feeds that may themselves be disrupted during the crises they aim to predict. Whether forecast-based systems are fiscally sustainable in the long term and whether they outperform simpler early warning approaches remains unclear.

3.3 Identification, verification and record linkage

Registration | Assessment of needs and enrolment | Provision of payments/ services | Monitoring and evaluation | Case manage-ment | Account-ability | Information system

Note: See Figure 4 above for the full mapping against social protection system functions and refer to *A Taxonomy for AI in Social Protection* for more details (DCI AI Hub, 2026a).

Social protection function

This category covers AI systems that confirm identity, validate claims or credentials, and link records across fragmented systems to create a unified and reliable view of individuals or households. They answer the question: is this person who they claim to be, and do they already exist in the system? Typical applications include biometric authentication (fingerprint, iris, or facial recognition), deduplication to remove ghost or duplicate records from national registries, document verification, and entity resolution that links records across separate databases to create unified citizen profiles. These systems sit at the identity layer of the delivery chain. When they work, they reduce fraud and ensure that benefits reach the right person. When they fail, they lock people out of the system entirely.

What the evidence shows

Across 11 countries, the evidence base includes 13 cases. Perception and extraction is the dominant capability (8 cases), reflecting the visual and biometric processing at the core of these systems.

National biometric identity systems form the largest cluster. In each of these, the AI component is the matching algorithm, typically a one-to-many comparison engine that scores similarity between biometric templates to identify duplicates or verify identity. The capture hardware, data storage, and issuance infrastructure around it are non-AI components that, in most cases, predate or exist independently of the AI matching layer. In **Nigeria**, the National Identification Number (NIN) system uses an Automated Biometric Identification System (ABIS) for duplicate detection at enrolment (NGA-001; Federal Republic of Nigeria, 2023), with 121 million NINs issued as of June 2025. In **Bangladesh**, the Smart National Identity (NID) system (BGD-002; Election Commission of Bangladesh, 2025) employs ABIS for biometric deduplication across the national identity register, achieving over 99% coverage of eligible adults by 2025, with a United Nations Development Programme (UNDP) audit finding of zero ghost voters. Here too the AI contribution is deduplication at enrolment rather than the identity programme itself. In **Pakistan**, NADRA's ABIS (PAK-002) underpins identity verification for approximately 120 million adults and achieved over 99.5% accuracy in fingerprint verification benchmarking. In **Nepal** (NPL-002), over 14 million citizens have been enrolled in the national biometric identity card programme, approximately 90% of the eligible population. **Ethiopia**'s Fayda

Digital ID (ETH-001) has registered over 30 million people as of early 2026 and integrated with over 90 agencies, with biometric deduplication performed by the ABIS layer. In **India**, AI is used for pension proof-of-life (Box 3)[4].

BOX 3. JEEVAN PRAMAAN: AI FOR PENSION PROOF-OF-LIFE IN INDIA

Until recently, millions of pensioners in India had to appear in person at a government office each year to prove they were still alive. Jeevan Pramaan (IND-003; Press Information Bureau, 2024) replaced that requirement with face authentication on a smartphone. The AI component is the face-matching algorithm that compares the submitted selfie against the pensioner's biometric record on file. The Employees' Provident Fund Organisation reports that face-authenticated digital life certificates grew from approximately 210,000 in FY 2022–2023 to 660,000 in FY 2023–2024, a 200% year-on-year increase. The system is now scaled and institutionalised across India's pension infrastructure. This is a different AI application from the ABIS-based enrolment systems above: rather than one-to-many deduplication at registration, Jeevan Pramaan performs one-to-one verification at the point of benefit claim, a use case that may be more easily transferable to other pension systems with existing biometric records.

Humanitarian contexts present a distinct cluster. In **Afghanistan**, UNHCR's Biometric Identity Management System (BIMS) provides biometric identity verification for the Forcibly Returned Afghans Enrollment (FARE) cash assistance programme (AFG-001; MacDonald, 2024), through which approximately 39,000 returnees had received verified cash transfers by early 2024. In **Ethiopia**, the same PRIMES/BIMS infrastructure authenticates refugees at 27 distribution locations (ETH-003; UNHCR, 2018). In the **Democratic Republic of Congo**, WFP's SCOPE platform uses biometric verification at food and cash distributions (COD-001).

Entity resolution forms a third cluster. In Telangana, **India**, the Samagra Vedika platform (IND-005; Al Jazeera, 2024) uses machine learning clustering to create unified citizen profiles by linking records across welfare databases. The government reports that Aasara pension processing flagged 6,625 of 65,693 applications as ineligible, saving an estimated INR 160 million per year. In **Kenya**, the Social Health Authority (KEN-001; Kenya Law, 2023) uses biometric verification at health facilities, with registrations surpassing 25 million by August 2025 and 40 facilities suspended over fraudulent claims.

Further social protection applications less documented in the literature include automated evidence validation (AI-assisted verification of supporting documents submitted with applications), and cross-registry linkage for unified citizen profiles beyond the biometric and deduplication use cases documented above. Entity resolution at scale across fragmented social registries is an area in which techniques exist, but operational deployment remains limited. For a full list see Section 1 of *A Taxonomy for AI in SP* (DCI AI Hub, 2026a).

4 In each of these national biometric identity systems, the AI component is the matching algorithm, typically a one-to-many comparison engine that scores similarity between biometric templates to identify duplicates or verify identity. The biometric capture hardware, data storage, and issuance infrastructure are non-AI components that predate or exist independently of the AI matching layer.

Important considerations

This category is dominated by LIC and LMIC contexts (12 of 13 cases), where foundational identity infrastructure is being built.

Identity verification and record linkage may directly affect rights and entitlements. Errors in identity verification or record linkage can result in wrongful exclusion, denial of benefits, or false fraud flags. Governance sensitivity is highest for biometric systems, given documented cases of exclusion among elderly, disabled, and manual-labour populations whose biometrics may be unreadable. However, the risks are not limited to biometrics: entity resolution systems that link records across databases (such as India's Samagra Vedika) raise distinct concerns about purpose limitation, consent, and the downstream consequences of incorrect linkages.

Documented exclusion patterns skew systematically: women, elderly citizens, and populations with worn fingerprints from manual labour are over-represented in authentication failures, such that errors in identity systems land disproportionately on groups already at higher risk of exclusion from benefits.

Ultimately, the purpose for which the AI-supported identification, verification and record linkage is used matters. As an example, when AI-supported record linkage enables interoperability across existing databases to support registration functions (by making existing data readily available) the governance risk is far less pronounced than when it is being used to support functions that more directly affect rights and entitlements (e.g. payment verification or removal of ghost or duplicate beneficiaries).

3.4 Decision support

Eligibility criteria and qualifying conditions

Assessment of needs and enrolment

Note: See Figure 4 above for the full mapping against social protection system functions and refer to *A Taxonomy for AI in Social Protection* for more details (DCI AI Hub, 2026a).

Social protection function

This category covers AI systems that assist human decision-makers in determining eligibility, setting benefit levels, or configuring benefit and services packages, typically through rule-based or model-assisted decision support. Unlike vulnerability assessment, which scores/ranks populations, decision support operates at the point of individual case determination. The AI produces a recommendation, a risk score, or a classification that a caseworker or adjudicator uses to inform their decision. The human retains formal authority. Common applications include household socioeconomic classification to determine household/individual eligibility for (proxy) means-tested programmes, clinical decision support for occupational health assessments, and profiling tools that help employment counsellors tailor services to individual circumstances.

What the evidence shows

Only six cases across five countries have decision support as their primary use. That thinness is itself a finding (see Important Considerations below).

The cases that do exist are strikingly diverse. In **Colombia**, the SISBEN IV system (COL-001; López & Castaneda, 2020) collects household socioeconomic status for targeting across at least 18 social programmes, with plans to reach 40.5 million people, roughly 84% of the population. The AI component is a supervised classification model that groups households into four welfare categories (extreme poverty, moderate poverty, vulnerable, non-poor) based on survey responses, replacing the deterministic rule-based scoring used in earlier SISBEN iterations. Caseworkers use the classification to determine eligibility across programmes rather than making eligibility judgments from scratch. The system has been scaled and institutionalised. In **Chile**, the ISL/ACHS eTóraxLaboral system (CHL-001) uses deep learning to analyse chest X-rays for pneumoconiosis in occupational health, with a clinical validation study reporting a positive likelihood ratio of 23 across 2,300 radiographs. In the **United States**, the Social Security Administration's Quick Disability Determinations model (USA-007; US Social Security Administration, 2019) expedites a subset of likely-favourable disability claims, operating at a national scale. Also in the US, the Artificial Intelligence Adjudicator Assistance (AIAA) prototype (USA-004) is exploring whether AI can help unemployment insurance adjudicators sort cases, although it remains at the research stage. In **Brazil**, the Meu INSS system (BRA-002; Dataprev, 2020) has deployed AI-enabled claims adjudication to address a backlog of approximately 2 million pending social insurance requests.

Further social protection applications **less documented in the literature (but possible) include tailored benefit/service package configuration at the individual or household level (e.g. transfer value determination, benefit/service type). The potential would be high, e.g. for tailored support to people with disabilities or those in other vulnerable categories. For a full list, see Section 1 of *A Taxonomy for AI in SP* (DCI AI Hub, 2026a).**

Important considerations

Several factors contribute to explaining why this use case category remains small:

- Governments are cautious about letting algorithms influence who receives benefits (or decide other crucial benefit/services parameters), and legal frameworks in many jurisdictions require a human decision-maker, which limits the formal role an AI system can play.
- Political sensitivity is high – governments face reputational risk if algorithmic recommendations are perceived to deny benefits to eligible people. This is particularly problematic in contexts where trust in government systems is already fragile.
- Similarly, accountability is harder to assign when a decision is shaped by a model that neither the caseworker nor the applicant fully understands. In fact, in the few cases where AI is already operating to support eligibility determination it tends to operate behind advisory labels, even when the recommendation is rarely overridden, because of the widely studied phenomenon of 'cognitive surrender', whereby humans assume that AI must be right, as it knows something they do not.

This 'thinness' of evidence also implies that options to use decision support functions beyond eligibility determination

(e.g. for setting transfer values, or choosing optimal benefit type, etc.) **are under-explored by governments globally,** which is another important finding.

One interesting exception is **Estonia**'s Otsustustugi (OTT) decision-support tool (EST-003; Vihalemm et al., 2025), which assists employment counsellors in tailoring service pathways to individual unemployed clients. OTT predicts each person's probability of finding employment within 180 days and their risk of returning to unemployment, then segments clients into risk categories to guide counsellor decisions about service intensity and type, making it one of the few cases where AI informs benefit/service package configuration rather than eligibility alone. Ninety-three per cent (93%) of welfare specialists considered its scoring accurate, although reception among counsellors was mixed, with some reporting that the tool improved workload planning, while others questioned its added value over professional judgment.

Having said this, all documented cases use human-in-the-loop (HITL) oversight.

AI assists, but does not determine. The primary risk is that decision-support tools become de facto decision makers when caseworkers routinely accept recommendations without independent assessment. Whether AI-assisted decision support improves the accuracy, speed, or consistency of eligibility determinations compared to unaided human judgment is to-date entirely undocumented.

3.5 Data quality and anomaly detection

Legal and policy framework | Financing | Functional and technical capacities | Governance and coordination | Registration | Provision of payments/ services | Assessment of needs and enrolment | Monitoring and evaluation | Information system

Note: See Figure 4 above for the full mapping against social protection system functions and refer to *A Taxonomy for AI in Social Protection* for more details (DCI AI Hub, 2026a).

Social protection function

This category covers AI systems the primary purpose of which is to improve the accuracy, completeness, and consistency of data within social protection registries and information systems. In other words, AI can support the preparation, validation and quality enhancement of data, while also detecting unusual or suspicious patterns, prioritising cases for data quality checks or investigation. These systems work on the data itself rather than on programme decisions, although their outputs directly affect the reliability of every downstream function, from targeting to fraud detection to payments.[5]

What the evidence shows

No cases in our evidence base had data quality as their primary purpose. The AI techniques exist: duplicate detection, anomaly flagging, and data validation all appear as secondary functions within compliance systems (section 3.9), identification/verification systems (section 3.3), and operational automation tools (section 3.8). For example, **Nigeria**'s ABIS performs biometric deduplication upon national ID enrolment (NGA-001), while in **India** Telangana's Samagra Vedika links records across welfare databases to flag duplicate or ineligible beneficiaries (IND-005). But in every instance, data quality is a means to another end, not the primary objective.

