OECD*publishing*

AI AND THE FUTURE OF SOCIAL PROTECTION IN OECD COUNTRIES

OECD ARTIFICIAL INTELLIGENCE PAPERS

June 2025 **No. 42**

OECD
BETTER POLICIES FOR BETTER LIVES

OECD Artificial Intelligence Papers

AI and the future of social protection in OECD countries

Disclaimers

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Member countries of the OECD.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Photo credits: Cover © Kjpargeter/Shutterstock.com

Corrigenda to OECD publications may be found at: www.oecd.org/en/publications/support/corrigenda.html.

AI and the future of social protection in OECD countries

Governments in OECD countries are increasingly making use of advanced technology and data to improve the coverage, effectiveness and efficiency of social programmes, yet they are proceeding with caution when introducing artificial intelligence (AI). Common AI uses in social protection include client support, automating back-office processes, and fraud detection. Looking ahead, there is significant potential for AI to help improve the performance of social programmes – including through predictive analytics, enhanced outreach, and better-tailored interventions – but governments must continue to build trust and foster transparency when using AI.

Acknowledgements

This paper was written by Valerie Frey, Doron Wijker and Raphaela Hyee in the OECD Directorate for Employment, Labour and Social Affairs. The authors thank the Delegates to the OECD Working Party on Social Policy, as well as Dorothy Adams, Marc Folch, Pauline Fron, Marguerita Lane, Annikka Lemmens, Mark Pearson, Veerle Miranda, Stefano Scarpetta, Jasmin Thomas, Monika Queisser and members of the OECD AI Review Board for their thoughtful contributions, feedback and suggestions. The discussions with policymakers and programme providers at the OECD's AI and the Future of Social Protection webinar series have fed into this paper and are much appreciated. Natalie Corry, Marie-Aurélie Elkurd and Hanna Varkki provided valuable logistical and publication support.

Table of contents

FIGURES

Executive summary

Governments in OECD countries are increasingly making use of advanced technology and data to improve the coverage, effectiveness, and efficiency of social programmes. Within this broader technological transformation, governments are proceeding with caution in their applications of artificial intelligence (AI) in social protection, reflecting the significant risks and challenges involved – including potential data privacy risks and errors in automated decision-making.

The use of AI in social programmes today largely entails client support, automating back-office processes, and fraud detection. Yet there is significant potential for AI to help improve the performance of social programmes. Expanding upon findings in the report *Modernising Access to Social Protection* (OECD, 2024[1]), this paper identifies current use cases and elaborates on potential new uses of AI for social protection, including predictive analytics to forecast demand and shocks; predictive analytics to improve client identification and early intervention; enhanced outreach and reducing non-take-up; better-tailored interventions to support client needs and meet programme goals; and addressing (real or perceived) human discrimination in claims processes.

Governments are proceeding with caution as they identify policy challenges that AI can meaningfully address (relative to other tools), establish supporting infrastructure, and trial small-scale projects to test whether processes can be safely rolled out at larger scale. Public support for the government's use of AI is tepid: evidence from the OECD's Risks that Matter Survey finds that across 27 countries, only 40% of respondents feel that governments' use of AI to help process and approve social programme applications is good for users. Greater engagement with the public is therefore needed to foster transparency and build trust in the use of AI for social protection.

Countries are harnessing data and technology to modernise social protection

OECD governments increasingly rely on advanced technologies and data to enhance users' experiences of public services, including social programme applications and delivery, and to ensure that people are aware of, and can access, the benefits and services for which they are eligible.

Digital technologies are changing the nature of the bureaucratic encounter between the state and people. More services are now available online, enabling agencies to focus (human) resources on people whose needs are not suited to automated systems, such as service users with complex needs. OECD countries are also increasingly linking administrative data across sources to lower the administrative burden on claimants, make information more readily available, and measure non-take-up (OECD, 2024[1]). Digital tools and strengthened data capacity are helping to improve programme enrolment and service delivery. They offer significant potential to improve operations efficiency and reduce costs – a particularly important goal in the face of ongoing budget constraints (OECD, 2024[2]).

In general, however, advanced uses of technology and data – including data linking – are less common in OECD countries' public sectors than in the private sector, and are less common in social protection than, for example, the health sector (OECD, 2024[3]).

In light of governments' interest in the potential of artificial intelligence (AI) to improve social programmes, the risks and implementation challenges involved, and the rapid pace and scope of AI technology development, this brief presents a snapshot of current use cases of AI in OECD countries' social protection systems. This paper expands upon findings in the report *Modernising Access to Social Protection* (OECD, 2024[1]) and suggests new ways by which AI could benefit social protection.

AI has tremendous potential to transform social programmes

AI is reshaping societies and economies, and it is opening new opportunities for governments. Proponents of AI argue that it could radically improve the efficiency and quality of public service delivery in areas such as education, healthcare and social protection, bringing significant social and economic benefits. Attention to the potential and use of AI in public services has consequently grown significantly. By May 2023, 51 countries globally had reported a national AI Strategy to the OECD, a dramatic increase from 2017 when only a handful of countries had an AI strategy (OECD, 2023[4]).

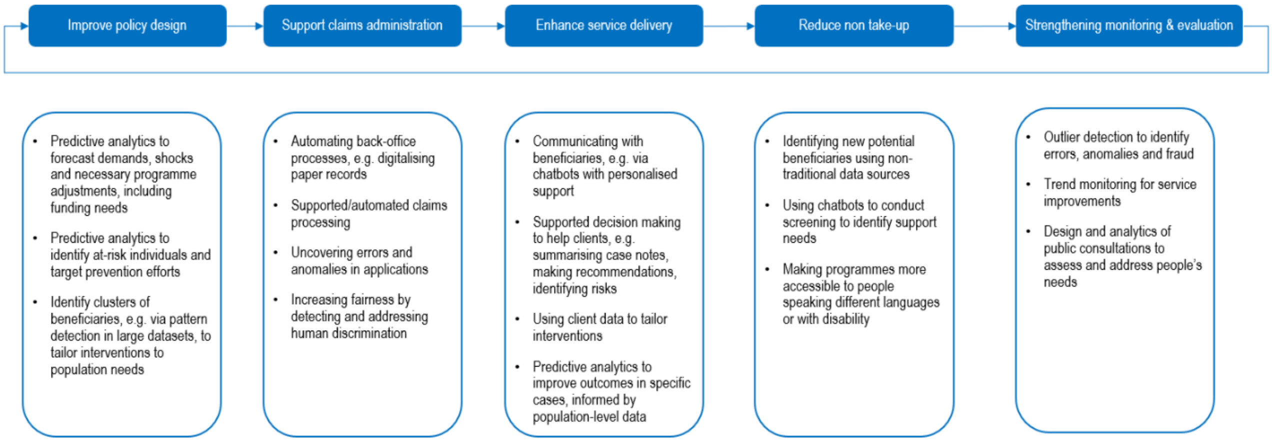
A few countries require government ministries or agencies to publish their AI use cases. One such example is the U.S. Department of Health and Human Services' "Artificial Intelligence Use Case Inventory," which published around 270 instances of AI uses in 2024 (HHS, 2025[5]) as mandated by executive order (Executive Office of the President, 2020[6]). The "Public Sector Tech Watch" Observatory, hosted by Interoperable Europe, had identified 58 AI use cases by national governments in the European Union related to social protection (as well as 40 use cases at the local level and 17 at the regional level) as of May 2025.

The **potential applications of AI in social protection systems are significant, with the potential for AI to improve policy design, support claims administration, enhance service delivery, predict programme demand, reduce non-take-up, and strengthen monitoring and evaluation** (for example to reduce fraud). These potential uses are previewed in Figure 1. AI could be used to improve access to and delivery of social protection, for example through more precise targeting of programmes to eligible beneficiaries, providing targeted information about available programmes and benefits, making faster and more accurate eligibility decisions, adjusting benefits in real-time, monitoring and managing benefit delivery, identifying at-risk populations to target prevention efforts, and automating or simplifying tasks, freeing up time for stretched social service workers to spend with clients.

Of course, AI models have broad applications (see Box 1) and given the pace of technological development, more applications will almost certainly emerge in the coming years (OECD, 2024[7]).

Figure 1. Current and potential uses of artificial intelligence in social protection

Schematic of current and likely upcoming use cases for AI and social protection



Note: OECD overview of current and hypothetical uses of AI for social protection.
Source: Authors' illustration.

