

# L'IA ET L'AVENIR DE LA PROTECTION SOCIALE DANS LES PAYS DE L'OCDE

---

DOCUMENTS DE L'OCDE SUR  
L'INTELLIGENCE ARTIFICIELLE

Juin 2025 N° 42

Documents de l'OCDE sur l'intelligence artificielle

# **L'IA et l'avenir de la protection sociale dans les pays de l'OCDE**



## Avertissements

Le présent document est publié sous la responsabilité du Secrétaire général de l'OCDE. Les opinions exprimées et les arguments avancés dans ce document ne reflètent pas nécessairement les vues officielles des pays membres de l'OCDE.

Le présent document, ainsi que les données et cartes qu'il contient, n'affectent en rien le statut ou la souveraineté des territoires, la délimitation des frontières internationales et des limites territoriales, ni le nom des territoires, villes ou régions.

**Crédits photos** : Couverture© Kjpargeter/Shutterstock.com

Les errata des publications de l'OCDE sont disponibles à l'adresse suivante : [www.oecd.org/en/publications/support/corrigenda.html](http://www.oecd.org/en/publications/support/corrigenda.html) .

© OCDE 2025



Attribution 4.0 International (CC BY 4.0)

Cette œuvre est mise à disposition selon les termes de la licence Creative Commons Attribution 4.0 International. En utilisant cette œuvre, vous acceptez d'être lié par les termes de cette licence

(<https://creativecommons.org/licenses/by/4.0/>).

**Attribution** : vous devez citer l'œuvre.

**Traductions** – vous devez citer l'œuvre originale, identifier les modifications apportées à l'original et ajouter le texte suivant : *En cas de divergence entre l'œuvre originale et la traduction, seul le texte de l'œuvre originale doit être considéré comme valide.*

**Adaptations** – vous devez citer l'œuvre originale et ajouter le texte suivant : *Il s'agit d'une adaptation d'une œuvre originale de l'OCDE. Les opinions exprimées et les arguments utilisés dans cette adaptation ne doivent pas être présentés comme représentant les vues officielles de l'OCDE ou de ses pays membres.*

**Contenu tiers** – la licence ne s'applique pas au contenu tiers présent dans l'œuvre. Si vous utilisez un tel contenu, vous êtes tenu d'obtenir l'autorisation du tiers et d'assumer toute responsabilité en cas de violation des droits d'auteur.

Vous ne devez pas utiliser le logo, l'identité visuelle ou l'image de couverture de l'OCDE sans autorisation expresse, ni suggérer que l'OCDE approuve votre utilisation de l'œuvre.

Tout litige découlant de la présente licence sera réglé par arbitrage conformément au Règlement d'arbitrage 2012 de la Cour permanente d'arbitrage (CPA). Le siège de l'arbitrage sera Paris (France). Le nombre d'arbitres sera de un.

# L'IA et l'avenir de la protection sociale dans les pays de l'OCDE

---

Les gouvernements des pays de l'OCDE recourent de plus en plus aux technologies de pointe et aux données pour améliorer la couverture, l'efficacité et l'efficience des programmes sociaux, mais ils font preuve de prudence lorsqu'il s'agit d'introduire l'intelligence artificielle (IA). L'IA est couramment utilisée dans le domaine de la protection sociale pour l'assistance à la clientèle, l'automatisation des processus administratifs et la détection des fraudes. À l'avenir, l'IA pourrait contribuer de manière significative à améliorer les performances des programmes sociaux, notamment grâce à l'analyse prédictive, à un meilleur rayonnement et à des interventions mieux adaptées, mais les gouvernements doivent continuer à renforcer la confiance et à favoriser la transparence lorsqu'ils utilisent l'IA.

---

## Remerciements

Cet article a été rédigé par Valerie Frey, Doron Wijkers et Raphaela Hyye, de la Direction de l'emploi, du travail et des affaires sociales de l'OCDE. Les auteurs remercient les délégués du Groupe de travail de l'OCDE sur la politique sociale, ainsi que Dorothy Adams, Marc Folch, Pauline Fron, Marguerita Lane, Annikka Lemmens, Mark Pearson, Veerle Miranda, Stefano Scarpetta, Jasmin Thomas, Monika Queisser et les membres du Comité d'examen de l'IA de l'OCDE pour leurs contributions, leurs commentaires et leurs suggestions. Les discussions avec les décideurs politiques et les responsables de programmes lors de [la série de webinaires de l'OCDE sur l'IA et l'avenir de la protection sociale](#) ont alimenté le présent document et sont très appréciées. Natalie Corry, Marie-Aurélien Elkurd et Hanna Varkki ont apporté un soutien logistique et éditorial précieux.