Further social protection applications that are not yet found in the literature as primary functions include operations to enhance the data environment (e.g. structuring unstructured data or supporting data imputation), transaction anomaly detection, as well as data consistency checks and quality audits (e.g. automated detection of duplicate records, flagging of anomalous or outdated entries, analysis of registry coverage gaps and data completeness, validation of data against external sources, and identification of deceased beneficiaries still listed as active, etc.). They could also include network or collusion analytics and audit prioritisation.[6] For a full list see Section 1 of *A Taxonomy for AI in SP* (DCI AI Hub, 2026a).

5 Note the use case boundary with the compliance and integrity use case below, as defined in *A Taxonomy for AI in SP* (see Section 1.2, DCI AI Hub, 2026a): "this use case (vis-a-vis the compliance and integrity one below) focuses on data and system integrity, i.e. answering questions such as – is this record accurate? Is this transaction pattern unusual? Is this dataset complete? The primary purpose here is data or system integrity".

6 Once again, many of these can enable detection of error and fraud (via AI analysing data patterns to flag irregularities), but these cases are classified under the 'compliance' primary use case.

Important considerations

The absence of primary cases is itself an analytical finding about how AI functions are categorised and reported. Data quality work is systematically invisible because it is embedded within larger systems rather than deployed as a standalone tool. When Nigeria's ABIS deduplicates identity records, the primary purpose is recorded as identification. When a fraud detection system flags anomalous payment patterns, the primary purpose is compliance. Data quality improvement happens as a by-product, and because no institution frames it as the headline objective, it did not appear as a primary use case in any of the documentation reviewed.

This matters because data quality problems in social protection registries are widely acknowledged, particularly in LMIC contexts where registries commonly contain duplicate records, outdated addresses, and deceased beneficiaries, for example. The techniques to address these problems exist and are already in use for other purposes. The gap is not technical. It is institutional: no agency has yet made a standalone business case for AI-enabled data quality as a primary investment, and no funder has prioritised it as a discrete intervention.

Whether framing data quality as a visible, primary function would attract resources and attention is an open question, with practical implications for registry modernisation.

3.6 User communication and interaction

Governance and coordination

Outreach and commu-nications

Registration

Account-ability

Note: See Figure 4 above for the full mapping against social protection system functions and refer to *A Taxonomy for AI in Social Protection* for more details (DCI AI Hub, 2026a).

Social protection functions

This category covers AI systems that communicate directly with beneficiaries, applicants, staff or the general public on behalf of a social protection institution. They answer questions (in multiple languages), provide guidance, take appointments, issue documents, or proactively check on vulnerable individuals. Typical applications include chatbots on government websites or messaging platforms, voice assistants ('voicebots') accessible by phone, and automated outreach systems (personalised messaging) that contact people who may not engage with services on their own. These systems handle the information and access layer of service delivery. They do not make eligibility or benefit decisions, but the quality of the information they provide can shape whether people apply for, understand, or retain their entitlements.

What the evidence shows

With 22 cases across 15 countries, user communication is the largest category in the evidence base. The dominant capabilities are perception and extraction (14 cases) and LLMs (8 cases).

Government chatbots and virtual assistants account for the majority of cases. **India**'s Pradhan Mantri Kisan Samman Nidhi (PM-KISAN) AI chatbot Kisan e-Mitra (IND-001; Press Information Bureau, 2023b) is an LLM-powered voice chatbot serving one of the world's largest cash transfer programmes, handling over 20,000 queries daily as of April 2025, with 9.5 million queries answered to date. In **Argentina**, Buenos Aires, the Boti chatbot (ARG-001; Microsoft, 2024) handles 2 million WhatsApp queries per month with an 82% automated resolution rate. **Brazil**'s INSS Helô (BRA-001) has processed over 100 million interactions across the Meu INSS platform. **Norway**'s NAV Frida (NOR-001) answered over 270,000 citizen enquiries during COVID-19, resolving approximately 80% without human escalation.

LLM-powered systems are the newest wave, although some earlier examples also use foundation models. Of the 22 cases in this category, 8 use LLMs as their primary capability and 14 use traditional perception and extraction techniques. **Estonia**'s Bürokratt (EST-001; European Commission, 2025) is building a cross-government virtual assistant network. The United Kingdom's (**UK**'s) GOV.UK Chat (GBR-002; Central Digital and Data Office, 2025) is a Retrieval-Augmented Generation (RAG)-based chatbot piloting across government information pages, with initial testing showing roughly 70% of users finding responses useful and accuracy reaching 90% after improvements. In Peterborough, UK, Hey Geraldine (GBR-006) replicated a specific local government officer's advice so accurately that staff reported they thought they were speaking with the real person. **Albania**'s Diella (ALB-001), deployed on the eAlbania government services platform, has processed 972,000 citizen interactions and issued 36,000 official documents. While Diella serves as a general government virtual assistant, its social protection relevance lies in its ability to guide citizens through social assistance applications and issue documents related to benefit entitlements.

Social insurance chatbots form a distinct cluster. **Belgium**'s ONEM Ori (BEL-001) handles unemployment benefits queries. **Argentina**'s SRT Julieta (ARG-002) answers occupational risk enquiries with 93% accuracy. **Mexico**'s ISSSTE ASISSSTE (MEX-002) manages 41,000 daily appointments. **Korea**'s NPS deploys AI kiosks with a real-time interpreter for foreign contributors (KOR-006).

The remaining eight cases, spanning **Saudi Arabia**, **Azerbaijan**, **Bangladesh**, **India**, **Singapore**, and **Uganda**, are predominantly structured chatbots handling benefits enquiries or appointment bookings.

Social care and wellbeing monitoring forms a distinct and emerging sub-cluster within user communication. **South Korea**'s CLOVA CareCall (KOR-003; The Pickool, 2025) makes regular AI-enabled phone calls to elderly people living alone, asking about meals, health, and daily life, and flagging concerns to social workers. Service areas show a 44.2% reduction in solitary deaths, compared to non-service areas, and a 9.2% reduction in emergency room visits. The Hyodol companion robot (KOR-004) provides physical companionship and care reminders to elderly people living alone, with a study of 278 users finding significant decreases in depression scores. Both cases sit at the boundary between social protection and social care, a space where AI-enabled proactive outreach may be particularly impactful given ageing populations and the global shortage of care workers.

Further social protection applications less documented in the literature include explainable decision notices (using AI to generate clear, personalised explanations of benefit decisions in plain language) and personalised proactive messaging that reaches people who may be eligible but have not applied. Translation and accessibility tools (text-to-speech, sign language) for social protection communications are also largely undocumented as primary deployments, despite clear potential. For a full list see Section 1 of *A Taxonomy for AI in SP* (DCI AI Hub, 2026a).

Important considerations

This category spans income levels more evenly than other use categories. HIC cases (12) tend to use LLMs for open-ended conversations. LMIC and LIC cases (4) more commonly use structured natural language processing or rule-based systems augmented with AI classification.

Using AI for communication and interaction with users is popular for a good reason: decision criticality is generally low to moderate. These systems provide information rather than making rights-impacting decisions. But the quality of that information matters: if a beneficiary relies on a chatbot response (or translation) to decide whether to apply for a benefit, an inaccurate answer has real consequences. Also, governance sensitivity increases when automated communication replaces, rather than supplements, human service channels, fuelling the already significant shortage of frontline workers serving the social protection sector. Ultimately, do chatbots reduce administrative burden, or do they simply redirect it? What impacts do chatbots have on user trust? The evidence base does not yet answer these questions.

Beyond decision criticality, other risks warrant attention. Language quality and accuracy are critical: a chatbot that provides incorrect benefit eligibility information or misinterprets a query in a second language can lead to missed entitlements or unnecessary applications. Accessibility risks arise when chatbot deployment coincides with a reduction in human service channels, particularly for elderly users, people with disabilities, or those with limited digital literacy. Data privacy concerns emerge when chatbot interactions generate detailed records of citizens' circumstances, queries, and vulnerabilities, data that is valuable for service improvement, but sensitive if mismanaged. Several cases in the evidence base report high automated resolution rates (80–93%), but what happens to users in the remaining 7–20% who need human assistance is largely undocumented.

3.7 Matching and recommendation

- Governance and coordination
- Functional and technical capacities
- Benefits/ services package
- Outreach and communications
- Assessment of needs and enrolment
- Provision of payments/ services
- Case management
- Information system

Note: See Figure 4 above for the full mapping against social protection system functions and refer to *A Taxonomy for AI in Social Protection* for more details (DCI AI Hub, 2026a).

Social protection functions

This category covers AI systems that connect individuals with opportunities, services, or programmes suited to their profile. They take information about a person, such as their skills, qualifications, work history, or circumstances, and match it against a database of available options. Typical applications include job matching platforms in public employment services, skills profiling tools that identify transferable competencies, and recommendation engines that suggest training programmes or social services. The output is a ranked list of options or a personalised recommendation. These systems support the case management and referral functions of social protection, particularly in labour market programmes.

What the evidence shows

The evidence base yields seven matching and recommendation cases, drawn from six countries. The dominant AI capabilities are classification and ranking, reflecting the task of connecting individual profiles to available opportunities.

Job matching in public employment services is the primary application. In **Korea**, 'The Work' (KOR-002; Kim, 2020) uses the national job information platform to make AI-enabled job recommendations. Between December 2018 and November 2019, approximately 7,600 individuals gained verified employment through companies recommended by the system, while average job search time fell from around 10 minutes across multiple sites to about 5 minutes. In **Japan**, Hello Work (JPN-001) is piloting AI-enhanced job matching at 10 locations to improve on the current 25.9% matching success rate. **Singapore**'s MyCareersFuture portal (SGP-001; GovTech Singapore, 2025) matches jobseekers to vacancies using AI-enabled skills-based recommendations. Since 2018, it has assisted over 2.7 million jobseekers and helped more than 77,000 individuals secure employment. Monthly job listings expanded from 23,000 to over 35,000 across 23 industries through partnerships with Indeed, LinkedIn, and other platforms. It is now scaled and institutionalised as core national infrastructure. Singapore has also launched the SkillsFuture Careers and Skills Passport (SGP-007), through which 315,000 Singaporeans have accessed AI-enabled skills profiling and learning recommendations, as of April 2025.

Skills profiling for refugees and displaced populations is an emerging application. In **Lebanon**, the ILO PROSPECTS programme used SkillLab's AI platform (LBN-001; SkillLab, 2025) to profile 500 Lebanese and Syrian individuals across three centres, generating multilingual CVs and identifying previously unarticulated skills. In **Egypt**, the same SkillLab tool (EGY-001; SkillLab, 2025) was deployed with a tracer study finding 64% of 400 users created a CV and 76% rated the app good or excellent for skills identification. In **India**, the Skill India Assistant (IND-009) provides multilingual chatbot guidance on the national skilling ecosystem.

Further social protection applications less documented in the literature (but possible, and important) include those focused on a) tailored case management and referrals, e.g. for social assistance, social insurance or social care services; b) matching staff profiles to training opportunities, recommending tailored learning journeys and supporting career pathway development for the social worker workforce; c) recommendation systems that match specific delivery chain solutions (e.g. specific communication channels, payment modalities, etc.) to the profiles of specific under-reached population segments. For a full list see Section 1 of *A Taxonomy for AI in SP* (DCI AI Hub, 2026a).

Important considerations

All seven cases in this category involve labour market programmes. The absence of matching and recommendation systems in social assistance (and social care services) or social insurance is notable.

The concentration in HIC and UMIC contexts raises questions about applicability in low-income labour markets, where formal vacancy databases may not exist and risks would over-ride benefits.

Governance concerns exist, but are moderate, and include algorithmic fairness in profiling, transparency of matching criteria to jobseekers, and the risk that profiling scores channel resources away from those assessed as having poor prospects. Governance sensitivity escalates when matching outputs directly determine programme referrals, job placements or service access without caseworker review. Whether or not AI-assisted matching outperforms experienced caseworker judgment remains untested.

A further consideration is whether matching and recommendation systems adequately serve populations with non-standard career histories, informal workers, care-givers returning to the workforce, people with disabilities, or those with interrupted education. Most matching systems are trained on formal labour market data and may systematically underserve populations whose skills and experience are not legible in structured databases. The SkillLab deployments in **Egypt** and **Lebanon** (EGY-001, LBN-001) represent an explicit attempt to address this by profiling transferable skills for refugee populations, but this approach has not been tested at scale.

3.8 Operational and process automation

Note: See Figure 4 above for the full mapping against social protection system functions and refer to *A Taxonomy for AI in Social Protection* for more details (DCI AI Hub, 2026a).