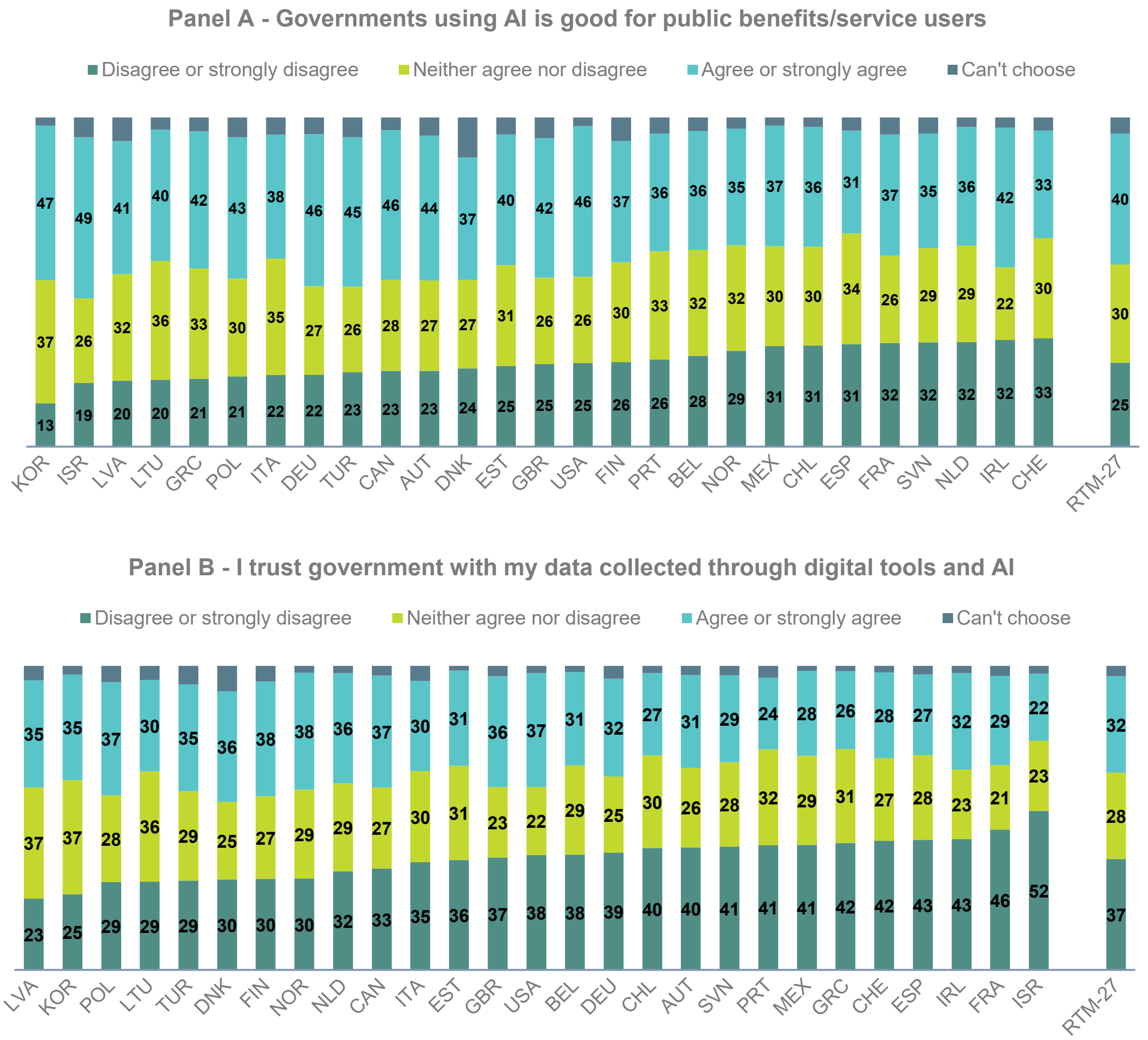
Yet **significant risks and challenges persist** around the use of AI in public services. The use of large data sets with sensitive information raises significant privacy concerns. Process automation could de-humanise services or decrease transparency, particularly if there is not enough human involvement or oversight, or if the AI systems used to make decisions are not well-understood or explained. The use of AI raises questions about who is responsible and liable if something goes wrong, and algorithms trained on unrepresentative or biased data sets risk (re)producing biased and discriminatory outcomes. For example, a risk when using AI to manage social benefits is that assessments of benefit eligibility may be incorrect or systematically biased against certain demographic groups (Adams, 2024[8]). There are also considerable environmental concerns given the significant computing power that will be needed to power AI systems (Hodgkinson, Jennings and Jackson, 2024[9]).

Perhaps reflecting these challenges, **much of the public does not yet trust the use of AI in social services**. Results from the 27-country OECD Risks that Matter (RTM) survey highlight skepticism about the benefits of governments' AI use for service users, as well as a lack of confidence in how the government will use the data collected about them through digital tools and AI. Across representative samples in 27 countries, only 40% of respondents feel that the use of AI to help process and approve social programme applications is good for users (Figure 2 Panel A). 30%, on average, express uncertainty, and 25% do not believe that the governments' use of AI is good for users of social programmes (OECD, 2025[10]).

When asked whether they trust the government with the data collected about them through digital tools and AI, 37% disagree or strongly disagree, while 32% indicate that they do trust the government with the data collected (Figure 2 Panel B). Fewer than half (44%) of respondents were confident that AI would only be deployed to assess public benefits applications when it is safe and trustworthy to do so ((OECD, 2025[10]). These findings suggest that governments have room to improve when communicating about, and building trust in, the use of AI in social policy.

Figure 2. Many people do not yet trust the use of AI in public and social services

Distribution of responses to the statements "Governments using AI to help process and approve applications is good for users of public benefits/services" (**Panel A)** and "I trust the government with the data they collect on me through digital tools and AI" (**Panel B**), in response to the question, "Thinking about the use of digital tools and artificial intelligence (AI) by government, to what extent do you agree or disagree with the following statements," 18- to 64-year-olds, 2024



Note: Respondents were asked "Thinking about the use of digital tools and artificial intelligence (AI) by government, to what extent do you agree or disagree with the following statements?" and presented with a list of statements regards government, digital tools, and artificial intelligence, including: "Governments using AI to help process and approve applications is good for users of public benefits services" and "I trust the government with the data they collect on me through digital tools and AI". Response options were: 1. Strongly disagree, 2. Disagree, 3. Neither agree nor disagree, 4. Agree, 5. Strongly agree, 6. Can't choose. RTM-27 refers to the unweighted average for the 27 OECD countries participating in OECD Risks that Matter survey 2024.
Source: 2024 OECD Risks that Matter Survey (https://oe.cd/rtm).

These findings mirror national surveys. A 2022 survey in Australia found that while 60% of Australians support the development of AI in general, this share drops significantly when asked about specific social service scenarios. Only 31-39% of respondents support the development of AI across different social service situations, with a solid majority of respondents (four out of five) favouring human contact and discretion in health and social care over the speed, accuracy and convenience benefits that AI might bring (Isbanner et al., 2022[11]). Respondents were particularly conscious of the accuracy and fairness of AI.

Concerns over the use of AI have loomed large in policymaking, too. Reasons for the relatively slow deployment in the government sector include ethical and legal concerns, as well as scepticism about whether computer-driven systems are appropriate in the sphere of public policy and administration (Ohlenburg, 2020[12]). Leveraging AI also of course requires complementary investments in data, skills training and digitalised workflows, as well changes to organisational processes.

Given the risks and challenges involved, countries are establishing the necessary foundations and capacity to safely advance the use of AI applications. Such examples include the United Kingdom's "AI Lighthouse Programme" and the 2020 United States Executive Order 139 060, "Promoting the use of trustworthy AI in the federal government". To increase transparency, the U.S. Executive Order requires making inventories of federal government use cases of AI available to the public[1] (e.g. (HHS, 2025[5])). Similarly, Scotland's "Artificial Intelligence Strategy" will require that "every project using AI […] be logged on the Scottish AI Register, a publicly-accessible database which provides a range of information about the use of the technology in projects developed by public bodies" (Government of Scotland, 2024[13]).

Countries are thinking carefully about how best to take advantage of new technologies and are proceeding with caution, implementing and monitoring small-scale projects before determining whether to scale up. Many uses of advanced technologies in social services, and AI in particular, continue to be small, ad hoc test cases to determine feasibility, functionality and scope of deployment. Currently the use of AI in social protection is focussed on client support, task automation and the detection of error and fraud, with more advanced applications – for example the use of predictive analytics for prevention – still somewhat limited.