# Table des matières

Remerciements	4
Résumé	6
Les pays exploitent les données et la technologie pour moderniser la protection sociale	7
L'accompagnement des clients et l'automatisation au cœur des préoccupations	11
Analyse prédictive, aide à la décision et interventions sur mesure : l'avenir de l'IA dans la protection sociale ?	13
Mise en place de garanties pour une utilisation sûre et fiable de l'IA	18
Conclusions	21
Références	22
Notes	27

## FIGURES

Figure 1. Utilisations actuelles et potentielles de l'intelligence artificielle dans la protection sociale	8
Figure 2. De nombreuses personnes ne font pas encore confiance à l'utilisation de l'IA dans les services publics et sociaux	9

## Résumé

Les gouvernements des pays de l'OCDE ont de plus en plus recours aux technologies de pointe et aux données pour améliorer la couverture, l'efficacité et l'efficience des programmes sociaux. Dans le cadre de cette transformation technologique plus large, les gouvernements font preuve de prudence dans l'application de l'intelligence artificielle (IA) à la protection sociale, compte tenu des risques et des défis importants que cela comporte, notamment les risques potentiels pour la confidentialité des données et les erreurs dans la prise de décision automatisée.

À l'heure actuelle, l'utilisation de l'IA dans les programmes sociaux concerne principalement l'aide à la clientèle, l'automatisation des processus administratifs et la détection des fraudes. Pourtant, l'IA offre un potentiel considérable pour améliorer les performances des programmes sociaux. S'appuyant sur les conclusions du rapport *Moderniser l'accès à la protection sociale* (OCDE, 2024<sup>[1]</sup>), le présent document recense les cas d'utilisation actuels et examine les nouvelles utilisations potentielles de l'IA dans le domaine de la protection sociale, notamment l'analyse prédictive pour anticiper la demande et les chocs ; l'analyse prédictive pour améliorer l'identification des clients et l'intervention précoce ; l'amélioration de la sensibilisation et la réduction du non-recours ; des interventions mieux adaptées pour répondre aux besoins des clients et atteindre les objectifs des programmes ; et la lutte contre la discrimination humaine (réelle ou perçue) dans les processus de demande.

Les gouvernements font preuve de prudence en identifiant les défis politiques auxquels l'IA peut apporter une réponse significative (par rapport à d'autres outils), en mettant en place les infrastructures nécessaires et en testant des projets à petite échelle afin de vérifier si les processus peuvent être déployés en toute sécurité à plus grande échelle. Le soutien du public à l'utilisation de l'IA par les pouvoirs publics est mitigé : selon l'enquête « Risks that Matter » de l'OCDE, dans 27 pays, seuls 40 % des personnes interrogées estiment que l'utilisation de l'IA par les pouvoirs publics pour traiter et approuver les demandes de programmes sociaux est bénéfique pour les utilisateurs. Il est donc nécessaire de renforcer la participation du public afin de favoriser la transparence et de renforcer la confiance dans l'utilisation de l'IA pour la protection sociale.

## Les pays exploitent les données et les technologies pour moderniser la protection sociale

Les gouvernements de l'OCDE s'appuient de plus en plus sur les technologies de pointe et les données pour améliorer l'expérience des utilisateurs des services publics, notamment en matière de demande et de prestation de programmes sociaux, et pour veiller à ce que les citoyens connaissent les prestations et les services auxquels ils ont droit et puissent y accéder.