Social protection functions

This category covers AI systems that handle internal administrative tasks within social protection agencies, enhancing productivity and back-office efficiency. These systems process documents, classify incoming correspondence, transcribe hearings, summarise case files, route workflows, or assist staff with routine tasks. Unlike decision support, these systems do not directly advise on eligibility or benefit levels. They reduce the administrative burden on staff so that human effort can focus on judgment-intensive work. Typical applications include automated claims processing, document extraction and classification, AI-assisted quality review, and generative AI tools that draft correspondence or summarise case notes for caseworkers.

What the evidence shows

Operational automation accounts for 11 cases across 8 countries, and 2 cases use LLMs as their primary capability, a lower share than in user communication (8 of 22), but notable because LLM use in back-office functions is new and growing rapidly.

Document processing and classification form the largest cluster. In **Germany**, the Cultural Funds Platforms (DEU-002) use a hybrid pipeline of classical machine learning and optical character recognition (OCR) to process applications for targeted cultural subsidies in Hamburg, a fee-waiver/subsidy programme classified as social assistance. Germany's Federal Employment Agency operates ADEST (DEU-004), which extracts structured information from unstructured job advertisements. In **Finland**, Kela's AI platform (FIN-001) supports attachment pre-processing, customer feedback analysis, and call transcription across the national social insurance system. In **Chile**, SUSESO (CHL-003) is developing machine-learning-based medical insurance claims processing. **Austria**'s KAI system (AUT-001; Dachverband der Sozialversicherungsträger, 2023) semi-automatically processes reimbursement claims for medical services across the country's social insurance system. By October 2022, two thirds of all reimbursement requests were processed semi-automatically, handling approximately 8% annual growth in claim volumes without additional staffing. The system uses perception and extraction to read claim documents, classify them, and route them for processing.

Generative AI for caseworker assistance is a rapidly emerging application. The **UK**'s Department for Work and Pensions has run five structured generative AI pilot projects, each with director-general sponsorship, testing LLM applications across the welfare system including case note summarisation and staff assistance (GBR-003; Trendall, 2024), reporting reductions in documentation time during 2024–2025. In the **United States**, the Social Security Administration's HeaRT system (USA-010; US Social Security Administration, 2025) uses AI to record and transcribe disability hearings on a national scale, projecting annual savings of approximately USD 5 million. The Social Security Administration's Insight system (USA-003) supports decision-quality review in disability adjudication. In **India**, the Centralised Public Grievance Redress and Monitoring System (CPGRAMS) system (IND-004; Press Information Bureau, 2023a) uses AI classification to triage and route citizen grievances, reducing average disposal time from 32 days in 2021 to 12 days in 2024 with a 98% resolution rate on over 2.4 million grievances. In Employment and Social Development **Canada**'s (ESDC's) machine learning system (CAN-001; Canada School of Public Service, 2023) triages employment insurance recalculation workloads, processing over 40,000 claims with a 90% accuracy rate in identifying no-change cases.

Further social protection applications less documented in the literature as primary functions include workload and resource forecasting (e.g. predicting operational demand, caseload volumes and staffing requirements). No cases in the evidence base have this as a primary function, although the Canada EI case (CAN-001) includes workload triage as a secondary element. Given the operational pressures facing most social protection agencies, this represents a significant gap between potential and practice. For a full list see Section 1 of *A Taxonomy for AI in SP* (DCI AI Hub, 2026a).

Important considerations

This category is predominantly seen in HICs (10 of 11 cases), reflecting the digital maturity required for automation to deliver gains.

A sub-category within operational automation is workload and resource forecasting: meaning the prediction of operational demand, caseload volumes, and staffing requirements for planning and resource allocation. Yet no cases in the evidence base fall under this sub-category. Whether this reflects genuine under-deployment or the embedding of such functions within broader operational systems is unclear, but the gap is notable given the widespread challenges of caseload management and resource planning in social protection agencies globally.

Decision criticality is generally low to moderate, as these systems assist internal processes. But errors in document processing or case summarisation can have downstream effects on decisions, and the evidence for actual efficiency gains, as opposed to vendor claims, is thin.

3.9 Compliance and integrity

Provision of payments/ services

Account- ability

Case manage- ment

Monitoring and evaluation

Note: See Figure 4 above for the full mapping against social protection system functions and refer to *A Taxonomy for AI in Social Protection* for more details (DCI AI Hub, 2026a).

Social protection functions

This category covers AI systems that monitor adherence to programme rules, requirements and conditions, including management of complaints and appeals (grievances). In other words, these are systems that can help detect, predict, or prevent fraud, error, and non-compliance within social protection programmes. They can be designed to score individuals or claims by risk of irregularity, flag anomalous patterns in payment or claims data, and monitor whether beneficiaries meet programme conditions. Typical applications thus include algorithmic risk scoring for welfare fraud investigation, anomaly detection in health insurance claims, and predictive models that prioritise audit targets. The output is a flag, a score, or a referral for human investigation.

This category has the highest governance stakes, as the processes in it are fundamentally rights-critical: false positives can lead (and have led) to benefit suspension, financial hardship, and the wrongful accusation of vulnerable people (see Important Considerations below).

What the evidence shows

Compliance and integrity is represented by 14 cases across 12 countries, with anomaly detection (8 cases) and prediction (3 cases) as the dominant AI capabilities.

This is the category where most 'failures' have occurred, because of human rights implications: three of the five suspended or halted cases in the entire evidence base are compliance systems. Beyond the renowned Netherlands SyRI case (Box 4), the **Netherlands** also suspended a second system. Rotterdam's municipal welfare fraud prediction algorithm (NLD-002; Lighthouse Reports, 2023c) scored approximately 30,000 social assistance recipients annually and referred 1,000–1,500 for investigation each year from 2017 to mid-2021. It was suspended after a Rekenkamer Rotterdam audit identified discriminatory profiling. In **Sweden**, the Försäkringskassan's machine learning risk-scoring system (SWE-001; Ikumi, 2025) was suspended after independent journalistic and regulatory scrutiny revealed that women were overrepresented by a factor of 1.5 and people with foreign backgrounds by a factor of 2.5 relative to the general population. In 2017, the system selected

5,082 of approximately 977,730 applications for investigation. In **Ireland**, Cogent facial recognition software for welfare fraud detection (IRL-001; O'Brien, 2015) was also suspended.

BOX 4. THE NETHERLANDS' SYRI CASE

On 5 February 2020, the District Court of The Hague ruled that the Dutch government's System Risk Indication (SyRI), a tool used to detect welfare fraud by linking and analysing data from multiple public sources (employment, tax, debts, housing, etc.), was unlawful. The Court found that the legislation enabling SyRI violated the right to privacy protected under Article 8 of the European Convention on Human Rights (ECHR) (Privacy International, 2020).

SyRI is included in this review because it used statistical risk indicators and pattern matching, meeting this review's boundary definition. However, the system can be more accurately characterised as rule-based algorithmic risk scoring rather than machine learning. In fact, The Hague District Court's ruling – which remains the most significant legal precedent for AI in social protection globally – addressed algorithmic decision-making broadly, not AI specifically.

Having said this, several AI-based fraud detection and risk-scoring systems remain operational around the globe, reporting important 'efficiency' achievements. In **France**, the National Family Allowance Fund's (CNAF's) algorithmic risk scoring (FRA-001; Lighthouse Reports, 2023a) generates more than 13 million suspicion scores monthly, with 7 of every 10 fraud investigations triggered by algorithmic flagging. France's CNAM (FRA-003) detected EUR 628 million in healthcare fraud in 2024, using risk-scoring algorithms that analyse claims data patterns to flag anomalous submissions for human investigation, a 35% increase from 2023. In **Korea**, the National Health Insurance Service machine learning system (KOR-001) detected 567,820 cases worth approximately USD 174 million between 2014 and 2021. In **Indonesia**, BPJS Kesehatan's DEFRADA system (IDN-001; ISSA, 2018) contributed 25–30% of total national health insurance efficiency gains in 2017 and reported preventing approximately IDR 10.5 trillion (USD 740 million) in fraudulent claims in 2019. In **India**, AB PM-JAY's fraud detection (IND-006; Press Information Bureau, 2022) has identified over 342,000 fraud cases across approximately 66.6 million processed claims, with 3,167 hospitals found guilty of irregularities. In **Ghana**, the National Health Insurance Authority (NHIA) (GHA-001; Ghana Data Protection Commission, 2025) reported saving approximately GHS 9.5 million through AI-enabled fraud detection in 2023. In the **UK**, the Department for Work and Pensions' (DWP's) Universal Credit fraud model (GBR-005; DWP, 2025) is approximately three times more effective than random selection, saving an estimated GBP 4.4 million since May 2022. **Spain**'s Instituto Nacional de la Seguridad Social (INSS) 'AI Doctor' (ESP-001) scores sick leave fraud risk, although after five years of operation it has been described as limited by chronic underfunding and inspector staff shortages. In **Germany**, KIRA (DEU-001) is piloting risk-based employer audit selection. In the **US**, the CalWORKs Data Mining Solution in Los Angeles (USA-006) identified collusive fraud rings with an 85% hit rate.

Further social protection applications **less documented in the literature include conditionality verification (automated monitoring of whether beneficiaries meet programme conditions such as school attendance or health check-ups) and systematic grievance pattern detection (identifying systemic issues from complaint data rather than individual case triage). For a full list see Section 1 of *A Taxonomy for AI in SP* (DCI AI Hub, 2026a).**

Important considerations

This category is concentrated in HIC settings (11 of 14 cases), posing questions about whether deployment in settings with lower governance standards could elevate risks further than they already are.

Decision criticality for this use case is very high: fraud or conditionality-related investigations can lead to benefit suspension, payment recovery, and prosecution. The consequences of false positives, when people are incorrectly flagged, include financial hardship, stigma, and erosion of trust in the social protection system. The above-mentioned SyRI Dutch childcare benefits scandal (*toeslagenaffaire*), excluded from our 99 cases because its algorithmic profiling system (FSV) was designed and deployed prior to this review's 2020–2025 scope, further illustrates what can happen when algorithmic risk scoring operates without adequate safeguards. In this case, more than 20,000 families were wrongly accused of childcare benefit fraud, many forced to repay tens of thousands of euros, and the resulting political crisis brought down the Dutch government in January 2021. Similar outcomes were found in Australia's Robo-debt scandal.

So why is algorithmic fraud detection so popular despite the governance risks? Three factors converge: fiscal pressure (fraud represents quantifiable losses that can justify investment), technical feasibility (anomaly detection in structured claims data is a well-established ML application), and political appeal (governments can demonstrate action against fraud). The more instructive question is why some systems continue operating, while others are halted. Of the five suspended cases in the entire evidence base, three are due to compliance (**NLD-001**, **NLD-002**, **SWE-001**), plus Ireland's facial recognition system (**IRL-001**). In each halted case, the common factor was not technical failure, but external accountability: court rulings (Netherlands), independent audit findings (Rotterdam, Sweden), and public and media scrutiny (Ireland). Systems that continue operating – France's CNAF, Denmark's Udbetaling Danmark (UDK), the UK's DWP – tend to operate in institutional environments where such external scrutiny has been less intense or where proactive transparency measures (such as the UK's published fairness assessments) have provided a degree of public legitimacy. Whether this distinction reflects genuine governance quality or simply the absence of challenge is an open question.

TABLE 2. THE FIVE HALTED AI SYSTEMS IN THE EVIDENCE BASE, AND THE ACCOUNTABILITY MECHANISM THAT ENDED EACH

CASE	COUNTRY	PROGRAMME	YEAR ENDED	TRIGGER
NLD-001	Netherlands	SyRI, System Risk Indication	2020	The District Court of The Hague struck down the legal basis on privacy grounds; no proven fraud case in the flagged deployments.
NLD-002	Netherlands	Rotterdam Municipal Social Assistance	2021	A Lighthouse Reports/WIRED investigation exposed discriminatory profiling; a subsequent audit confirmed bias against women and residents with foreign backgrounds.
SWE-001	Sweden	Försäkring-skassan VAB risk scoring	2023	A journalistic investigation and regulatory review found women were over-flagged 1.5x and people with foreign backgrounds 2.5x, with 5,082 of 977,730 applications selected in a single year.
IRL-001	Ireland	Cogent Facial Image Matching (welfare fraud)	2023	A Data Protection Commission audit found inadequate lawful basis and unreliable identity-matching performance.
DNK-002	Denmark	Gladsaxe Model (predictive profiling of vulnerable children)	2018	A national data-protection and ethics review halted the pilot before wider rollout, citing chilling effects and insufficient evidentiary basis for predictive profiling of children.