Box 1. What is AI? Categories, applications and uses of Artificial Intelligence

Definitions of AI are often broad, and it can be tempting to categorise most examples of analytics as AI. However, not all analytical methods deploy AI techniques. The OECD describes an AI system as a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. The OECD aimed for a description that was understandable, technically accurate, technology-neutral and applicable to short- and long-term time horizons.[2]

Different AI systems vary in their levels of autonomy and adaptiveness after deployment. They can use machine and/or human-based inputs to perceive real and/or virtual environments; abstract such perceptions into models (in an automated manner e.g. with machine learning[3] or manually); and use model inference to generate content, and/or formulate options for information or action. AI systems are designed to operate with varying levels of autonomy.

AI systems are typically built on one or more models that infer outputs based on specified inputs, including:

Machine learning (ML) techniques, which use algorithms to identify and learn patterns and generate models from data, learning and improving their performance over time. These models are 'trained' on a dataset to generate outputs such as predictions or decisions, without explicit instructions from a human operator, and feedback can help the model to 'learn' and become more accurate. For example, machine learning techniques can be used to predict which households are at risk of homelessness based on historical data and patterns that have been associated with homelessness risk, such as missed utility payments. ML techniques often teach machines to reach an outcome by showing them many examples of correct outcomes.

Symbolic or knowledge-based AI systems rely on explicit descriptions of variables and their relationships to make logical and/or probabilistic inferences from data.

Application areas of AI models are broad, including natural language processing, intelligent decision support systems and computer vision (OECD, 2024[14]). Natural language processing (NLP), for example, is a subset of AI that automates natural language functions such as analysing, generating or responding to text (OECD, 2023[15]). This includes large language models (LLMs) such as ChatGPT that use machine learning techniques – trained on vast amounts of data – to understand and respond to text, generating conversational interactions between machines and people. Generative AI systems that produce 'content' such as text or images have become such an important and widespread type of AI that content was included as its own output category in the OECD's 2023 revised definition of AI.

Source: Adapted from OECD (2024[7]), Explanatory memorandum on the updated OECD definition of an AI system, https://doi.org/10.1787/623da898-en; OECD (2023[15]), AI language models: Technological, socio-economic and policy considerations, https://doi.org/10.1787/13d38f92-en.

Client support and automation at the centre

AI-powered chatbots and digital assistance

The use of chatbots and digital assistance to improve online customer services, with 24/7 availability and in different languages, is increasingly common. These intelligent assistants mimic human behaviour and autonomously respond to user requests. Client-facing digital assistants emerge as a relatively common use of AI in OECD countries (OECD, 2024[1]) and worldwide. In a 2021 ISSA AI adoption survey of 166 government agencies across the world, chatbots emerged as the frontrunner, with 26% of respondents already implementing them and another 59% planning to implement them within three years. And in a review of 230 AI-enabled public services across the EU, chatbots emerged as the most common use, accounting for over one-fifth of use cases (ISSA, 2022[16]).

Chatbots are used at least in **Austria**, **Australia, Belgium, Finland, France, Germany, Korea, Norway, Spain** and the **United States** to support social service delivery at the national or subnational level. In most cases chatbots have been deployed to provide information to citizens and respond to their queries, while a handful of countries use chatbots to support prevention, for example to check on people's well-being. The Social Insurance Institution (Kela) in **Finland**, for example, has two chatbots, Kela-Kelpo and FPA-Folke, to help clients find information about benefits on Kela's self-service web portal. The chatbots speak two languages – Finnish and Swedish – and they also understand English, helping people obtain information and complete benefit applications. They also provide customised tips based on contextual variables as clients fill out applications for specific programmes, such as parental benefits. **In Germany,** the "Digital First Contact" chat and telephone bot is being developed to improve clients' access to the most appropriate counselling services (Public Sector Tech Watch, 2025[17]). In **Korea**, an AI-driven personalised conversation service (which remembers past conservations and uses them for the next call) is used to check on people's well-being once or twice a week, chatting with people for about two minutes (OECD, 2024[1]). In the **United States**, the Department of Health and Human Services has used AI-powered chatbots for a range of tasks, including answering questions by grant applicants and providing support to users of the "Child Welfare Information Gateway" (U.S. Administration for Children and Families, 2024[18]).

Chatbots were particularly valuable in providing information during the COVID-19 pandemic, signalling their significant potential to scale up access to support, and to free up caseworker time for more complex tasks. Between March and May 2020, **Norway's** Labour and Welfare (NAV) administration's intelligent conversation assistant responded to more than 8 000 daily requests, compared to 2000 before COVID-19.

A review found the main success factors were training the assistant based on a knowledge base updated daily and having a permanent link between the assistant and a human expert. New topics were added, especially to help employers and freelancers (ISSA, 2020[19]). **Belgium** and **Finland** also deployed chatbots to ease contact centre pressures during COVID-19 (ibid).

Not all chatbots, however, are developed using AI technology. Some chatbots only replicate human actions and tasks in digital systems, typically tasks that are repetitive and rule-based i.e. they do not think or learn. The **Canada Revenue Agency** (CRA), for example, is developing Chatbot and Online Chat solutions which are based on varying maturity levels of AI. The existing chatbot is rules-based using a question-and-answer model. It can only respond from its knowledge base with answers that match a specific set of topics.

Automating back-office processes

Governments are using AI to automate back-office processes and reduce the time spent by civil servants on administrative tasks. Examples include processing large amounts of data from traditional databases and unstructured text and images from scanned paper media. This has the potential benefit of freeing up staff hours to work with beneficiaries who may need support.

Austria, **Canada, Finland** and the **United States** are currently using AI techniques to automate back-office processes in social protection. Employment and Social Development Canada (ESDC), for example, has leveraged natural language processing (NLP) to automate the review of free-text comments received on records of employment, helping to reduce the manual workload of Service Canada officers (OECD, 2024[1]).

Finland's Social Insurance Institution (*Kansaneläkelaitos*) uses AI image recognition to automate administrative processes through document recognition, while the Austrian Social Insurance Agency (*Dachverband der österreichischen Sozialversicherungsträger*) has used a voice recognition system to support call centre services by automatically forwarding requests to the appropriate offices, and has used AI to automatically distribute emails to relevant departments, with an accuracy rate of around 93%. In addition, there is an ongoing project to implement a semi-automated AI-based medical reimbursement process where AI is used to automate several tasks such as recognising submitted documents, classifying diagnostics according to ICD-10 codes, and extracting the data required for reimbursement (e.g. invoice amount and IBAN). Semi-automatic processing speeds up the reimbursement process and supports the staff involved (ISSA, 2020[19]). In Brazil, the National Social Security Institute uses AI to speed up the identification of deceased beneficiaries to avoid undue payments, which in many countries is processed manually (and slowly) (ISSA, 2020[19]).

AI is being used to streamline claims processing, too. COVID-19 triggered an unprecedented volume of assistance and benefit claims. In Canada, the focus during the pandemic on implementing the Emergency Response Benefit (ERB) and simplified Employment Insurance (EI) claims contributed to a backlog of claim reviews when regular processing resumed. Implementation of a Pre-ERB EI Recalculation Outcome Prediction Machine Learning model sought to minimise the number of older claims (pre-March 2020) requiring review by an officer. The model was used to predict the most probable outcome of each recalculation and triage the associated work items accordingly, taking into account if recalculations were likely to impact claimants (OECD, 2024[1]).

In the United States, the Department of Health and Human Services has deployed AI for a range of back office, "mission-enabling" tasks, including document summarizing, drafting and editing content, qualitative data analysis, literature review support, designing the interior layout of public facilities (e.g. for housing shelter), and others. Out of the 272 listed for 2024, 133 use cases are categorised as "mission-enabling" (HHS, 2025[5]). Japan, too, notes that the AI is being used in claims processing to identify cases that need human review, and that this review rate has declined over the past two years.

Error and fraud detection by providers, and feedback by clients

AI also has considerable potential for both government and clients to improve the quality and accuracy of claims and to improve service delivery. Some countries are already using AI techniques for error and fraud detection. The **United Kingdom**, for example, is using AI to help detect fraud in social benefit claims (Marr, 2018[20]). **Korea** has made considerable advancements in using big data for error and fraud detection among health insurance claims. Korea's National Health Insurance Service (NHIS) houses big data on a range of socio-economic, health behaviour, healthcare utilisation and long-term care variables to which smart audit algorithms are applied to predict healthcare facilities with high probability of fraudulent claims, thereby pre-emptively supporting investigators (ISSA, 2022[21]).

Fraud detection is one area where the use of AI techniques is likely to increase rapidly. Just as AI will likely drive an increase in the volume and sophistication of fraud and scams, it will also provide detection and prevention solutions. In Portugal, for example, AI-based face or voice recognition is being used for proof of life checks for civil servants living abroad to keep collecting their pension.