Les technologies numériques modifient la nature des interactions bureaucratiques entre l'État et les citoyens. Davantage de services sont désormais disponibles en ligne, ce qui permet aux administrations de concentrer leurs ressources (humaines) sur les personnes dont les besoins ne sont pas adaptés aux systèmes automatisés, telles que les utilisateurs de services ayant des besoins complexes. Les pays de l'OCDE relient également de plus en plus les données administratives provenant de différentes sources afin de réduire la charge administrative qui pèse sur les demandeurs, de rendre l'information plus facilement accessible et de mesurer le non-recours (OCDE, 2024<sup>[1]</sup>). Les outils numériques et le renforcement des capacités en matière de données contribuent à améliorer l'inscription aux programmes et la prestation des services. Ils offrent un potentiel considérable pour améliorer l'efficacité opérationnelle et réduire les coûts, un objectif particulièrement important compte tenu des contraintes budgétaires actuelles (OCDE, 2024<sup>[2]</sup>).

Toutefois, d'une manière générale, les utilisations avancées des technologies et des données – y compris le recoupement des données – sont moins courantes dans le secteur public des pays de l'OCDE que dans le secteur privé, et moins courantes dans le domaine de la protection sociale que dans le secteur de la santé, par exemple (OCDE, 2024<sup>[3]</sup>).

Compte tenu de l'intérêt des gouvernements pour le potentiel de l'intelligence artificielle (IA) en matière d'amélioration des programmes sociaux, des risques et des défis liés à sa mise en œuvre, ainsi que du rythme rapide et de l'ampleur du développement des technologies d'IA, cette note présente un aperçu des cas d'utilisation actuels de l'IA dans les systèmes de protection sociale des pays de l'OCDE. Ce document développe les conclusions du rapport *Moderniser l'accès à la protection sociale* (OCDE, 2024<sup>[1]</sup>) et propose de nouvelles façons dont l'IA pourrait bénéficier à la protection sociale.

### ***L'IA recèle un énorme potentiel de transformation des programmes sociaux.***

L'IA est en train de transformer les sociétés et les économies, et elle ouvre de nouvelles perspectives pour les gouvernements. Ses partisans affirment qu'elle pourrait améliorer radicalement l'efficacité et la qualité des services publics dans des domaines tels que l'éducation, la santé et la protection sociale, apportant ainsi des avantages sociaux et économiques considérables. L'intérêt pour le potentiel et l'utilisation de l'IA dans les services publics s'est donc considérablement accru. En mai 2023, 51 pays dans le monde avaient communiqué à l'OCDE leur stratégie nationale en matière d'IA, ce qui représente une augmentation spectaculaire par rapport à 2017, où seuls quelques pays disposaient d'une telle stratégie (OCDE, 2023<sup>[4]</sup>).

Quelques pays exigent des ministères ou des agences gouvernementales qu'ils publient leurs cas d'utilisation de l'IA. Citons par exemple l'« Artificial Intelligence Use Case Inventory » (inventaire des cas d'utilisation de l'IA) du ministère américain de la Santé et des Services sociaux, qui a publié environ 270 cas d'utilisation de l'IA en 2024 (HHS, 2025<sup>[5]</sup>), conformément à un décret présidentiel (Executive Office of the President, 2020<sup>[6]</sup>). L'observatoire « Public Sector Tech Watch », hébergé par Interoperable Europe, avait recensé 58 cas d'utilisation de l'IA par les gouvernements nationaux de l'Union européenne dans le domaine de la protection sociale (ainsi que 40 cas au niveau local et 17 au niveau régional) en mai 2025.