What share of AI-flagged individuals turn out to be genuine cases of fraud?

The evidence base rarely answers this question. The UK's DWP is a notable exception: its published fairness assessment of the Universal Credit advances model (DWP, 2025) reports accuracy and demographic breakdown metrics. Indonesia's BPJS Kesehatan has published efficiency gain figures through the International Social Security Association (ISSA). However, in most cases, false positive rates are either not measured, not reported, or not broken down by demographic group. The independent assessment of false positive rates by demographic group is almost entirely absent, as is evidence on whether AI-based fraud detection actually deters fraud, or merely shifts it to less visible channels.

As one last point, the DCI AI Hub's *A Taxonomy for AI in SP* (2026a) includes grievance and complaint management within this category, but no documented case has AI-enabled grievance handling as its primary function. The closest analogue is India's CPGRAMS (IND-004), which uses AI to triage and route citizen grievances, but this is classified as operational automation. AI-enabled grievance mechanisms, for instance, automated first-response to complaints, intelligent routing of appeals, or pattern detection across complaint data to identify systemic issues, represent a significant gap in both deployment and documentation.

3.10 Policy analysis, learning and M&E

Legal and policy framework

Financing

Benefits/ services package

Provision of payments/ services

Monitoring and evaluation

Note: See Figure 4 above for the full mapping against social protection system functions and refer to *A Taxonomy for AI in Social Protection* for more details (DCI AI Hub, 2026a).

Social protection functions

This category covers AI systems that support the design, evaluation, or oversight of social protection policies and programmes. These systems analyse programme performance, simulate the effects of policy changes, detect bias in existing algorithms, or extract insights from large volumes of administrative or evaluation data. Unlike operational tools, these systems are not embedded in day-to-day service delivery. They sit in the policy and learning layer, informing decisions about how programmes should be designed, adjusted, or monitored. Typical applications include microsimulation enhanced by machine learning, natural language processing analysis of programme evaluation text, and algorithmic auditing tools that test existing AI systems for fairness.

What the evidence shows

Three cases across three countries use AI for policy analysis, making this the smallest category, but one with clear potential for growth as techniques mature.

The three cases are diverse. In the **United Kingdom**, UKMOD-Essex (GBR-004; Frimpong & Richiardi, 2025) uses machine learning to enhance regional tax-benefit microsimulation, generating synthetic datasets for policy modelling. Macro-validation shows strong alignment with external benchmarks. In **Mexico** (MEX-001), natural language processing text mining analyses free-text performance evaluation justifications across social programmes, enabling the large-scale identification of common implementation issues. In **Togo** (TGO-004; Aiken et al., 2022), researchers conducted model performance monitoring and bias detection on the Novissi poverty-targeting algorithms, finding that phone-based machine learning targeting reduced exclusion errors by 4–21% relative to geographic alternatives, although traditional PMT outperforms machine learning where registry data is available.

Further social protection applications less documented in the literature include: a) AI-assisted microsimulation to assess distributional and financial impacts of proposed policy, legal or regulatory changes before adoption (with impacts on design-choices), b) AI-generated dashboards and performance analytics, and c) out-of-the-box approaches such as system stress-testing under varied scenarios, bias/fairness monitoring, counterfactual analysis and synthetic evaluation. For a full list see Section 1 of *A Taxonomy for AI in SP* (DCI AI Hub, 2026a).

Important considerations

The DCI AI Hub's *A Taxonomy for AI in SP* (2026a) identifies potential applications in this category that are not yet documented in the evidence base

including AI-assisted regulatory impact assessment, automated programme evaluation synthesis, and real-time fairness monitoring of deployed AI systems. These represent areas where techniques are maturing rapidly, but operational deployment in social protection contexts has not yet been documented.

Why are there only three cases of policy analysis, learning and monitoring and evaluation (M&E), despite the clear benefits and low governance criticality?

Policy analysis AI may be embedded in research institutions rather than operational agencies, making it less visible as a social protection deployment. It may also reflect the recency of applicable techniques. The evidence gaps encompass nearly the entire category: whether or not AI-assisted policy analysis actually improves decision making, and whether or not bias detection tools are feasible in operational settings, remain open questions.

CHAPTER 4. CROSS-CUTTING PATTERNS AND GOVERNANCE

The 99 uses cases in this review do not exist in isolation. Many systems serve multiple functions, and the secondary use case assignments reveal how AI deployments connect across the social protection delivery chain. These connections matter for practitioners, because a failure or bias in one system propagates downstream through others.

The most common overlaps are between user communication and operational automation (chatbots that also route cases), identification and compliance (biometric verification that feeds fraud detection), and vulnerability assessment and decision support (poverty maps that inform eligibility).

4.1 Income and geography patterns

A clear income-geography gradient emerges from the data.

LIC and LMIC cases are concentrated in 'upstream' delivery chain functions: e.g. vulnerability assessment, where international organisations fund poverty mapping, and identification, where foundational ID systems are being built with development assistance.

HIC cases are concentrated 'downstream': e.g. fraud detection, operational automation, and policy analysis. This reflects both funder priorities and available digital infrastructure.

The upper-middle-income gap deserves emphasis. Only 11 of 99 cases are from UMICs. Brazil, Mexico, Argentina, and Colombia each have one or two cases, but China, Turkey, Thailand, and South Africa are absent entirely. We suspect this is partly a documentation gap rather than an absence of AI deployment, but it limits the evidence base for the income group with the largest number of social protection beneficiaries globally.

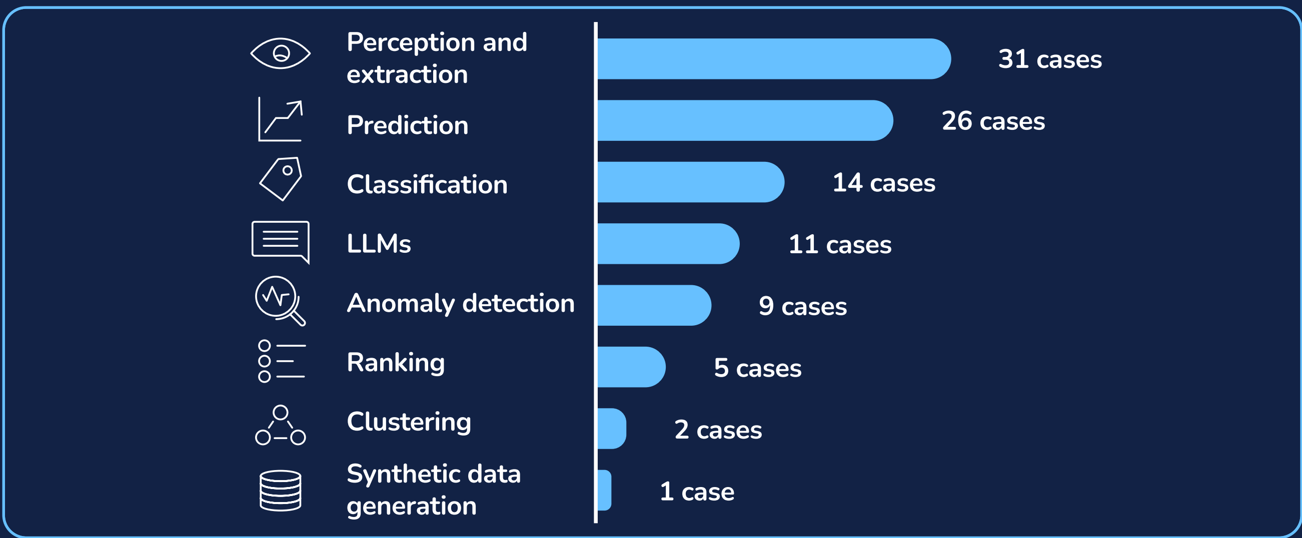
4.2 AI capability distribution

Perception and extraction is the most common primary capability (31 cases), driven by biometric verification, document processing, and chatbot natural language processing. Prediction ranks second (26 cases), spread across vulnerability assessment, forecasting, and fraud risk scoring. Classification (14 cases), LLMs (11 cases), and anomaly detection (9 cases) follow. Clustering (2 cases), ranking (5 cases), and synthetic data generation (1 case) are rare.

The same capability often appears across categories. Although technical expertise may be transferable, the governance requirements differ sharply.

Prediction underpins poverty mapping in Togo, fraud detection in France, and displacement forecasting in South Sudan. Yet this technical transferability masks the different governance requirements. A prediction model for poverty mapping in Togo operates with aggregate-level outputs, with errors affecting resource allocation, but not individual rights. The same predictive capability deployed for fraud detection in France generates individual-level suspicion scores that can trigger benefit suspension and investigation. The safeguards, transparency requirements, appeal mechanisms, bias auditing, and human override protocols must differ commensurately.

FIGURE 5. AI CAPABILITY DISTRIBUTION



4.3 Types of social protection and implementation

Social assistance dominates with 56 cases, followed by social insurance with 33 and labour market programmes with 10. Social insurance cases are concentrated in compliance and operational automation. Labour market cases cluster in matching and recommendation. The under-representation of labour market programmes, with only 10 cases across all categories, is a clear gap.

Classical machine learning accounts for 60 of the 99 cases. Deep learning (19 cases) is concentrated in perception tasks. Foundation models (11 cases) are predominantly found in chatbots and back-office automation.

4.4 Data sovereignty patterns

The distribution discussed above has practical implications for data sovereignty: the dominance of classical machine learning means that most systems can run on modest domestic infrastructure, reducing the need for cross-border data transfers that foundation model deployments more frequently require (see *AI in Social Protection: Data Sovereignty Considerations* published by the AI Hub for Social Protection of the DCI, 2026b).

This finding matters for policy. It means that for the 60% of AI deployments that use classical ML, digital sovereignty is technically achievable: these systems do not require the large-scale compute infrastructure or cloud-based API access that characterises foundation model deployments. Governments in lower-income contexts can, in principle, run these models on domestic infrastructure, preserving data residency, reducing vendor dependency, and maintaining sovereign control over the systems that govern their citizens' entitlements.

More generally, data sovereignty and residency are under-documented. Most cases do not provide enough information to determine the compute environment, data residency, or cross-border transfer arrangements. Where information is available, HIC deployments tend to use domestic infrastructure while LIC and LMIC deployments funded by international organisations more frequently involve cross-border data processing. Vendor lock-in and hosting opacity compound the problem: the DCI AI Hub's *AI in Social Protection: Data Sovereignty Considerations*[7] report (2026b) found that two thirds of cases offer no public documentation of where data is stored or processed, and that at least 8% of systems show confirmed indicators of vendor lock-in.

4.5 Governance, oversight and outcomes

Most cases have humans in or on the loop (HITL or HOTL). Fifty-four (54) cases reported HITL oversight, 42 reported HOTL, and 3 human-out-of-the-loop (HOOTL).

Yet formal oversight classifications may be unreliable. A system described as HITL is only as effective as the human in the loop. But if the reviewer faces time pressure, lacks training to evaluate the AI's output, or routinely approves recommendations without independent assessment, the effective oversight is weaker than the formal description suggests. Our evidence base rarely reports override rates, reviewer training, or the conditions under which humans disagree with the AI.

Decision criticality

Decision criticality is rated high in 39 cases, moderate in 31, and low in 29. High-criticality cases are concentrated in identification (authentication failure prevents benefit access), compliance (fraud flags trigger investigation), and vulnerability assessment (targeting determines who receives assistance). Low-criticality cases are predominantly found in user communication.

Documented benefits

Where benefits are documented, they are concrete. Maidstone's homelessness prediction system (GBR-001) achieved a 40% reduction in the local homelessness rate and estimated savings of GBP 225,000. **Korea**'s CLOVA CareCall (KOR-003) is associated with a 44% reduction in solitary deaths in service areas. In Togo, AI-assisted targeting reached approximately 140,000 beneficiaries

7 The report examines these implications in detail, and includes a sovereignty risk framework that maps data sensitivity against infrastructure intensity and model training arrangements.

who could not have been identified through existing registries. **India**'s CPGRAMS (IND-004) reduced grievance disposal time from 32 days to 12 days.

However, most documented benefits come from implementer self-reporting, not independent evaluation. The evidence base for cost-effectiveness is still thin.