Fraud detection is also an area where ethical concerns have been prominent. In the Netherlands, between 2005 and 2019, many families were falsely accused of fraud due to discriminatory algorithms (discussed below). As governments look to scale up the use of AI in this space, it will be particularly important to implement appropriate safeguards against bias.

Another promising use of AI to promote government accountability and quality improvement – currently not routinely deployed – is complaints monitoring. Machine learning techniques have been used to identify factors that increase the likelihood of a complaint, which could help policymakers and practitioners identify potential gaps in service delivery and prioritise areas for improvement (Ohlenburg, 2020[12]).

Predictive analytics, supported decision-making and tailored interventions: The future of AI in social protection?

Potentially one of the most rewarding use cases for AI in social protection lies in the domain of predictive analytics. Predictive analytics could play a significant role in preventing and responding to shocks or crises; identifying at-risk individuals and targeting preventative efforts; identifying groups of (potential) beneficiaries (e.g. by region or sociodemographic traits) to tailor interventions better; and improving outreach and reducing non-take-up of social programmes.

Predictive analytics to forecast demand and shocks

Predictive analytics are already being used to identify individuals at risk of experiencing a shock or adverse event, to enable action to avoid the event, or to mitigate the impact of the shock where it cannot be prevented. AI tools are being used to forecast the likelihood of occupational accidents in **Germany**, for example, and are helping to prevent accidents by identifying companies with the greatest need for guidance, and allocating site inspectors accordingly. The model aggregates data from building sites, workers and employers to identify risks on building sites, effectively implementing a risk-based approach to allocate scarce resources in circumstances where inspectors can reach only 1% of companies with inspections (GIZ, 2024[22]). Inspectors report with voice notes which are transformed into written reporting, with the results used to improve the machine learning model. The project seems to be unlocking value, with automated recommendations freeing up time for workers to spend on inspection, reported to have reached over 4 000 sites that had not previously been inspected, and generating estimated savings of around EUR 250 million in compensation and rehabilitation costs over a five year period (GIZ, 2024[22]). AI could possibly be used in the future for similar predictive analytics by labour inspections, e.g. to predict the likelihood of labour violations and estimate demand for inspections. This type of AI system could also be used, for example, to establish which day-care services most merit inspection in situations of limited resources.

Predictive analytics have also been used to mitigate and help cushion the impact of shocks where they cannot be avoided, including, for example, in Togo where it has been used to forecast floods, to notify citizens to take action to mitigate the impact of flooding, and to make financing available before a flood hits (Ohlenburg, 2020[12]). Similar applications have been explored in **Europe**, where researchers have collaborated with the European Space Agency to explore the use of AI to develop a prediction and early warning system for floods, and in Ireland, where government-backed, not-for-profit 'CeADAR' has used AI to predict future flooding in flood-prone areas (Lancaster University, 2021[23]); (CeADAR, n.d.[24]).

Predictive analytics for identification and early intervention

The possibilities for new technologies to identify and enrol potential beneficiaries in social programmes is significant, although it is worth noting that major gains could also be made using existing tools and good-quality, linked personal data (Frey and Hyee, 2024[25]). Looking ahead to ongoing and potential use cases for AI, there are opportunities both in identifying and enrolling people who have not been in touch with social protection agencies, as well as people who are already integrated in some social programmes – and who are likely eligible for others.

Predictive analytics could also help to improve outcomes by supporting providers' decision-making. By quickly and systematically processing and summarising large, complex and unstructured datasets, AI could help policymakers and practitioners to identify and prioritise risks and develop targeted and personalised interventions.

Given the substantial returns on investment of some preventive interventions in social services, the social and economic benefits of predictive analytics could be significant. One area where this already is being done, with predictive modelling, is in homelessness prevention and alleviation. In Los Angeles, **United States,** academic researchers and the local government have partnered on a programme to identify and reach out to residents at the highest risk of first-time homelessness (California Policy Lab, 2024[26]). An individual's risk is predicted using data on risk factors among single adults already receiving "mainstream" social services, with the predictive model suggesting that high (multiple), increasing and spikes in service use[4] provide warning signs of high homelessness risk for people experiencing deep poverty (California Policy Lab, 2020[27]). A similar initiative was implemented in the United Kingdom, where Maidstone Borough Council and external partners developed 'OneView', a data and predictive analytics tool that aggregates data from different agencies to enable frontline officers to identify and reach out to residents at risk of homelessness (Šuica et al., 2024[28]). The platform sends risk alerts to frontline officers for events that might indicate a risk of homelessness, such as missed utility payments (Šuica et al., 2024[28]), with promising results (Maidstone Borough Council, 2022[29]) (Ernst & Young, 2022[30]).

Machine learning techniques have also been used to **improve predictions of domestic violence recurrence**, improving upon existing risk assessments. Leveraging police data in the **United Kingdom**, researchers have found that machine learning algorithms that consider criminal histories perform better than traditional risk assessment tools in predicting violent recidivism in domestic abuse cases, assuming that over detection is better than under detection (Grogger, Ivandic and Kirchmaier, 2020[31]). The negative prediction error for the algorithm – the share of cases that would wrongly have been predicted not to have violence – is a relatively low 6.3%, compared to the 11.5% result of the traditional screening tool (Centre for Economic Performance, 2020[32]). Researchers suggest that machine learning could help to overcome low predictive power and considerable inconsistencies in the application of existing screening tools, and could help to prioritise incoming domestic abuse calls based on risk, to ensure that the highest risk calls get the quickest responses. The researchers nonetheless call for development of a more sensitive screening instrument to address false positives that emerge from such screening, which could be particularly important. Tolerance of false positives will likely also hinge on what action a risk flag generates:

tolerance for false positives would likely be higher if the prediction is used to triage calls than if the prediction is used for more severe action.

Another area with enormous potential for prevention and early intervention is in **care and income support for older people.** Countries such as **Sweden** and the **United Kingdom** are using AI to identify rehabilitation needs and injury risk, respectively, helping to tailor early interventions. An AI model has been used in Sweden to forecast both the need for and potential of preventative services, aiming to improve resource allocation through targeted preventive support, while the United Kingdom (England) is rolling out an AI tool in home care visits with a reported 97% accuracy in predicting patients' risk of falls, a leading cause for emergency hospital admissions for older people (Šuica et al., 2024[28]); (NHS England, 2025[33]). This is an area where prevention and efficient resource allocation are becoming increasingly important, with recent OECD work showing that population ageing and declining fertility will significantly increase demand for health and social care in the coming years, while shrinking both the workforce and the funding available for care (OECD, 2024[2])**. AI is also being used to identify poverty and debt risks preventatively** through the analysis of data from different agencies in different municipalities in the **Netherlands** (Public Sector Tech Watch, 2025[17]).

Algorithm-based decision support systems have also been developed or trialled to **help social workers identify possible child maltreatment and/or child abuse** in countries such as **Australia, the Netherlands, New Zealand and the United States**, though some researchers suggest that additional work is needed to ensure that such tools are fit-for-purpose (Gillingham, 2019[34]) (Hall et al., 2024[35]). Machine learning techniques were piloted to identify and flag possible cases of child abuse in the Netherlands, for instance, where researchers developed a decision support system combining data on child height and weight with free text comments from health practitioners, to alert workers to possible cases of child abuse (Amrit et al., 2017[36]), cited in (Robila and Robila, 2020[37]). Such approaches offer the potential for improved identification and intervention without the need to collect additional data (Robila and Robila, 2020[37]), and they could have wide applications in social services to supplement and/or improve existing risk assessment methods.

The risk of false positives necessitates caution and risk mitigation on the part of policy makers and service providers. Flagging a family as being at high risk of adverse outcomes comes with serious implications – for families' relationship with the state and for welfare in the family, to name a few. The reported use of software to monitor school devices to prevent self-harm among students in the United States, for example, has received a mixed response. In some cases, the detection has enabled a timely response to children and young people at risk of self-harm or suicide, but many advocates and parents are concerned about false positives (e.g. flagging a student as suicidal when they are not) and ongoing surveillance of young people (Barry, 2024[38]); (American Academy of Pediatrics, 2025[39]).

Improving outreach and reducing non-take-up in programmes

In addition to improving the identification and prevention of risks amongst individuals and households already in contact with social programmes, AI could help to proactively identify and reach out to *potential* beneficiaries of social programmes to address non-take-up by people who are likely eligible.