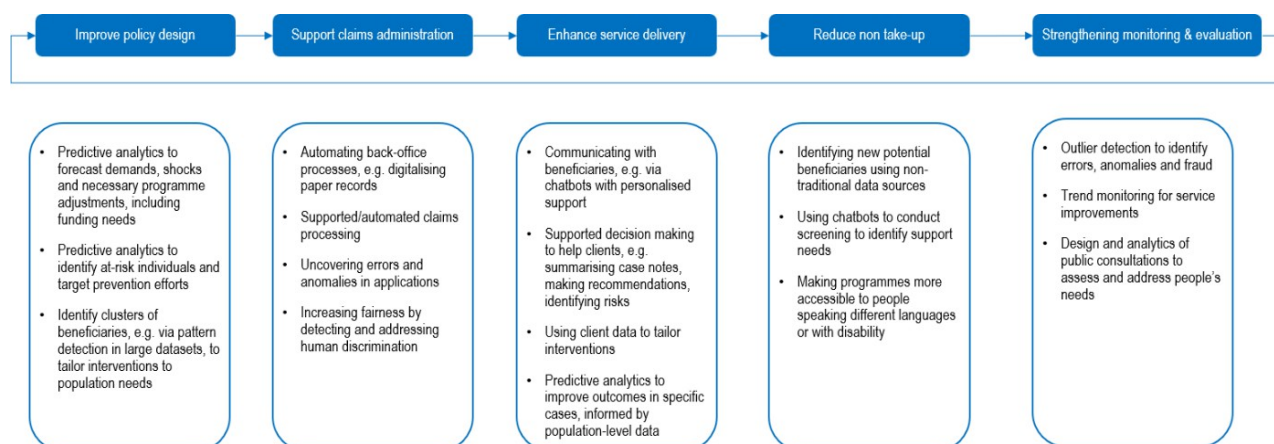
**Les applications potentielles de l'IA dans les systèmes de protection sociale sont importantes, l'IA pouvant améliorer la conception des politiques, faciliter la gestion des demandes, améliorer la prestation des services, prévoir la demande de programmes, réduire le non-recours et renforcer le suivi et l'évaluation** (par exemple pour réduire la fraude). Ces utilisations potentielles sont présentées dans la figure 1. L'IA pourrait être utilisée pour améliorer l'accès à la protection sociale et sa prestation, par exemple en ciblant plus précisément les programmes aux bénéficiaires éligibles, en fournissant des informations ciblées sur les programmes et les prestations disponibles, en prenant des décisions plus rapides et plus précises en matière d'éligibilité, en ajustant les prestations en temps réel, en surveillant et en gérant la prestation des prestations, en identifiant les populations à risque afin de cibler les efforts de prévention, et en automatisant ou en simplifiant les tâches, ce qui permettrait de libérer du temps pour les travailleurs sociaux surchargés afin qu'ils puissent le consacrer à leurs clients.

Bien entendu, les modèles d'IA ont de nombreuses applications (voir encadré 1) et, compte tenu du rythme des progrès technologiques, d'autres applications verront très certainement le jour dans les années à venir (OCDE, 2024<sup>[7]</sup>).



Figure 1. Utilisations actuelles et potentielles de l'intelligence artificielle dans la protection sociale

Schéma des cas d'utilisation actuels et probables de l'IA et de la protection sociale



Remarque : aperçu de l'OCDE des utilisations actuelles et hypothétiques de l'IA pour la protection sociale.

Source : illustration des auteurs.