Documented harms

Documented harms are concentrated in three areas. First, biometric exclusion, where individuals cannot authenticate due to physical limitations, environmental conditions, or system errors – and this is documented across multiple identification cases. Second, discriminatory profiling. Algorithmic risk scoring disproportionately targets specific demographic groups, documented most starkly in the **Netherlands** (NLD-001, NLD-002) and **Sweden** (SWE-001), where women and people with foreign backgrounds were systematically over-flagged. Third, system discontinuation. Five cases were suspended or halted after governance failures, resulting in wasted investment and disrupted services.

FIGURE 6. DOCUMENTED BENEFITS, HARMS AND KEY EVIDENCE GAPS

Documented benefits

- 40% reduction in homelessness (UK)
- 44% reduction in solitary deaths (Korea)
- Faster grievance resolution (India)
- Efficiency gains in operations

Documented harms

- Biometric exclusion and false mismatches
- Discriminatory profiling and bias
- Benefit delays or errors
- System discontinuation and disruption

Key evidence gaps

- Few independent evaluations or impact studies
- Rare reporting of override rates or human decisions
- Limited long-term outcome data
- Insufficient transparency on harms

Gendered and equity patterns

Where the evidence permits analysis, this review found that AI-enabled social protection systems reproduce and, at times, amplify gendered and group-specific patterns in the populations they serve. In the **Netherlands** (NLD-001, NLD-002) and **Sweden** (SWE-001), algorithmic fraud detection over-flagged women by a factor of roughly 1.5 and people with foreign backgrounds by a factor of 2.5 to 3 – patterns that survived court and regulatory scrutiny and contributed to each system's eventual suspension. Positive equity outcomes are also documented: CLOVA CareCall (KOR-003) is associated with a 44% reduction in solitary deaths among elderly people living alone in service areas, and Hyodol (KOR-004) showed significant reductions in loneliness and depressive symptoms in a user base that skews female and elderly. Fourteen (14) cases in the evidence base engage directly with disability-related entitlements, predominantly eligibility determination or case triage, but none in the available documentation disaggregate outcomes by disability type. Gender, disability and ethnic disaggregation of AI-system outcomes is the single most common reporting gap in the evidence base, and it is where the next wave of deployments should demand evaluation design upfront.

Beyond these documented harms, the combination of biometric identity systems, predictive targeting, and compliance monitoring creates a comprehensive data infrastructure that, without adequate safeguards, could enable the disproportionate surveillance of vulnerable populations. This is a plausible systemic risk rather than a documented impact in the mapped evidence: no case in the evidence base reports surveillance as a realised harm. *The AI Hub Risk Assessment Framework for Social Protection* (DCI AI Hub, 2026c) examines this risk in detail. The distinction matters. The three harms above are evidenced. The surveillance concern is inferred from the architecture of the systems and should be treated as a governance priority, not a finding.

The evaluation gap

The independent evaluation of AI in social protection is rare, and this is the most important cross-cutting finding after governance. Most evidence comes from implementer reports and technical documentation. Peer-reviewed evaluation exists for a small subset of cases. Randomised controlled trials or quasi-experimental evaluations are almost entirely absent. Vendor claims of accuracy and efficiency are frequently unverified. We do not know the long-term effects of AI deployment on institutional capacity, staff skills, beneficiary experience, or system resilience. Until this evaluation gap is closed, the evidence base for AI in social protection rests on a foundation of self-reported outcomes.

CHAPTER 5. CONCLUSIONS AND EVIDENCE GAPS

AI in social protection is real, growing, and unevenly distributed. It is concentrated in user communication, vulnerability assessment, compliance monitoring, and identification. It is predominantly concentrated in HICs, which mainly use classical machine learning, most commonly in social assistance.

The four most common use case categories offer clear benefits to governments, but pose risks to the people governments serve (with the exception of user communication and interaction, where risks are lower, but still exist). Going forward it will be critical to adopt this lens when assessing AI readiness for future AI use cases in the social protection sector[8] – so as to systematically mitigate emerging risks, and halt applications that are cause for concern.

Ultimately, governance determines outcomes. Prediction models map poverty in **Togo** (TGO-001, TGO-005; Aiken et al., 2022), reaching approximately 140,000 of the poorest beneficiaries with emergency cash. The same family of techniques detected fraud in the **Netherlands** (NLD-001; District Court of The Hague, 2020), where the system was struck down by a court for violating privacy rights without detecting a single case of fraud. There are five such cases in which AI uses in social protection were suspended or halted, all for governance failures.

The difference in each case is not the technical sophistication of the model. It is the design choices, oversight arrangements, institutional capacity, and legal safeguards surrounding it.

The AI Hub *Risk Assessment Framework for Social Protection* (DCI AI Hub, 2026c), and *AI in Social Protection: Data Sovereignty Considerations* report (DCI AI Hub, 2026b) together provide the tools and evidence base for navigating these choices. This review supplies the empirical foundation; the companion products supply the frameworks for action.

As a desk-research baseline, this review documents what we found, not everything that exists. UMICs, labour market programmes, AI-assisted casework, longitudinal evaluation, and the lived experience of affected populations are all areas where the evidence base is thin. Whether this reflects genuine under-deployment or simply under-documentation is an open question, and distinguishing between the two should be a priority for the DCI AI Hub and the social protection sector's future research agenda.

8 The DCI AI Hub is designing guidance on this topic specifically tackled through the AI Hub Clinic Series, which is designed to allow for knowledge sharing and an open discussion of challenges with government counterparts and technical audiences.

Priorities for further research

Four research priorities emerge from this review.

First, evidence priorities: the independent evaluation of AI systems in social protection is almost entirely absent. Rigorous impact evaluations, including randomised and quasi-experimental designs, are needed for the most widely-deployed system types, particularly chatbots, vulnerability targeting models, and fraud detection algorithms. Comparative cost-effectiveness studies that measure AI-assisted approaches against simpler alternatives would be especially valuable.

Second, geographic priorities: targeted case documentation in UMICs, francophone Africa, and South and Southeast Asia would address the most visible gaps in the current evidence base. Field research that distinguishes genuine under-deployment from documentation scarcity is needed to calibrate the evidence map accurately.

Third, method priorities: longitudinal tracking of deployed systems over multiple years would reveal how AI performance, governance arrangements, and institutional capacity evolve after initial deployment. Beneficiary-perspective research, capturing how affected populations experience and understand AI-mediated decisions, is almost entirely missing and should be a priority for any programme that claims to be people-centred.

Fourth, governance priorities: the systematic measurement of false positive rates by demographic group in compliance and targeting systems, documentation of effective human oversight practices that go beyond formal classification labels, and analysis of data sovereignty arrangements in cross-border AI deployments would each strengthen the evidence base for the governance choices that this review identifies as decisive.

REFERENCES

Agyemang, F. S. K., Memon, R., Wolf, L. J. and Fox, S., 'High-Resolution Rural Poverty Mapping in Pakistan with Ensemble Deep Learning', *PLOS ONE*, Vol. 18, No. 4, 2023, e0283938, https://doi.org/10.1371/journal.pone.0283938

Aiken, E., Bellue, S., Karlan, D., Udry, C. and Blumenstock, J., 'Machine learning and phone data can improve targeting of humanitarian aid', *Nature*, 603, 864–870, 2022, doi:10.1038/s41586-022-04484-9

Al Jazeera, 'How an algorithm denied food to thousands of poor in India's Telangana', *Al Jazeera*, 24 January 2024, https://www.aljazeera.com/economy/2024/1/24/how-an-algorithm-denied-food-to-thousands-of-poor-in-indias-telangana (accessed 24 March 2026)

AlgorithmWatch and Bertelsmann Stiftung, *Automating society report 2020 – Denmark section*, Berlin, AlgorithmWatch, 2020, https://automatingsociety.algorithmwatch.org/report2020/denmark (accessed 31 October 2025)

Amnesty International, *Coded injustice: Surveillance and discrimination in Denmark's Automated Welfare State*, Copenhagen, Amnesty International Denmark, 2024, https://amnesty.dk/wp-content/uploads/2024/11/Coded-Injustice-Surveillance-and-discrimination-in-Denmarks-automated-welfare-state.pdf (accessed 31 October 2025)

Armed Conflict Location & Event Data (ACLED), 'Helping the Danish Refugee Council anticipate and prevent forced displacement', ACLED, 2024, https://acleddata.com/helping-danish-refugee-council-anticipate-and-prevent-forced-displacement (accessed 23 March 2026)

Boost.ai, 'Conversational AI supports Norway's COVID response', Boost.ai Case Studies, 2024, https://boost.ai/case-studies/how-conversational-ai-is-helping-norways-citizens-with-covid/ (accessed 26 March 2026)

California Policy Lab, *The Homelessness Prevention Unit: A proactive approach to preventing homelessness*, Los Angeles, CPL, 2024 https://capolicylab.org/wp-content/uploads/2024/12/Homelessness-Prevention-Unit-Report.pdf (accessed 31 October 2025)

Canada School of Public Service, *Government of Canada Data Conference 2023: ESDC Integrity Services Branch's use of quantitative and qualitative data* (transcript), 2023, https://www.csps-efpc.gc.ca/video/data-conference2023/integrity-services-eng.aspx (accessed 31 October 2025)

Central Digital and Data Office (CDDO), *Artificial intelligence playbook for the UK government*, London, CDDO, 2025, https://www.gov.uk/government/publications/ai-playbook-for-the-uk-government/artificial-intelligence-playbook-for-the-uk-government-html (accessed 24 March 2026)

Centre for Humanitarian Data, 'Peer review of 510's Typhoon Model and its use in the Philippines', The Centre for Humanitarian Data Blog, 5 October 2022, https://centre.humdata.org/peer-review-of-510s-typhoon-model-and-its-use-in-the-philippines/ (accessed 31 October 2025)

Chandran, R., 'Hyodol AI robots ease loneliness for South Korea's seniors', Rest of World, 14 January 2025, https://restofworld.org/2025/korea-ai-robot-senior-care-hyodol/ (accessed 25 March 2026)

County of Los Angeles DPSS and Office of the CIO, *Approve amendment number two to Agreement 77217 with SAS Institute Inc. for Data Mining Solution (DMS)*, CIO Analysis Report CIO 12-07, Los Angeles, County of Los Angeles, 2012, https://file.lacounty.gov/SDSInter/bos/supdocs/68310.pdf (accessed 15 May 2026)

Daros, G., 'Brazil's AI-powered social security app is wrongly rejecting claims', *Rest of World*, 24 April 2025, https://restofworld.org/2025/brazil-ai-social-security-app-rejected/ (accessed 30 March 2026)

Dataprev, 'Conheça Helô, a atendente virtual que a Dataprev desenvolveu para o INSS', dataprev.gov.br, 2020

DCI AI Hub for Social Protection, *A Taxonomy for AI in Social Protection: Defining Capabilities and Application Areas*, Deutsche Gesellschaft für Internationale Zusammenarbeit, Germany, 2026a.

DCI AI Hub for Social Protection, *AI in Social Protection: Data Sovereignty Considerations*, Deutsche Gesellschaft für Internationale Zusammenarbeit, Germany, 2026b.

DCI AI Hub for Social Protection, *AI Hub Risk Assessment Framework for Social Protection*. Deutsche Gesellschaft für Internationale Zusammenarbeit, Germany, 2026c.

Department for Work and Pensions (DWP), *Fairness assessment including statistical analysis of the Universal Credit advances machine learning model: 1 April 2024 to 31 March 2025*, London, DWP, 2025, https://www.gov.uk/government/publications/fairness-assessment-including-statistical-analysis-of-the-universal-credit-advances-machine-learning-model-1-april-2024-to-31-march-2025 (accessed 22 March 2026)

Deutsche Rentenversicherung Bund, *Künstliche Intelligenz entlastet Mitarbeitende und schützt das Sozialversicherungssystem*, KIRA Digitalstrategie, 2024, https://www.deutsche-rentenversicherung.de/Bund/DE/Ueber-uns/Digitalstrategie/KIRA.html (accessed 19 March 2026).

Deutscher Bundestag, 'Einsatz Künstlicher Intelligenz in Jobcentern', Kurzmeldungen, bundestag.de, 2024, https://www.bundestag.de/presse/hib/kurzmeldungen-1019588 (accessed 26 March 2026).