Non-take-up of social programmes remains a persistent challenge throughout OECD countries (Frey and Hyee, 2024[25]). Some populations that are most in need of social benefits and services can be difficult to reach and unlikely to actively seek out help. For instance, young people not in education and training hugely benefit from early interventions to prevent long-term joblessness, but it can be difficult to enrol them in social programmes as they are unlikely to be aware of, or reach out to, the relevant agencies (OECD, 2024[40]). Some countries do actively try to contact these young people, but it is research intensive, and many jobless young people increasingly spend time online instead of in public spaces. In a few cases, AI is being used to help identify people in need of support and proactively offer it.

In **Spain**, the social services department of Madrid City Council uses an AI-powered virtual assistant to **phone people over the age of 75 to assess whether they might be at risk of loneliness**. The AI-powered tool asks a set of structured questions to identify whether these people feel lonely, have family, friends or someone they could turn to, or wish to have a follow-up from municipal services. As a result, more than 600 people have been identified and contacted by caseworkers from the social services department (Lara-Montero, 2024[41]). **Greece**, too, has personalised interventions that help to improve safety, reduce feelings of loneliness and strengthen independence (Public Sector Tech Watch, 2025[17]).

Of course, AI is not a *necessary* tool for identifying groups of people in need. As with many of the other examples presented here, more traditional data analysis has been used to identify groups of individuals or households with similar needs, often using survey data to produce probabilistic estimates that can be disaggregated by different sociodemographic profiles (Frey, Hyee and Minondo Canto, 2024[42]).

Tailoring interventions to the needs of clients

Another potential use of AI in social programmes is to better tailor interventions to the needs of users. AI could support greater personalisation at a population level – for example, clustering beneficiary groups for social programmes to tailor service offerings to their needs (Ohlenburg, 2020[12]) – or at an individual level. In **Korea,** the Workers' Compensation and Welfare Service (COMWEL) has developed the Intelligent Rehabilitation Recommendation System (IRRS) to enhance its support to injured workers. While COMWEL has been implementing customised rehabilitation plans for injured workers since 2011, the process has relied on limited information and the experience of managers in charge, resulting in variable service quality and timeliness. IRRS, an AI-based system, was developed to select the injured workers with the potential to be active, and design scientifically tailored rehabilitation services for them. The IRRS calculates a vulnerability index based on administrative data on 98 million workers accumulated since 2011, comprising details about workers' compensation, unemployment insurance and rehabilitation case management, using rule-based filtering and case-based reasoning methodology.

Korea's IRRS also suggests a rehabilitation plan based on the AI model. The workers selected for rehabilitation and return to work undergo consultation with the rehabilitation experts of COMWEL before AI-generated plans are finalised. The IRRS is reported to have helped COMWEL improve the consistency of service quality nationally while ensuring timely and appropriate interventions to ultimately improve the return-to-work ratio (ISSA, 2022[21]).

The **Danish Agency for Labour Market and Recruitment** (STAR) has developed a profiling model using machine learning techniques that predict the likelihood of people becoming long-term (>26 weeks) unemployed. The model combines data from administrative records and an online survey that gathers behavioural information. In collaboration with the University of Copenhagen, a new survey instrument is currently being developed that aims to capture structural personality traits such as time and risk preferences. The system is voluntary for jobseekers to use but if they do, they get full access to the model's results. The system does not automatically refer jobseekers to active labour market programmes (ALMPs), rather it supports caseworkers who keep full discretionary responsibility (Desiere, Langenbucher and Struyven, 2019[43]).

The use of AI in Public Employment Services (PES) is extensive. OECD research on the use of AI in OECD countries finds that almost half of PES are utilising AI to enhance their services, most commonly to match jobseekers with vacancies and to identify jobseekers' needs for support using profiling tools (Brioscú et al., 2024[44]). In the **Netherlands**, the "CompetentNL" programme uses AI to facilitate skills matching with jobs (CompetentNL, 2025[45]), while in **Portugal**, the employment agency uses AI to better identify individuals at high risk of long-term unemployment and support the more efficient allocation of agency resources to respond to the needs of people who are unemployed (Public Sector Tech Watch, 2025[17]). For a comprehensive review, see (Brioscú et al., 2024[44]).

Governments are also starting to implement AI to help case workers in their day-to-day work, including by supporting their decision-making. In **Barcelona, Spain**, for example, an algorithm trained on 300 000 interviews proposes resources and services to the 700 staffers serving around 50 000 clients per year, though the ultimate decision comes from the staff member (Public Sector Tech Watch, 2025[17]). Several local councils in the United Kingdom have piloted an AI note-taking tool that can capture, transcribe and produce recommendations from client meetings, freeing up more time for workers to spend on frontline care and support. The tool can produce a proposed list of actions for workers to review and decide on. Camden Council in London, England (**United Kingdom**), for example, has piloted the tool to create a summary of conversations with residents, carers and other service providers, and to produce a suggested list of actions (Camden Care Choices, 2024[46]). At Ealing Borough Council in London, England (**United Kingdom**), social workers use the tool to transcribe client meetings – with clients' consent – which is reported to have led to at least a 40% reduction in the time social workers spend on administration (Local Government Association, 2025[47]) (OECD, 2025[48]).

Combatting (actual or perceived) human discrimination in the claims process

The fact that AI can be prone to racial, ethnic or gender bias is much discussed (Parikh, Teeple and Navathe, 2019[49]) (Ntoutsi et al., 2020[50]) (Booth, 2024[51]). Of course, human decision-making is not devoid of bias, either, and AI could in fact be deployed to minimise bias or the appearance of bias – though this use is not yet reported in practice by ministries in OECD countries.

Many social programmes have clear and objective entitlement rules – such as maximum income or assets, or past contributions – but other rules can be open to interpretation. For instance, there is an element of judgement in determining sufficient job search efforts for unemployment benefit recipients. Many benefits have 'behavioural conditions' that recipients must meet, such as attending meetings with caseworkers or timely submission of updated information. Non-adherence to these conditions results in sanctions, such as partial or full benefit withdrawals (Hyee et al., 2024[52]). Since there is room for discretion in deciding whether behavioural requirements are being met or not, there is room for racial / ethnic / religious / gender-based discrimination in benefit sanctions.

Discrimination in the granting / withdrawal of social benefits can be due to (explicit or implicit) bias on the part of caseworkers. This is more likely in the decision of whether a claimant is entitled to the benefit or if it is legitimately complex and ambiguous (see, for example, (Assouline, Gilad and Ben-Nun Bloom, 2021[53]) and (Emeriau, 2022[54]).

AI could help detect and flag discrimination in benefit-granting decisions, especially where decisions are made in ambiguous circumstances (such as disability, but also job-search intensity or similar), or by less skilled or experienced caseworkers.

According to the limited literature, such discrimination is mainly statistical, and not taste-based (see, e.g. (Bell and Jilke, 2024[55])), meaning that detection and reporting could improve the accuracy of decisions. For example, when access to a benefit or service is limited – e.g. a certain programme is oversubscribed – caseworkers often prioritise clients with the highest probability of success. Caseworkers at public employment services may assign jobseekers they perceive as having the highest probability of finding a job to training programmes to increase the probability to get jobseekers off their books, a phenomenon known as "cream-skimming" or "creaming" (Carter and Whitworth, 2014[56]). If the true determinants for success in a given intervention are unknown, caseworkers may resort to using group-level means, and thus statistical discrimination.

Even if there is no actual discrimination in the claims process, potential claimants from disadvantaged groups can be reluctant even to apply for social benefits because they fear that they will be discriminated against. In the OECD's 2024 Risks that Matter Survey, people self-identifying as a minority based on ethnicity or skin colour were on average 3 percentage points more likely to feel that they *would not* be treated fairly by the government office processing their claim for public benefits or services, compared to people who did not self-identify as a minority based on ethnicity or skin colour.[5] This aligns with findings from the OECD Trust Survey which show that – on average across 29 OECD countries – there is a 14 percentage point gap in trust in government institutions between people who self-identify as belonging to a disadvantaged group (29.6%) and those who do not (43.3%) (OECD, 2024[57]).

The fact that bias in the claims process is likely to emerge in situations of ambiguity and limited information shows that there is scope for AI to improve fairness in the claims process, so long as appropriate oversight and safeguards are in place.

Putting safeguards in place to ensure the safe and trustworthy use of AI

While advanced technologies can play an important role in improving the design, delivery and coverage of social programmes, they also create complex challenges for governments, and the risks involved in adopting new digital and data technologies can be significant (Verhagen, 2024[58]). Governments must have in place appropriate accountability frameworks and procedures; without them, technology and data-driven innovations risk disempowering and disengaging people and eroding public trust and confidence. Principle 1.5 of the OECD's AI principles specifies that AI actors should be accountable for the proper functioning of AI systems, based on their roles, the context, and consistent with the state of art (OECD, 2019[59]).