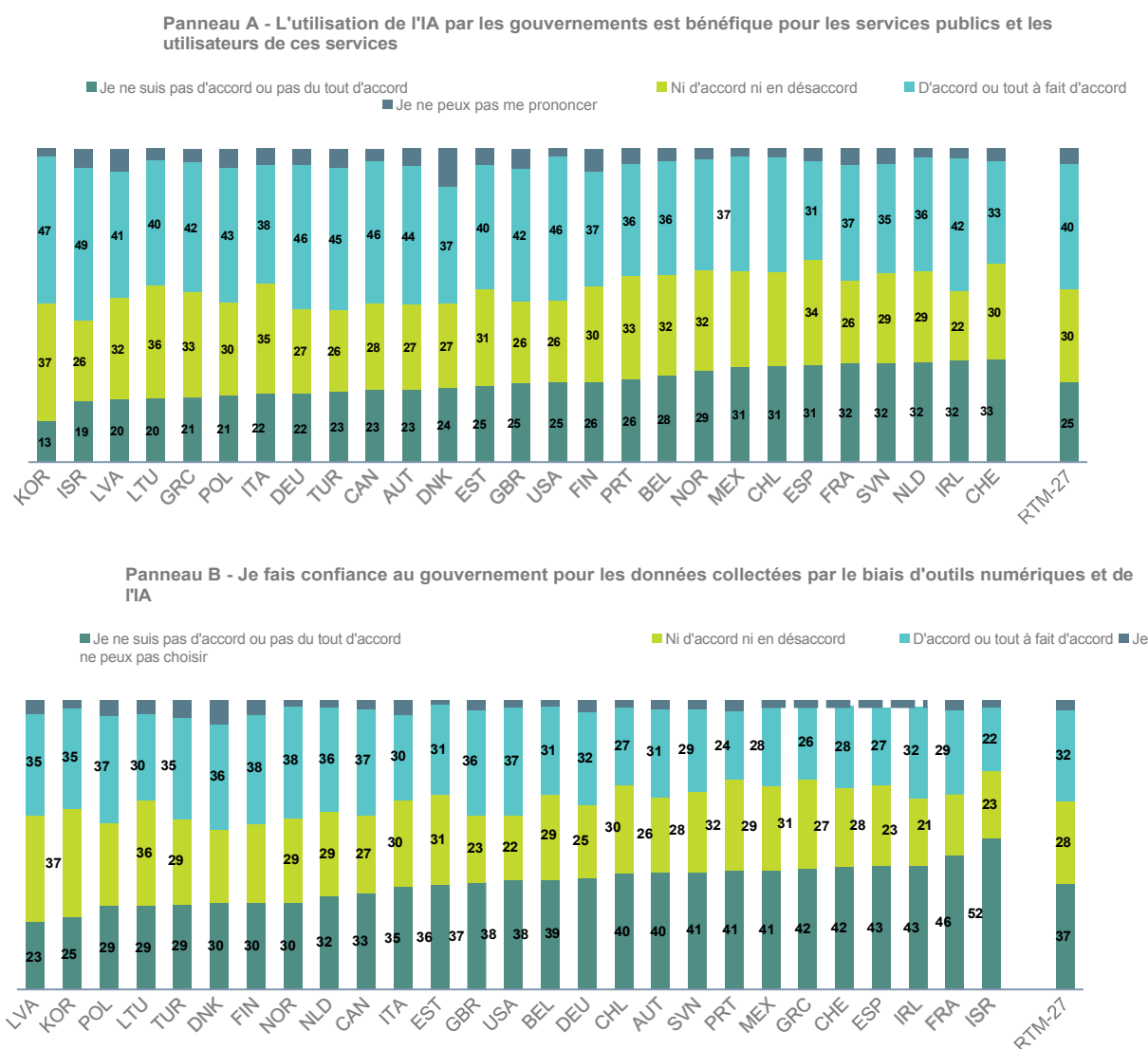
Cependant, l'utilisation de l'IA dans les services publics **continue** de présenter **des risques et des défis importants**. L'utilisation de grands ensembles de données contenant des informations sensibles soulève d'importantes questions en matière de protection de la vie privée. L'automatisation des processus pourrait déshumaniser les services ou réduire la transparence, en particulier si l'intervention ou la supervision humaine est insuffisante, ou si les systèmes d'IA utilisés pour prendre des décisions ne sont pas bien compris ou expliqués. L'utilisation de l'IA soulève des questions quant à la responsabilité en cas de problème, et les algorithmes entraînés à partir de jeux de données non représentatifs ou biaisés risquent de (re)produire des résultats biaisés et discriminatoires. Par exemple, l'utilisation de l'IA pour gérer les prestations sociales comporte le risque que l'évaluation de l'éligibilité à ces prestations soit incorrecte ou systématiquement biaisée à l'égard de certains groupes démographiques (Adams, 2024<sup>[8]</sup>). L'importance de la puissance de calcul nécessaire pour faire fonctionner les systèmes d'IA soulève également des préoccupations environnementales considérables (Hodgkinson, Jennings et Jackson, 2024<sup>[9]</sup>).

Peut-être en raison de ces défis, **une grande partie du public ne fait pas encore confiance à l'utilisation de l'IA dans les services sociaux**. Les résultats de l'enquête « Risques qui comptent » (RTM) menée par l'OCDE dans 27 pays mettent en évidence le scepticisme quant aux avantages de l'utilisation de l'IA par les gouvernements pour les utilisateurs de services, ainsi qu'un manque de confiance dans la manière dont les gouvernements utiliseront les données collectées à leur sujet grâce aux outils numériques et à l'IA. Dans un échantillon représentatif de 27 pays, seuls 40 % des personnes interrogées estiment que l'utilisation de l'IA pour aider à traiter et approuver les demandes de programmes sociaux est bénéfique pour les utilisateurs (figure 2, panneau A). En moyenne, 30 % expriment leur incertitude et 25 % ne pensent pas que l'utilisation de l'IA par les gouvernements soit bénéfique pour les utilisateurs des programmes sociaux (OCDE, 2025<sup>[10]</sup>).