Diario y Radio Universidad Chile, *Alerta Infancia: el software que expone los datos personales de ninos y ninas en riesgo social*, Diario y Radio Universidad Chile, 29 January 2019, https://radio.uchile.cl/2019/01/29/alerta-infancia-el-software-que-expone-los-datos-personales-de-ninos-y-ninas-en-riesgo-social/ (accessed 30 March 2026)

District Court of The Hague, *Judgment in case C-09-550982-HA ZA 18-388 (NJCM v. the State of the Netherlands)*, 5 February 2020

Election Commission of Bangladesh, *Bangladesh NID application system*, Dhaka, NIDW, 2025, https://services.nidw.gov.bd/nid-pub/ (accessed 31 October 2025)

Eubanks, V., *Automating inequality: How high-tech tools profile, police, and punish the poor*, St Martin's Press, 2019)

European Commission, *Government Virtual Assistant Bürokratt (OSOR case)*, Brussels, Interoperable Europe Portal, 2025, https://interoperable-europe.ec.europa.eu/collection/open-source-observatory-osor/government-virtual-assistant-burokratt (accessed 31 October 2025)

Federal Republic of Nigeria, *Nigeria Data Protection Act*, 2023, Abuja, Government Printer, 2023, https://cert.gov.ng/ngcert/resources/Nigeria_Data_Protection_Act_2023.pdf (accessed 31 October 2025)

Financial Protection Forum, 'According to plan: Second activation of Uganda's Displacement Crisis Response Mechanism (DCRM)', Washington, DC, World Bank SPJ Platform, 2023, https://www.financialprotectionforum.org/blog/according-to-plan-second-activation-of-ugandas-displacement-crisis-response-mechanism-dcrm (accessed 24 March 2026)

Fondation IFRAP, 'Assurance-maladie : 628 millions d'euros de fraudes détectées… sur 4 milliards de fraudes estimées', Fondation IFRAP, 2025, https://www.ifrap.org/emploi-et-politiques-sociales/assurance-maladie-628-millions-deuros-de-fraudes-detecteessur-4-milliards-de-fraudes-estimees (accessed 26 March 2026)

Frimpong, R. & Richiardi, M., *Machine learning regionalisation of input data for microsimulation models: An application of a hybrid GBM/IPF method to build a tax-benefit model for the Essex region in the UK*, CeMPA Working Paper 9/25, University of Essex, 2025

Ghana Data Protection Commission, *Data Protection in Ghana – Role and Mandate of the Commission, Accra*, DPC, 2025, https://dataprotection.org.gh/ (accessed 31 October 2025)

Gholami, S., Knippenberg, E., Campbell, J., Andriantsimba, D., Kamle, A., Parthasarathy, P., Sankar, R., Birge, C. and Lavista Ferres, J.M., 'Food security analysis and forecasting: A machine learning case study in southern Malawi', *Data & Policy*, 4, e38, 2022, doi:10.1017/dap.2022.24

Gobierno de México and ISSSTE, 'Abre Issste servicios digitales Asisssste con portal, línea y chat', gob.mx, 11 April 2022, https://www.gob.mx/issste/prensa/abre-issste-servicios-digitales-asisssste-con-portal-linea-y-chat (accessed 26 March 2026)

Government of Ethiopia, *Ethiopian Digital Identification Proclamation No. 1284/2023*, Addis Ababa, Ministry of Justice, 2023, https://justice.gov.et/en/law/ethiopian-digital-identification-proclamation/ (accessed 29 October 2025)

Government of Korea, *Digital Government Act and National Data-Innovation Strategy*, Seoul, Ministry of the Interior and Safety, 2024, https://www.mois.go.kr/eng/ (accessed 31 October 2025)

GovTech Singapore, 'MyCareersFuture. Singapore: Government Technology Agency', 2025, https://www.tech.gov.sg/products-and-services/for-citizens/employment/mycareersfuture/ (accessed 27 March 2026)

GovTech Singapore, 'Career Kaki', 2025, https://www.tech.gov.sg/products-and-services/for-citizens/employment/career-kaki/ (accessed 27 March 2026)

Gualavisi, M. and Newhouse, D., 'Integrating survey and geospatial data for geographical targeting of the poor and vulnerable: Evidence from Malawi', *The World Bank Economic Review*, 39(2), 377–409, 2025, doi:10.1093/wber/lhf008

Ikumi, S., 'Swedish welfare authorities suspend "discriminatory" AI model', *Computer Weekly*, 19 June 2025, https://www.computerweekly.com/news/366634703/Swedish-welfare-authorities-suspend-discriminatory-AI-model (accessed 25 March 2026)

Innovationsfördernde Öffentliche Beschaffung (IÖB), *Artificial intelligence (AI) in der Kostenerstattung der SV*, Vienna, IÖB, n.d., https://www.ioeb.at/erfolgreiche-projekte-detail/artificial-intelligence-ai-in-der-kostenerstattung-der-sv (accessed 24 March 2026)

International Labour Organization (ILO), *Tracer study: Skills profiling application in Egypt for refugees and host communities*, Geneva, ILO PROSPECTS, 2022, https://www.ilo.org/publications/tracer-study-skills-profiling-application-egypt-refugees-and-host (accessed 31 October 2025)

ISSA, *DEFRADA (Deteksi Potensi Fraud Dengan Analista Data Klaim): The development of a fraud detection tool in hospital services*, ISSA Good Practice, Geneva, ISSA, 2018, https://ww1.issa.int/gp/173411 (accessed 26 March 2026)

ISSA, *Innovating the National Pension Service with artificial intelligence: Building global accessibility for the future, Good Practice in Social Security,* International Social Security Association, 2024, https://www.issa.int/gp/257677 (accessed 26 March 2026)

ISSA, *AI applications in social security: Building evidence and insights from the TechByte*, International Social Security Association, Geneva, 2025

ISSA, 'Julieta Lanteri, primer chatbot de la administración publica nacional argentina', ISSA Good Practice gp/198017, n.d., https://www.issa.int/gp/198017 (accessed 26 March 2026)

Kaye, K. 'AI governance on the ground: Chile's Social Security and Medical Insurance agency grapples with balancing new responsible ai criteria and vendor cost', World Privacy Forum, 1 November 2024, https://worldprivacyforum.org/posts/ai-governance-on-the-ground-chiles-social-security-and-medical-insurance-agency-grapples-with-balancing-new-responsible-ai-criteria-and-vendor-cost/ (accessed 30 March 2026)

Kela, 'AI register and principles for the responsible use of AI', About Kela, Helsinki, Kela, 2025, https://www.kela.fi/ai-register (accessed 24 March 2026)

Kenya Law, *Social Health Insurance Act, No. 16 of 2023*, Nairobi, National Council for Law Reporting, 2023, https://new.kenyalaw.org/akn/ke/act/2023/16 (accessed 31 October 2025)

Kim, H. '"The Work", AI job recommendation service using the National Job Information Platform', Observatory of Public Sector Innovation (OPSI), OECD, 5 August 2020, https://oecd-opsi.org/innovations/the-work/ (accessed 22 March 2026)

Lighthouse Reports, *France's Digital Inquisition*, Lighthouse Reports, December 2023a, https://www.lighthousereports.com/investigation/frances-digital-inquisition/ (accessed 26 March 2026)

Lighthouse Reports, 'Spain's AI Doctor', Suspicion Machines, April 2023b, https://www.lighthousereports.com/investigation/spains-ai-doctor/ (accessed 11 May 2026)

Lighthouse Reports, 'Suspicion Machines', Lighthouse Reports, March 2023c, https://www.lighthousereports.com/investigation/suspicion-machines/ (accessed 26 March 2026)

Lindert, K., Linder, A., Hobbs, J. and de la Brière, B., *Sourcebook on the foundations of social protection delivery systems*, Washington, DC, World Bank, 2020, doi:10.1596/978-1-4648-1577-5

Local Government Association, 'Peterborough City Council: Hey Geraldine, a personalised AI assistant', LGA Case Studies, December 2024, https://www.local.gov.uk/case-studies/peterborough-city-council-hey-geraldine-personalised-ai-assistant (accessed 27 March 2026)

Lopez, J. and Castaneda, J.D., 'Automation, digital technologies and social justice: The case of SISBEN in Colombia', in: Cerrillo-i-Martinez, A. and Fabra-Abat, P. (eds) AI, *Human Rights and Social Justice*, Karisma Foundation/CETyS Universidad de San Andres, 2020, https://web.karisma.org.co/ (accessed 30 March 2026)

MacDonald, A., 'UNHCR biometric program facilitates cash aid delivery to 39K Afghan returnees', Biometric Update, 22 February 2024, https://www.biometricupdate.com/202402/unhcr-biometric-program-facilitates-cash-aid-delivery-to-39k-afghan-returnees (accessed 24 March 2026)

Maidstone Borough Council, *Data protection impact assessment: 'EY & Xantura: Homelessness predictive analytics*, Maidstone, MBC, 2019, https://maidstone.gov.uk/__data/assets/pdf_file/0006/416553/FOI-4452-Appendix-1.pdf (accessed 31 October 2025)

Microsoft, 'Buenos Aires City: How generative AI is revolutionizing the lives of millions with Azure OpenAI Services', Microsoft Customer Stories, 2024, https://www.microsoft.com/en/customers/story/21596-government-of-the-city-of-buenos-aires-azure-open-ai-service (accessed, 25 March 2026)

Ministry of Health, Labour and Welfare, Press release: '将来を見据えたハローワークにおけるAI活用について' を公表 [Looking Ahead: AI Utilization in Hello Work], Tokyo, MHLW, 22 April 2025, https://www.mhlw.go.jp/stf/houdou/newpage_57223.html (accessed 24 March 2026)

Ministry of Labour and Social Protection of Population, 'AI-based "iDost" Digital Assistant launched on the e-Social Platform', sosial.gov.az, 6 January 2026, https://sosial.gov.az/en/media/news/ai-based-idost-digital-assistant-launched-on-the-e-social-platform-8871 (accessed 15 May 2026)

National Academy of Social Insurance, *Phase one report: Task Force on Artificial Intelligence, Emerging Technology, and Disability Benefits*, Washington, DC, NASI, 2025 https://www.nasi.org/wp-content/uploads/2025/04/Phase-One-Report-Task-Force-on-Artificial-Intelligence-Emerging-Technology-and-Disability-Benefits.pdf (accessed 31 October 2025)

Nepal Law Commission, *National Identity and Civil Registration Act, 2076* (2020), Kathmandu, Nepal Law Commission, 2020, https://lawcommission.gov.np/content/12265/12265-national-identity-card-and-vit/ (accessed 30 October 2025)

NIST, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, Gaithersburg, MD, National Institute of Standards and Technology, 2023.

O'Brien, C., 'Facial recognition helps detect welfare benefits fraud', *The Irish Times*, 5 September 2015, https://www.irishtimes.com/news/ireland/irish-news/facial-recognition-helps-detect-welfare-benefits-fraud-1.2340900 (accessed 27 March 2026)

OECD, *AI and the future of social protection in OECD countries*, OECD Artificial Intelligence Papers, No. 42, Paris, OECD Publishing, 2025

OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, Paris, OECD, 2019

Office of Management and Budget, 'M-24-10: Advancing governance, innovation, and risk management for agency use of AI', Washington, DC, The White House, 2024, https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf (accessed 31 October 2025)

Perrigo, B., 'Albania's AI-powered minister tests the future of government', *TIME*, September 2025, https://time.com/7324934/albania-ai-minister-diella/ (accessed 25 March 2026)

Pickool, The 'Naver AI CareCall Reduces Solitary Deaths by 44.2%', The Pickool [online], 2025, https://www.thepickool.com/naver-ai-care-call-reduces-solitary-deaths-by-44-2/ (accessed 25 March 2026)

Press Information Bureau, 'Anti-fraud system for India's National Health Insurance Scheme (AB-PMJAY)', Press Release, New Delhi, PIB, Ministry of Health and Family Welfare, 2 August 2022, https://www.pib.gov.in/ (accessed 31 October 2025)

Press Information Bureau, 'Dr Jitendra Singh launches the Intelligent Grievance Monitoring System (IGMS) 2.0 Public Grievance portal and Automated Analysis in Tree Dashboard portal of DARPG', Press Release, New Delhi, Government of India, 29 September 2023a, https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=1962142 (accessed 23 March 2026)

Press Information Bureau, 'Ministry of skill development & entrepreneurship launches AI Assistant "SIA" on WhatsApp and Skill India Digital Hub', Press Release, New Delhi, Government of India, 2025 https://pib.gov.in/PressReleasePage.aspx?PRID=2146573 (accessed 27 March 2026)

Press Information Bureau, 'Union Minister of State for Agriculture and Farmers Welfare, Shri Kailash Choudhary launches AI chatbot for PM-KISAN scheme today', Press Release, New Delhi, Government of India, 2023b, https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1959461 (accessed 24 March 2026)

Press Information Bureau, 'Use of facial authentication technology by the EPS pensioners for Digital Life Certificate Submission', Press Release, New Delhi, Government of India, 2024, https://pib.gov.in/PressReleasePage.aspx?PRID=2023558 (accessed 24 March 2026)