High-profile implementation issues have brought the risks of AI into public focus, with automated fraud and debt recovery schemes in Australia and the Netherlands resulting in significant unintended consequences. In the Netherlands, nearly 26 000 families were falsely accused of fraud between 2005 and 2019 by the Dutch tax authorities due to discriminatory algorithms. Risk profiles were created for individuals applying for childcare benefits in which "foreign-sounding names" and "dual nationality" were used as indicators of potential fraud. Thousands of low- and middle-income families were then subjected to scrutiny, falsely accused of fraud, and asked to pay back benefits they had obtained legally, which in some cases amounted to thousands of euros. As a result, many went into debt, with some ending up in poverty and some losing their homes and/or jobs. More than 1 000 children were placed in state custody as a result (The European Parliament, 2022[60]). When the mistakes were revealed, the Dutch cabinet resigned, and subsequent assessment of the operation of the fraud detection programme led to the "painful" conclusion in 2022 that institutional racism had taken place (Government of the Netherlands, 2022[61]). The Netherlands today extensively tracks AI use in government; the Employee Insurance Agency, for example, has installed an advisory board on ethical AI that advises on all AI projects.

Australia's Robodebt scheme, introduced in 2015 to assess entitlements to payments, highlights the challenges of detecting a systemic issue in an automated (albeit not AI) system. From 2015 to 2019 the Department of Human Services implemented a debt recovery scheme – Robodebt – to recover overpayments to welfare recipients dating back to 2010-11. To calculate the overpayments, social security payment data was matched with annual income data from the Australian Taxation Office and a process known as "income averaging" was used to assess income and benefit entitlement. Debt notices would then be issued to affected welfare recipients who would have to prove they did not owe a debt. The process both produced inaccurate results and did not comply with the income calculation provisions of the Social Security Act 1991. Despite public criticism and adverse findings by the Administrative Appeals Tribunal to some individual cases, the systemic nature of the problem was not identified immediately, and the programme continued to operate until 2019. A class action was settled in June 2020 and an apology issued by the then-Prime Minister, with a subsequent Royal Commission into the Robodebt Scheme in 2022 making 57 recommendations.

Both the Robodebt and the Dutch childcare benefit events highlight the critical importance of transparency and explainability and the need for meaningful human involvement, particularly when automated decisions can potentially and significantly impact people's lives. The lack of transparency contributed to the Dutch scandal, with insufficient human involvement in, and oversight of, automated decision making (Błażej Kuźniacki, 2023[62]).

While risks exist in the deployment of any advanced technology, some have argued that the use of “black box” AI techniques should be avoided in areas such as social protection, resulting in some governments establishing additional measures to ensure AI use is advanced safely. The Department of Work and Pensions (DWP) in the United Kingdom for example has created an Artificial Intelligence (AI) Lighthouse Programme to safely explore their use of emerging Generative AI technology. Recognising opportunities and risks, DWP has established a framework and process to explore such technology in a safe, ethical and transparent way (Adams, 2024[8]).

Transparency and explainability is a key principle of trustworthy AI in the OECD’s AI principles (Box 2), which suggest that there should be a degree of human involvement in automated decision-making, proportionate to the potential impact of the outputs generated. Principle 1.2(b) of the OECD’s AI principles specifies that AI actors should implement mechanisms and safeguards, such as capacity for human determination, that are appropriate to the context and consistent with the state of art (OECD, 2019[59]). International regulations have also sought to establish guardrails for the safe and trustworthy use of AI and personal data, placing both restrictions and responsibilities on specified (and certainly high-risk) uses of AI in public and social services. This includes the European AI Act, which explicitly categorises AI systems intended to be used to evaluate, grant, reduce, revoke or reclaim eligibility or access to essential public assistance benefits and services (including social security benefits) as ‘high risk,’ and sets out a series of requirements for the permissible deployment of such high risk uses, covering requirements for both the AI systems and the organisations deploying them. This may impact how certain applications of AI in social protection can be designed and used.

Box 2. Transparency and explainability: Key principles of trustworthy AI

The OECD AI Principles, the first intergovernmental standard on AI, have been adopted by OECD Member countries and additional adherents to promote the innovative, trustworthy use of AI that respects human rights and democratic values. Transparency and explainability are key to trustworthy AI, as set out in the Principles, alongside inclusive growth, sustainable development and well-being; human rights and democratic values, including fairness and privacy; robustness, security and safety; and accountability.

Transparency involves disclosing when automated systems are being used, for example to make a prediction, recommendation or decision, with disclosure being proportionate to the importance of the interaction. Transparency also includes being able to provide information that allows those affected by an AI system to understand the output, including information about how an automated system was developed and deployed, what information was provided and why. An additional aspect of transparency is facilitating, as necessary, public, multi-stakeholder engagement to foster general awareness and understanding of automated systems and to increase acceptance and trust (OECD, 2022[63]).

Explainability is the idea that the outcome of an automated system or algorithm can be explained in a way that “makes sense” to people, enabling those who have been affected by an output to understand and challenge it. This includes providing – in clear and simple terms, and as appropriate in the context – the main factors included in a decision, the determinant factors, and the data, logic or algorithm used to reach a decision (OECD, 2022[63]). Some algorithms are more readily explainable but potentially less accurate (and vice versa). While requiring explainability may negatively affect the performance of an algorithm, it may in some cases be an outweighing factor.

The OECD AI Principles and specified data protection regulations also emphasise the need to implement appropriate safeguards to ensure accountability and fairness, particularly with respect to automated decision-making. In the EU, the General Data Protection Regulation (GDPR) enshrines a right for individuals not to be subject to (solely) automated decisions with only some exceptions, for example, where the use is legally permitted, *and* Member State lays down 'suitable measures' to safeguard the person's rights, freedoms and legitimate interests. The right to review an automated decision or output is an important feature of an accountability framework. Those negatively impacted by automated decision-making should be able to appeal a decision and know how to do that. As the OECD's 2019 Recommendation on AI specifies, those who are adversely affected by an AI system should be able to challenge the outcome(s) of the system based on easy-to-understand information about the factors that served as the basis for the prediction, recommendation or decision (OECD, 2019[59]).

Grievances and investigations should be taken seriously and made publicly available together with the outcome(s) so that lessons can be learned and shared with others undertaking similar work. Yet, while inserting humans into the loop of automated systems is a crucial way of helping to achieve accountability and oversight, this does not come without challenges. For example, what level of oversight, accountability and liability are attached to human-made decisions? What qualifications and/or expertise is required to question an automated decision? Some groups may not know they have been overlooked or have the resources to address any issues (Lokshin and Umapathi, 2022[64]; Barca and Chirchir, 2019[65]). Complaint processes should account for this with public agencies ensuring that marginalised and excluded populations are supported in making any application for a review of a decision.

In addition to appropriate safeguards to ensure accountability and fairness, the OECD AI Principles also emphasise:

- The need to ensure safety and security by implementing robust systems that can withstand adverse conditions such as digital security risks, and by ensuring that AI systems can be replaced or decommissioned in the event that they cause undue harm or exhibit undesired behaviour; and
- The need to build capacity in labour markets to interact with AI systems and adapt to AI-generated changes in labour markets.

The need for training and skills development goes hand in hand with the need to ensure transparency and explainability. Staff who engage with social security applicants need to be able to explain how a decision was reached and provide information about how that decision can be reviewed. This requires staff to be adequately trained and for there to be sufficient complaint processes in place. Furthermore, public agencies should consider developing algorithms in-house using internal experts and/or understand and be able to explain algorithms developed by external partners (OECD, 2019[66]), and it may also be necessary for lawyers, judges or other arbitrators to receive training on the functioning and fallibility of algorithms to be able to respond appropriately to any claims brought related to adverse outcomes (Citron, 2007[67]; Gilman, 2020[68]).

With effective policies and governance in place, AI could help to ensure that social protection systems are well-prepared to meet the challenges of today and tomorrow. Economic and sociodemographic megatrends will expand and change the need for support in the years to come, as climate change, digitalisation and population ageing will increase the need for support while constraining the resources available to provide it (Frey and Hyee, 2024[69]). By helping to improve the targeting, delivery and efficiency of services, AI has the potential to get better and more timely support to people when they need it most.

Conclusions

The use of AI in social protection systems holds potential for improving service delivery, closing coverage gaps, and freeing up time for civil servants – but this technology also presents significant risks, including in data privacy and automated decision-making. OECD governments are moving carefully in this space, given the high stakes, and current applications of AI have focused on client support, automating administrative processes, and detecting fraud.

There are promising potential use cases. These include using predictive analytics to forecast demand and social shocks and identify vulnerable clients earlier; reducing non-take-up through targeted outreach; and tailoring interventions to meet specific individual needs and reach programme goals. AI may also be able to help address concerns about bias or inconsistency in (human) decision-making processes. However, realising these benefits will require both investment in digital infrastructure and careful evaluation of when and how AI adds value compared to other tools.

To move forward responsibly, governments should continue to pilot AI applications and identify where AI is the most appropriate tool, as significant gains can still be achieved through other methods (e.g. facilitating enrolment through linked administrative data). Governments should establish a clear rationale for AI use in each context, grounded in policy goals. Governments must also engage the public early and transparently. Public trust remains fragile – only 40% of respondents across 27 OECD countries believe AI-supported application processing is beneficial for users. Meaningful engagement, data security, transparency, and a focus on fairness will help governments foster confidence in AI-supported social protection systems going forward.

References

Adams, D. (2024), *Managing the challenges of leveraging technology and data advances to improve social protection*, OECD Publishing, https://doi.org/10.1787/af31746d-en. [8]

American Academy of Pediatrics (2025), *Monitoring Software for Suicide Risk in Schools*, https://www.aap.org/en/patient-care/media-and-children/center-of-excellence-on-social-media-and-youth-mental-health/qa-portal/qa-portal-library/qa-portal-library-questions/monitoring-software-for-suicide-risk-in-schools/?srsltid=AfmBOoo4l2IMc19OlP9npFh-Q7AVUI_3zCvaBdAxfNBmB_Pb9eIJ2d19 (accessed on 19 June 2025). [39]

Amrit, C. et al. (2017), "Identifying child abuse through text mining and machine learning", *Expert Systems with Applications*, Vol. 88, pp. 402-418, https://doi.org/10.1016/j.eswa.2017.06.035. [36]

Assouline, M., S. Gilad and P. Ben-Nun Bloom (2021), "Discrimination of Minority Welfare Claimants in the Real World: The Effect of Implicit Prejudice", *Journal of Public Administration Research and Theory*, Vol. 32/1, pp. 75-96, https://doi.org/10.1093/jopart/muab016. [53]

Barca, V. and R. Chirchir (2019), *Building an integrated and digital social protection information system*, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), https://www.giz.de/en/downloads/giz2019-en-integrated-digital-social-protection-information-system.pdf. [65]

Barry, E. (2024), *Spying on student devices, schools aim to intercept self-harm before it happens*, https://www.nytimes.com/2024/12/09/health/suicide-monitoring-software-schools.html. [38]

Bell, E. and S. Jilke (2024), "Racial Discrimination and Administrative Burden in Access to Public Services", *Scientific Reports*, Vol. 14/1, https://doi.org/10.1038/s41598-023-50936-1. [55]

Błażej Kuźniacki (2023), *The Dutch childcare benefit scandal shows that we need explainable AI rules*, University of Amsterdam, https://www.uva.nl/en/shared-content/faculteiten/en/faculteit-der-rechtsgeleerdheid/news/2023/02/childcare-benefit-scandal-transparency.html?cb (accessed on November 2023). [62]

Booth, R. (2024), "Revealed: bias found in AI system used to detect UK benefits fraud", *The Guardian*, https://www.theguardian.com/society/2024/dec/06/revealed-bias-found-in-ai-system-used-to-detect-uk-benefits (accessed on 13 December 2024). [51]

Brioscú, A. et al. (2024), "A New Dawn for Public Employment Services: Service Delivery in the Age of Artificial Intelligence", OECD, Paris, https://doi.org/10.1787/5dc3eb8e-en. [44]

California Policy Lab (2024), *Predicting and preventing homelessness*, https://capolicylab.org/topics/homelessness/. [26]

California Policy Lab (2020), *Fact Sheet: What Research Says About Homelessness Prevention Programs*, https://capolicylab.org/topics/homelessness/ (accessed on 8 January 2025). [27]

Camden Care Choices (2024), *Care and support rights, Adult Social Care use of BEAM/Magic Notes*, https://camdencarechoices.camden.gov.uk/care-and-support-rights/adult-social-care-use-of-beammagic-notes/ (accessed on 19 June 2025). [46]

Carter, E. and A. Whitworth (2014), “Creaming and Parking in Quasi-Marketised Welfare-to-Work Schemes: Designed Out Of or Designed In to the UK Work Programme?”, *Journal of Social Policy*, Vol. 44/2, pp. 277-296, https://doi.org/10.1017/s0047279414000841. [56]

CeADAR (n.d.), *CAMEO – Online Platform of Earth Observation Data*, https://ceadar.ie/blog/cameo-online-platform-of-earth-observation-data/ (accessed on 1 April 2025). [24]

Centre for Economic Performance (2020), *In brief... Tackling domestic violence using machine learning*, https://cep.lse.ac.uk/pubs/download/cp579.pdf (accessed on 13 December 2024). [32]

Citron, D. (2007), “Technological due process”, *Washington University Law Review*, Vol. 85, pp. 1249-1313, https://ssrn.com/abstract=1012360. [67]

CompetentNL (2025), *CompetentNL website*, https://competentnl.nl/. [45]

Desiere, S., K. Langenbucher and L. Struyven (2019), “Statistical profiling in public employment services: An international comparison”*, OECD Social, Employment and Migration Working Papers*, No. 224, OECD Publishing, Paris, https://doi.org/10.1787/b5e5f16e-en. [43]

Emeriau, M. (2022), “Learning to be Unbiased: Evidence from the French Asylum Office”, *American Journal of Political Science*, Vol. 67/4, pp. 1117-1133, https://doi.org/10.1111/ajps.12720. [54]

Ernst & Young (2022), *How can data stop homelessness before it starts?*, Ernst & Young, https://www.ey.com/en_uk/insights/government-public-sector/how-can-data-stop-homelessness-before-it-starts (accessed on 13 December 2024). [30]

Executive Office of the President (2020), *Executive Order 13960: Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government*, https://www.federalregister.gov/d/2020-27065. [6]

Frey, V. and R. Hyee (2024), *Non-take-up and the digital transformation of social programmes in OECD countries*, OECD Publishing, https://doi.org/10.1787/af31746d-en. [25]

Frey, V. and R. Hyee (2024), *Overview: Megatrends and the future of social protection*, OECD, https://doi.org/10.1787/6c9202e8-en. [69]

Frey, V., R. Hyee and P. Minondo Canto (2024), *National frameworks to identify potential beneficiaries and integrate them into social protection*, OECD Publishing, https://doi.org/10.1787/af31746d-en. [42]

Gillingham, P. (2019), “Can Predictive Algorithms Assist Decision-Making in Social Work with Children and Families?”, *Child Abuse Review*, Vol. 28/2, pp. 114-126, https://doi.org/10.1002/car.2547. [34]

Gilman, M. (2020), *Poverty Algorithms: A poverty lawyer’s guide to fighting automated decision-making harms on low-income communities*, Data & Society, https://datasociety.net/library/poverty-lawgorithms/. [68]

GIZ (2024), *Preventing Accidents with AI*. [22]

Government of Scotland (2024), *Increasing AI transparency*, https://www.gov.scot/news/increasing-ai-transparency/. [13]

Government of the Netherlands (2022), *Letter to Parliament on Fraud Detection Facility and the issue of institutional racism*, https://www.rijksoverheid.nl/documenten/kamerstukken/2022/05/30/kamerbrief-reactie-op-verzoeken-over-fraudesignaleringsvoorziening (accessed on 8 January 2025). [61]

Grogger, J., R. Ivandic and T. Kirchmaier (2020), *Comparing Conventional and Machine-Learning Approaches to Risk Assessment in Domestic Abuse Cases: CEP Discussion Paper No 1676*, Centre for Economic Performance, https://cep.lse.ac.uk/pubs/download/dp1676.pdf (accessed on 20 December 2024). [31]

Hall, S. et al. (2024), "A Systematic Review of Sophisticated Predictive and Prescriptive Analytics in Child Welfare: Accuracy, Equity, and Bias", *Child and Adolescent Social Work Journal*, Vol. 41/6, pp. 831-847, https://doi.org/10.1007/s10560-023-00931-2. [35]

HHS, U. (2025), *Department of Health and Human Services: Artificial Intelligence Use Cases Inventory*, https://www.hhs.gov/programs/topic-sites/ai/use-cases/index.html. [5]

Hodgkinson, I., N. Jennings and T. Jackson (2024), *Everyone must understand the environmental costs of AI*, https://oecd.ai/en/wonk/understand-environmental-costs (accessed on 20 January 2025). [9]