Privacy International, *The SyRI case: A landmark ruling for benefits claimants around the world*, 2020 https://privacyinternational.org/news-analysis/3363/syri-case-landmark-ruling-benefits-claimants-around-world (accessed 14 May 2026)

Radio Pakistan, 'NADRA launches Iris Recognition System for identity verification, citizen de-duplication', Radio Pakistan, 4 June 2023, https://www.radio.gov.pk/04-06-2023/nadra-launches-iris-recognition-system-for-identity-verification-citizen-de-duplication (accessed 31 October 2025)

Ruiz, E.R., Arellano, C.A., Archila, C.A., Llobet, C., Carrasco, G. and Pinochet, F., 'Clinical validation of an AI system for pneumoconiosis detection using chest X-rays', *Journal of Occupational and Environmental Medicine*, 67(4), e250–e256, 2025, doi: 10.1097/JOM.0000000000003329

SAP, 'City of Hamburg: Leveraging BTP to build a Cultural Covid-19 Aid-Platform', SAP Customer Story, Walldorf: SAP SE, 2025, https://www.sap.com/asset/dynamic/2025/01/043d156f-ed7e-0010-bca6-c68f7e60039b.html (accessed 31 October 2025)

SkillLab, 'Enhancing job-seekers work-readiness in Lebanon using the language of skills', SkillLab News, January 2025, https://skilllab.io/en-us/news/enhancing-job-seekers-work-readiness-in-lebanon (accessed 24 March 2026)

SkillsFuture Singapore (SSG), *Enhanced MySkillsFuture portal to provide personalised recommendations to guide Singaporeans towards achieving career and skills goals*, Singapore, SSG, 2019, https://www.ssg.gov.sg/newsroom/enhanced-myskillsfuture-portal-to-provide-personalised-recommendations-to-guide-singaporeans-towards-achieving-career-and-skills-goals/ (accessed 22 March 2026)

Trendall, S., 'EXCL: DWP and £11m supplier prep five gen AI projects "with director general sponsorship"', PublicTechnology, 4 December 2024, https://www.publictechnology.net/2024/12/04/society-and-welfare/excl-dwp-and-11m-supplier-prep-five-gen-ai-projects-with-director-general-sponsorship/ (accessed 22 March 2026)

UNDP Innovation Toolkit, 'Text mining performance evaluation', in: *UNDP Accelerator Labs Innovation Toolkit for Signature Solutions*, Governance Chapter 6.8, 2023, https://undp-accelerator-labs.github.io/Innovation-Toolkit-for-UNDP-Signature-Solutions/6.Governance/6.8%20TextMiningPerformance.html (accessed 23 March 2026)

UNHCR, *Comprehensive registration and PRIMES in Ethiopia*, December 2018 Newsletter, ReliefWeb, 2018, https://reliefweb.int/report/ethiopia/comprehensive-registration-and-primes-ethiopia-december-2018-newsletter (accessed 23 March 2026)

US Department of Labor, Employment and Training Administration, 'Introducing Artificial Intelligence Adjudicator Assistance (AIAA): A Research Initiative Exploring Ways to Streamline Work for Adjudicators', Washington, DC, US Department of Labor, 2024, https://www.dol.gov/agencies/eta/ui-modernization/aiaa (accessed, 24 March 2026)

US Social Security Administration, *Quick Disability determinations* (QDD), 2019, https://www.ssa.gov/disabilityresearch/qdd.htm (accessed 23 March 2026)

US Social Security Administration, *Social Security announces AI enhancements for hearings recordings*, Press Release, 13 March 2025, https://www.ssa.gov/news/en/press/releases/2025-03-13.html (accessed 25 March 2026)

Vaithianathan, R., Putnam-Hornstein, E., Jiang, N., Nand, P. and Maloney, T., *Developing predictive risk models to support child maltreatment hotline screening decisions: Allegheny County methodology and implementation*, Auckland, Centre for Social Data Analytics, Auckland University of Technology, 2017, https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Methodology-V1-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL.pdf (accessed 30 October 2025)

Vihalemm, T., Manniste, M., Trumm, A. and Solvak, M., 'Specialists and algorithms: Implementation of AI in the delivery of unemployment services in Estonia', in: *Participatory AI in Public Social Services*, Cham, Springer, 2025, https://link.springer.com/chapter/10.1007/978-3-031-71678-2_5

World Food Programme, *Biometric Deduplication (ABIS) – SCOPE user manual*, Rome, WFP, n.d., https://usermanual.scope.wfp.org/cash-accounts/content/intros_to_sections/deduplication.htm (accessed 30 October 2025)

World Food Programme, *Machine Learning for Early Warning Systems* (Factsheet), Rome, WFP, 2023, https://www.wfp.org/publications/2023-machine-learning-early-warning-systems (accessed 30 October 2025)

World Food Programme, 'Mozambique – Emergency Response', WFP Drones, Rome, WFP, 2019, https://drones.wfp.org/mozambique-emergency-response (accessed 31 October 2025)

ANNEX 1. CASE INDEX

This annex lists all 99 cases reviewed in this document, with country, programme name and primary citation. Case IDs follow the DCI AI Hub convention (country ISO3 code + sequence number, e.g. UGA-001). Full case descriptions, sources and classifications are available on the AI Hub Tracker (forthcoming). In terms of the strength of the evidence, 25 cases are rated A, 73 are rated B and 1 is rated C, reflecting an evidence base that is dominated by implementer and government reporting rather than independent evaluation (A = peer-reviewed study or independent audit; B = implementer evaluation, government or multilateral report, working paper, or legal/regulatory document; C = media or single-source account).

TABLE A1. CASE INDEX: 99 CASES REVIEWED WITH COUNTRY, PROGRAMME AND PRIMARY SOURCE

CASE ID	COUNTRY	PROGRAMME/SYSTEM	PRIMARY SOURCE	EVIDENCE
AFG-001	Afghanistan	UNHCR Afghanistan – Forcibly Returned Afghans Enrollment (FARE) Cash Assistance Programme	MacDonald (2024)	A
ALB-001	Albania	eAlbania Digital Government Services Platform	Perrigo (2025)	C
ARG-001	Argentina	Buenos Aires City Citizen Services Platform (Boti WhatsApp Chatbot)	Microsoft (2024)	B
ARG-002	Argentina	Sistema de Riesgos del Trabajo (Occupational Risk System)	ISSA (n.d.)	B
ARM-001	Armenia	Artificial Intelligence in Social Security / AI for Predictive Social Protection	Caucasus Watch (2022)	B
AUS-001	Australia	Centrelink Income Support System (research prototype for ML-based recertification targeting)	Sansone & Zhu (2021)	A
AUT-001	Austria	KAI System – AI-based Semi-Automatic Reimbursement of Medical Service	Innovationsfördernde Öffentliche Beschaffung (IÖB) (n.d.)	B

CASE ID	COUNTRY	PROGRAMME/SYSTEM	PRIMARY SOURCE	EVIDENCE
AZE-001	Azerbaijan	142 Call Center and Social Bot/ iDost Digital Assistant	Ministry of Labour & Social Protection of Population (2026)	B
BEL-001	Belgium	ONEM/RVA Unemployment Insurance Information Services	ONEM (2025)	B
BGD-002	Bangladesh	Smart National Identity (Smart NID) System	Election Commis-sion of Bangladesh (2025)	B
BGD-003	Bangladesh	National Helpline 333	a2i (2025)	B
BGD-004	Bangladesh	GiveDirectly MobileAid Phone-Data (CDR (Call Detail Records)) Targeting Pilot, Cox's Bazar	Aiken (2025)	B
BRA-001	Brazil	Meu INSS (My INSS) Digital Platform, Helô Virtual Assistant	Dataprev (2020)	B
BRA-002	Brazil	INSS Social Security Benefits (Retirement, Sick Pay, Disability) via Meu INSS Application	Daros (2025)	A
CAN-001	Canada	Employment Insurance (EI) Machine Learning Workload	Canada School of Public Service (2023)	B
CHL-001	Chile	Chilean Occupational Health Surveillanc, Pneumoconiosis Screening	Ruiz et al. (2025)	A
CHL-002	Chile	Oficinas Locales de la Ninez (OLN) – Child Protection Services under the Subsecretaria de la Ninez	Diario y Radio Universidad Chile (2019)	B
CHL-003	Chile	SUSESO Medical Insurance Claims Processing System	Kaye (2024)	B
COD-001	Congo, Dem. Rep.	WFP SCOPE – Biometric Verification at Food/Cash Distributions (DRC)	World Food Programme (n.d.)	B
COL-001	Colombia	SISBEN IV (Sistema de Identifica-cion de Potenciales Beneficiarios de Programas Sociales) – Cross-cutting targeting system for 18+ social programmes including Familias en Accion	Lopez & Castaneda (2020)	A
DEU-001	Germany	KIRA (AI for Risk-Based Employer Audits)	Deutsche Renten-versicherung Bund (2024)	B

CASE ID	COUNTRY	PROGRAMME/SYSTEM	PRIMARY SOURCE	EVIDENCE
DEU-002	Germany	Cultural Funds Platforms (Cultural Energy Fund & Cultural Events Fund)	SAP (2025)	B
DEU-003	Germany	BG BAU Occupational Safety and Health Prevention Programme	Hellmann (2024)	B
DEU-004	Germany	VerBIS Job Placement and Counselling System	Deutscher Bundes-tag (2024)	B
DNK-001	Denmark	Udbetaling Danmark (UDK) – Algorithmic Fraud Detection System	Amnesty Interna-tional (2024)	B
DNK-002	Denmark	Gladsaxe Model (Municipal Predictive Profiling Pilot for Vulnerable Children)	AlgorithmWatch/ Bertelsmann Stiftung (2020)	B
EGY-001	Egypt, Arab Rep.	ILO PROSPECTS Egypt – Skills Profiling Application (SkillLab)	International Labour Organization (2022)	B
ESP-001	Spain	Incapacidad Temporal (Temporary Disability/Sick Leave Benefits)	Lighthouse Reports (2023b)	A
ESP-002	Spain	Paloma/Paloma 2.0 – AI Virtual Assistant for Elderly Loneliness Detection	Ayuntamiento de Madrid (2024)	B
EST-001	Estonia	KrattAI/Bürokratt (public-sector virtual assistant network)	European Commission (2025)	B
EST-003	Estonia	OTT (Otsustustugi) Decision-Support Tool	Vihalemm et al. (2025)	A
ETH-001	Ethiopia	Fayda Digital ID for Inclusion and Services Project (P179040)	Government of Ethiopia (2023)	B
ETH-002	Ethiopia	SEWAA, Strengthening Early Warning Systems for Anticipa-tory Action	World Food Programme (2023)	B
ETH-003	Ethiopia	Refugee assistance biometric authentication at distributions (UNHCR PRIMES: BIMS + Global Distribution Tool)	UNHCR (2018)	B
FIN-001	Finland	Kela Benefit Administration AI Platform	Kela (2025)	A
FRA-001	France	Allocations Familiales and Housing Benefits (administered by Caisses d'Allocations Familiales / CAF)	Lighthouse Reports (2023a)	B

CASE ID	COUNTRY	PROGRAMME/SYSTEM	PRIMARY SOURCE	EVIDENCE
FRA-003	France	Assurance Maladie (French National Health Insurance)	Fondation IFRAP (2025)	B
GBR-001	United Kingdom	Homelessness Preventive Analytics ('One View')	Maidstone Borough Council (2019)	B
GBR-002	United Kingdom	GOV.UK Chat	Central Digital & Data Office (2025)	B
GBR-003	United Kingdom	DWP Generative-AI Pilots	Trendall (2024)	B
GBR-004	United Kingdom	UKMOD (UK Tax-Benefit Microsimulation Model)	Frimpong & Richiardi (2025)	A
GBR-005	United Kingdom	Risk-based Review and Fraud/ Error Targeting Models for Universal Credit (UC Advances ML Model and related models)	Department for Work & Pensions (2025)	B
GBR-006	United Kingdom	Peterborough City Council Adult Social Care	Local Government Association (2024)	B
GHA-001	Ghana	National Health Insurance Scheme (NHIS) – e-Claims Fraud and Audit Analytics	Ghana Data Protection Commission (2025)	A
IDN-001	Indonesia	Jaminan Kesehatan Nasional (JKN) – DEFRADA Fraud Detection System	International Social Security Association (2018)	B
IND-001	India	Pradhan Mantri Kisan Samman Nidhi (PM-KISAN)	Press Information Bureau (2023b)	B
IND-003	India	Jeevan Pramaan – Digital Life Certificate via Aadhaar Face Authentication	Press Information Bureau (2024)	B
IND-004	India	Centralised Public Grievance Redress and Monitoring System (CPGRAMS)	Press Information Bureau (2023a)	B
IND-005	India	Samagra Vedika – Integrated Data Platform (Telangana)	Al Jazeera (2024)	B
IND-006	India	Ayushman Bharat – Pradhan Mantri Jan Arogya Yojana (AB PM-JAY)	Press Information Bureau (2022)	B
IND-008	India	Aadhaar (Unique Identification Programme)	Press Information Bureau (2025a)	B

CASE ID	COUNTRY	PROGRAMME/SYSTEM	PRIMARY SOURCE	EVIDENCE
IND-009	India	Skill India Digital Hub (SIDH)/ Skill India Mission	Press Information Bureau (2025b)	B
IRL-001	Ireland	Cogent Facial Image Matching Software for Welfare Fraud Detection	O'Brien (2015)	B
JPN-001	Japan	Hello Work AI-Enhanced Job Matching and Career Guidance System	Ministry of Health, Labour & Welfare (2025)	B
KEN-001	Kenya	Social Health Authority (SHA) – Universal Health Coverage through Biometric Health Identification	Kenya Law (2023)	B
KOR-001	Korea, Rep.	National Health Insurance Service (NHIS) – Machine Learning-Based Fraud Detection and Prediction System	Government of Korea (2024)	B
KOR-002	Korea, Rep.	'The Work' – AI Job Recommendation Service using the National Job Information Platform (OECD OPSI)	Kim (2020)	A
KOR-003	Korea, Rep.	Municipal Elderly Welfare Check-In Programmes (Various Local Governments)	The Pickool (2025)	B
KOR-004	Korea, Rep.	Hyodol AI Care Robot Distribution Programme for Elderly Living Alone	Chandran (2025)	A
KOR-005	Korea, Rep.	Industrial Accident Compensation Insurance, Customized Integrated Services (CIS) Rehabilitation Programme	ISSA (2021)	B
KOR-006	Korea, Rep.	National Pension Service, AI-Equipped KIOSK and Real-Time AI Interpreter Service	ISSA (2024)	B
LBN-001	Lebanon	ILO PROSPECTS Lebanon – SkillLab Skills Profiling Pilot	SkillLab (2025)	B
MEX-001	Mexico	Mexico Performance Evaluation System (Sistema de Evaluacion del Desempeno, SED), NLP Text-Mining Component	UNDP Innovation Toolkit (2023)	B
MEX-002	Mexico	ISSSTE Social Insurance Services (ASISSSTE Omnichannel Digital Platform)	Gobierno de México & ISSSTE (2022)	B

CASE ID	COUNTRY	PROGRAMME/SYSTEM	PRIMARY SOURCE	EVIDENCE
MLI-003	Mali	Ignitia Climate Advisory (SMS weather advice for small-holders)	WFP Innovation (2025)	B
MOZ-002	Mozambique	WFP Emergency Drone and AI-Assisted Damage Assessment Programme (Cyclones Idai/Kenneth Response)	World Food Programme (2019)	A
MWI-001	Malawi	Geospatial poverty mapping for village-level area targeting (research using household survey, satellite imagery, and partial registry data)	Gualavisi & Newhouse (2025)	A
MWI-002	Malawi	United in Building and Advancing Life Expectations (UBALE) – MIRA sentinel site monitoring component	Gholami et al. (2022)	A
NGA-001	Nigeria	National Identity Number (NIN) – Foundational Identity System	Federal Republic of Nigeria (2023)	B
NLD-001	Netherlands	SyRI – System Risk Indication (Systeem Risico Indicatie)	District Court of The Hague (2020)	A
NLD-002	Netherlands	Rotterdam Municipal Social Assistance (Bijstandsuitkering)	Lighthouse Reports (2023)	B
NOR-001	Norway	NAV Multi-Channel Service Delivery, Frida Virtual Agent	Boost.ai (2024)	A
NPL-002	Nepal	National Identity Card (Rastriya Parichaya Patra) / National Identity Management Information System (NIDMIS)	Nepal Law Commission (2020)	B
PAK-001	Pakistan	Sindh Social Protection Strategic Unit – High-Resolution Rural Poverty Mapping Pilot	Agyemang et al. (2023)	A
PAK-002	Pakistan	NADRA Automated Biometric Identification System (ABIS) for CNIC Enrolment Deduplication	Radio Pakistan (2023)	B
PHL-001	Philippines	Philippines Anticipatory Action Framework for Tropical Cyclones (CERF/HCT)	Centre for Humanitarian Data (2022)	A
SAU-002	Saudi Arabia	GOSI Social Insurance Programmes – Amin AI Digital Human Customer Service	Saudi Press Agency (SPA) (2025)	B

CASE ID	COUNTRY	PROGRAMME/SYSTEM	PRIMARY SOURCE	EVIDENCE
SGP-001	Singapore	MyCareersFuture	GovTech Singapore (2025)	B
SGP-002	Singapore	Career Kaki	GovTech Singapore (2024)	B
SGP-003	Singapore	Ask MSF Chatbot	Ministry of Social & Family Develop-ment (2025)	B
SGP-004	Singapore	mindline.sg + Wysa Digital Mental-Health Chatbot	MOH Office for Healthcare Trans-formation (2020)	A
SGP-007	Singapore	SkillsFuture – MySkillsFuture Portal and Careers & Skills Passport (AI-Enabled Skills Profiling and Learning Recommendations)	SkillsFuture Singa-pore (SSG) (2019)	A
SSD-001	South Sudan	AHEAD – Anticipatory Humanitarian Action for Displacement (Foresight Model)	ACLED (2024)	A
SWE-001	Sweden	Tillfällig föräldrapenning (Temporary Parental Benefit/ VAB)	Ikumi (2025)	B
TGO-001	Togo	Novissi Emergency Cash Transfer Programme – Model 2 (Geospatial Area Targeting Component)	Aiken et al. (2022)	A
TGO-002	Togo	CNSS 'Biosecu' – online facial-recognition proof-of-life for pensioners	République Togo-laise (2022)	B
TGO-004	Togo	Novissi Emergency Cash Transfer Programme – ML Performance Monitoring Component	Aiken et al. (2022)	A
TGO-005	Togo	Novissi Emergency Cash Trans-fer Programme – Phase 2 (Rural Expansion with ML Targeting)	Aiken et al. (2022)	A
UGA-001	Uganda	Displacement Crisis Response Mechanism (DCRM) under DRDIP	Financial Protection Forum (2023)	B
UGA-002	Uganda	National Integrated Early Warning System (NIEWS) with SEWAA-linked trigger design for anticipatory/adaptive social protection response	IGAD Climate Prediction & Appli-cations Centre (2025)	B
UGA-003	Uganda	National Social Security Fund (NSSF) Uganda	Intelligent CIO Africa (2022)	B

CASE ID	COUNTRY	PROGRAMME/SYSTEM	PRIMARY SOURCE	EVIDENCE
USA-001	United States	Los Angeles County Homelessness Prevention Unit (HPU)	California Policy Lab (2024)	B
USA-002	United States	Allegheny Family Screening Tool (AFST)	Vaithianathan et al. (2017)	B
USA-003	United States	SSA disability adjudication quality-review process (Insight subsystem)	National Academy of Social Insurance (2025)	B
USA-004	United States	Unemployment Insurance (UI) Adjudication – AIAA Prototype	US Department of Labor & Training Administration (2024)	B
USA-005	United States	USCIS Refugee, Asylum, and International Operations (RAIO) Officer Training Programme	Office of Management & Budget (2024)	B
USA-006	United States	CalWORKs Stage 1 Child Care Program	County of Los Angeles DPSS / Office of the CIO (2012)	B
USA-007	United States	Social Security Administration (SSA) – Quick Disability Determinations (QDD) Predictive Model	US Social Security Administration (2019)	B
USA-010	United States	Social Security Disability Insurance (SSDI) hearings process	US Social Security Administration (2025)	B

Note. The risk categorisation refers to governance sensitivity. The levels reflected in this visual show the default tier; ratings for specific use cases within these categories escalate where outputs feed individual eligibility, sanctions or automated adverse action.

ANNEX 2. GLOSSARY

Anomaly detection
AI capability that identifies data points, patterns, or observations that deviate significantly from expected behaviour. Used in fraud detection, data quality monitoring, and compliance systems.

Classical ML
Machine learning approaches that use hand-engineered features and traditional algorithms such as gradient boosting, random forests, logistic regression, and support vector machines. Distinguished from deep learning by the absence of neural network architectures.

Classification
AI capability that assigns inputs to predefined categories based on learnt patterns. Used in eligibility assessment, document routing, and risk categorisation.

Clustering
AI capability that groups similar data points without predefined categories. Used in beneficiary segmentation and population profiling.

Decision criticality
Assessment of how directly an AI system's output affects beneficiary outcomes. Rated as high (directly determines access to benefits), moderate (influences but does not determine decisions), or low (informational only).

Deep learning
Machine learning using neural networks with multiple layers. Used primarily in perception tasks such as image recognition, biometric verification, and natural language processing.

Delivery chain
The sequence of operational functions in social protection programme delivery: outreach, intake and registration, needs assessment, eligibility determination, enrolment, benefit provision, beneficiary management, and exit. These functions group into four higher-level stages (as per the *Sourcebook on the Foundations of Social Protection Delivery Systems,* Lindert et al., 2020) – assess, enrol, provide, and manage – which are used as a filter dimension in the AI Hub Tracker.

Foundation model
Large-scale AI model pre-trained on broad data that can be adapted for specific tasks. Includes large language models (LLMs) and multimodal models.

HIC/UMIC/LMIC/LIC
World Bank income classifications: high-income, upper-middle-income, low and middle-income, and low-income countries.

HITL (human-in-the-loop)
Oversight model where a human reviews and approves every AI output before it affects beneficiaries.

HOTL (human-on-the-loop)
Oversight model where the AI operates with periodic human review and the ability for humans to intervene.

HOOTL (human-out-of-the-loop)
Oversight model where the AI operates autonomously without routine human review of individual outputs.

LLM
Large language model: a foundation model trained on text data, capable of generating, summarising, and transforming text. Used in chatbots, document processing, and content generation.

Perception and extraction
AI capability that processes unstructured inputs (images, text, audio) to extract structured information. The most common primary capability in the evidence base.

Prediction
AI capability that estimates future outcomes or unknown values based on historical patterns. Includes forecasting (time-series prediction) and regression (estimating continuous values).

Ranking and decision systems
AI capability that orders options or alternatives according to learnt preferences or criteria. Used in matching, recommendation, and priority scoring.

SP pillar
Social protection pillar: social assistance (non-contributory), social insurance (contributory), or labour market programmes (active labour market policies).

Synthetic data generation
AI capability that creates artificial data resembling real data. Used for privacy-preserving analysis and training data augmentation.

IMPRINT

Published by
Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

Registered offices
Bonn and Eschborn

Friedrich-Ebert-Allee 32 + 36
53113 Bonn, Germany
T +49 228 44 60-0
F +49 228 44 60-17 66
info@giz.de

Dag-Hammarskjöld-Weg 1–5
65760 Eschborn, Germany
T +49 6196 79-0
F +49 6196 79-11 15

Required citation
DCI AI Hub for Social Protection, *AI Adoption in Social Protection: A Global Evidence Review (2020–2025)*, Deutsche Gesellschaft für Internationale Zusammenarbeit, Germany, 2026.

Authors
Thomas Byrnes
Avril James

Responsible for the content
Bernd Schramm, GIZ

Editing
Susan Sellars

Layout
EYES-OPEN, Berlin

Bonn, August 2026

AI Hub for Social Protection

The AI Hub is a global partnership, with close cooperation among nine core implementing partners: German Social Accident Insurance Institution for the Construction Sector (BG BAU), Government of Canada – Employment and Social Development Canada (ESDC), Dataprev (the Social Security Information and Technology Enterprise of the Brazilian government), Federation of German Pension Insurance Institutions (DRV Bund), Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), International Labour Organization (ILO), International Social Security Association (ISSA), Organisation for Economic Co-operation and Development (OECD) and the World Bank.

The AI Hub is implemented by:

BG BAU Berufsgenossenschaft der Bauwirtschaft

Government of Canada / Gouvernement du Canada

DATAPREV

Deutsche Rentenversicherung

giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

issa

OECD

THE WORLD BANK

International Labour Organization