Hyee, R. et al. (2024), "How Reliable Are Social Safety Nets in Situations of Acute Economic Need? Extended Estimates for 14 OECD countries", *IZA Discussion paper No 17477*, https://docs.iza.org/dp17477.pdf. [52]

Isbanner, S. et al. (2022), "The Adoption of Artificial Intelligence in Health Care and Social Services in Australia: Findings From a Methodologically Innovative National Survey of Values and Attitudes (the AVA-AI Study)", *Journal of Medical Internet Research*, Vol. 24/8, p. e37611, https://doi.org/10.2196/37611. [11]

ISSA (2022), *Artificial intelligence in social security institutions: The case of intelligent chatbots*. [16]

ISSA (2022), *Data-driven innovation in social security: Good practices from Asia and the Pacific*. [21]

ISSA (2020), *Artificial Intelligence in Social Security: Background and Experiences*. [19]

Lancaster University (2021), *Developing a new AI early warning system for flooding*. [23]

Lara-Montero, A. (2024), *The transformation potential of AI on social services*, European Public Mosaic, https://eapc.gencat.cat/ca/publicacions/epum. [41]

Local Government Association (2025), *Peer Challenge Progress Report: London Borough of Ealing*, https://www.ealing.gov.uk/download/downloads/id/20338/review_of_ealings_adult_social_care_-_january_2025.pdf (accessed on 19 June 2025). [47]

Lokshin, M. and N. Umapathi (2022), *AI for social protection: Mind the people*, Brookings, https://www.brookings.edu/articles/ai-for-social-protection-mind-the-people/. [64]

Maidstone Borough Council (2022), *MBC Housing Team wins national award for preventing homelessness*, https://news.maidstone.gov.uk/home/news/2022/november/mbc-housing-team-wins-national-award-for-preventing-homelessness (accessed on 13 December 2024). [29]

Marr, B. (2018), *How The UK Government Uses Artificial Intelligence To Identify Welfare And State Benefits Fraud*, Forbes. [20]

NHS England (2025), *Nationwide roll out of artificial intelligence tool that predicts falls and viruses*, https://www.england.nhs.uk/2025/03/nationwide-roll-out-of-artificial-intelligence-tool-that-predicts-falls-and-viruses/ (accessed on 1 April 2025). [33]

Ntoutsi, E. et al. (2020), "Bias in data-driven artificial intelligence systems—An introductory survey", *WIREs Data Mining and Knowledge Discovery*, Vol. 10/3, https://doi.org/10.1002/widm.1356. [50]

OECD (2025), *More Effective Social Protection for Stronger Economic Growth: Main Findings from the 2024 OECD Risks that Matter Survey*, OECD Publishing, Paris, https://doi.org/10.1787/3947946a-en. [10]

OECD (2025), *Saving time, improving support: How AI is helping social workers free up time for care*, OECD, https://www.oecd.org/en/about/projects/ai-and-the-future-of-social-protection.html. [48]

OECD (2024), *AI in Health: Huge potential, huge risks*, OECD, https://www.oecd.org/en/publications/2024/01/ai-in-health-huge-potential-huge-risks_ff823a24.html. [3]

OECD (2024), "Explanatory memorandum on the updated OECD definition of an AI system", *OECD Artificial Intelligence Papers*, No. 8, OECD Publishing, Paris, https://doi.org/10.1787/623da898-en. [7]

OECD (2024), "Explanatory memorandum on the updated OECD definition of an AI system", *OECD AI Papers*, Vol. 8, pp. 1-11, https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/03/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_3c815e51/623da898-en.pdf. [14]

OECD (2024), *Investing in Youth: North Denmark Region*, Investing in Youth, OECD Publishing, Paris, https://doi.org/10.1787/0856ce12-en. [40]

OECD (2024), *Megatrends and the Future of Social Protection*, OECD Publishing, Paris, https://doi.org/10.1787/6c9202e8-en. [2]

OECD (2024), *Modernising access to social protection: Strategies, technologies and data advances in OECD countries*, OECD, https://doi.org/10.1787/af31746d-en. [1]

OECD (2024), *OECD Survey on Drivers of Trust in Public Institutions – 2024 Results: Building Trust in a Complex Policy Environment*, OECD Publishing, Paris, https://doi.org/10.1787/9a20554b-en. [57]

OECD (2023), "AI language models: Technological, socio-economic and policy considerations", *OECD Digital Economy Papers*, No. 352, OECD Publishing, Paris, https://doi.org/10.1787/13d38f92-en. [15]

OECD (2023), "The state of implementation of the OECD AI Principles four years on", *OECD Artificial Intelligence Papers*, No. 3, OECD Publishing, Paris, https://doi.org/10.1787/835641c9-en. [4]

OECD (2022), “OECD Framework for the Classification of AI systems”, *OECD Digital Economy Papers*, No. 323, OECD Publishing, Paris, https://doi.org/10.1787/cb6d9eca-en. [63]

OECD (2019), *Harnessing new social data for effective social policy and service delivery*, OECD, Paris, https://www.oecd.org/els/soc/Workshop-NewSocialData-16Oct2019-BackgroundNote.pdf. [66]

OECD (2019), *Recommendation of the Council on Artificial Intelligence*, OECD, Paris, https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449. [59]

Ohlenburg, T. (2020), *AI in Social Protection - Exploring Opportunities and Mitigating Risks*, GIZ; ADB, Bonn, https://www.adb.org/publications/ai-social-protection-exploring-opportunities-mitigating-risks (accessed on 13 December 2024). [12]

Parikh, R., S. Teeple and A. Navathe (2019), “Addressing Bias in Artificial Intelligence in Health Care”, *JAMA*, Vol. 322/24, p. 2377, https://doi.org/10.1001/jama.2019.18058. [49]

Public Sector Tech Watch (2025), *Public Sector Tech Watch website*, https://interoperable-europe.ec.europa.eu/collection/public-sector-tech-watch. [17]

Robila, M. and S. Robila (2020), “Applications of Artificial Intelligence Methodologies to Behavioral and Social Sciences”, *Journal of Child and Family Studies*, Vol. 29/10, pp. 2954-2966, https://doi.org/10.1007/s10826-019-01689-x. [37]

Šuica, D. et al. (2024), *Social Rights: Key Challenges*, https://eapc.gencat.cat/ca/detalls/Noticia/EPum_23_noticia#googtrans(ca|en) (accessed on 13 December 2024). [28]

The European Parliament (2022), *The Dutch childcare benefit scandal, institutional racism and algorithms*, https://www.europarl.europa.eu/doceo/document/O-9-2022-000028_EN.html (accessed on November 2023). [60]

U.S. Administration for Children and Families (2024), *Child welfare information gateway*, https://www.childwelfare.gov/stay-connected/contact/. [18]

Verhagen, A. (2024), “Using AI to manage minimum income benefits and unemployment assistance: Opportunities, risks and possible policy directions”, *OECD Artificial Intelligence Papers*, No. 21, OECD Publishing, Paris, https://doi.org/10.1787/718c93a1-en. [58]

Notes

[1] Section 5(e) clarifies: “Within 120 days of the completion of their inventories, agencies shall make their inventories available to the public, to the extent practicable and in accordance with applicable law and policy, including those concerning the protection of privacy and of sensitive law enforcement, national security, and other protected information” (Executive Office of the President, 2020[6]).

[2] The OECD has recently clarified the definition of an AI system contained in the 2019 OECD Recommendation on AI (the “AI Principles”) to support their continued relevance. The following updated definition was adopted by the OECD Council on 8 November 2023: “An AI system is a machine-based system that can, for a given set of human-defined explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as makes predictions, content, recommendations, or decisions that can influence physical real or virtual environments. Different AI systems are designed to operate with varying vary in their levels of autonomy and adaptiveness after deployment.”

[3] Machine learning (ML) is a set of techniques to allow machines to learn in an automated manner through patterns and inferences rather than through explicit instructions from a human. ML approaches often teach machines to reach an outcome by showing them many examples of correct outcomes. ML contains numerous techniques that have been used by economists, researchers and technologists for decades. These range from linear and logistic regressions, decision trees and principal component analysis to deep neural networks (OECD, 2022[63]).

[4] Warning signs for homelessness amongst those living in deep poverty included “sharp spikes in service use, increasingly frequent service use, and the receipt of multiple services from a single agency.”

[5] 13 country average. Only countries with sufficiently large numbers of respondents who identified as a minority based on ethnicity or skin colour were included in the average (cell sizes larger than 10 observations for numerators, denominators and residuals).