Lorsqu'on leur demande s'ils font confiance au gouvernement pour traiter les données collectées à leur sujet à l'aide d'outils numériques et de l'IA, 37 % des personnes interrogées se disent en désaccord ou fortement en désaccord, tandis que 32 % indiquent qu'elles font confiance au gouvernement pour traiter ces données (figure 2, panneau B). Moins de la moitié (44 %) des personnes interrogées sont convaincues que l'IA ne sera utilisée pour évaluer les demandes d'aides publiques que lorsque cela sera sûr et fiable (OCDE, 2025<sup>[10]</sup>). Ces résultats suggèrent que les gouvernements ont des progrès à faire en matière de communication et de renforcement de la confiance dans l'utilisation de l'IA dans les politiques sociales.

**Figure 2. De nombreuses personnes ne font pas encore confiance à l'utilisation de l'IA dans les services publics et sociaux**

Répartition des réponses aux affirmations « L'utilisation de l'IA par les gouvernements pour aider à traiter et approuver les demandes est une bonne chose pour les utilisateurs des prestations/services publics » (**panneau A**) et « Je fais confiance au gouvernement pour les données qu'il collecte à mon sujet à l'aide d'outils numériques et de l'IA » (**panneau B**), en réponse à la question « En ce qui concerne l'utilisation des outils numériques et de l'intelligence artificielle (IA) par le gouvernement, dans quelle mesure êtes-vous d'accord ou en désaccord avec les affirmations suivantes ? », auprès de personnes âgées de 18 à 64 ans, 2024



Remarque : les personnes interrogées ont été invitées à répondre à la question suivante : « En ce qui concerne l'utilisation des outils numériques et de l'intelligence artificielle (IA) par les pouvoirs publics, dans quelle mesure êtes-vous d'accord ou en désaccord avec les affirmations suivantes ? » Elles ont ensuite reçu une liste d'affirmations relatives aux pouvoirs publics, aux outils numériques et à l'intelligence artificielle, notamment : « L'utilisation de l'IA par les pouvoirs publics pour traiter et approuver les demandes est bénéfique pour les utilisateurs des services publics » et « Je fais confiance aux pouvoirs publics pour traiter les données qu'ils collectent à mon sujet à l'aide d'outils numériques et de l'IA ». Les réponses possibles étaient les suivantes : 1. Pas du tout d'accord, 2. Pas d'accord, 3. Ni d'accord ni pas d'accord, 4. D'accord, 5. Tout à fait d'accord, 6. Je ne peux pas me prononcer. RTM-27 fait référence à la moyenne non pondérée des 27 pays de l'OCDE participant à l'enquête « Risques importants » de l'OCDE 2024.

Source : Enquête de l'OCDE sur les risques importants 2024 (<https://oe.cd/rtm>).

Ces résultats reflètent ceux d'enquêtes nationales. Une enquête réalisée en Australie en 2022 a révélé que si 60 % des Australiens soutiennent le développement de l'IA en général, cette proportion diminue considérablement lorsqu'on les interroge sur des scénarios spécifiques liés aux services sociaux. Seuls 31 à 39 % des personnes interrogées soutiennent le développement de l'IA dans différents services sociaux

, une solide majorité d'entre eux (quatre sur cinq) préférant le contact humain et la discrétion dans les domaines de la santé et des services sociaux aux avantages que l'IA pourrait apporter en termes de rapidité, de précision et de commodité (Isbanner et al., 2022<sup>[11]</sup>). Les personnes interrogées étaient particulièrement sensibles à la précision et à l'équité de l'IA.

Les préoccupations relatives à l'utilisation de l'IA ont également occupé une place importante dans l'élaboration des politiques. Les raisons de la lenteur relative du déploiement dans le secteur public comprennent des préoccupations éthiques et juridiques, ainsi que le scepticisme quant à l'adéquation des systèmes informatiques dans le domaine des politiques publiques et de l'administration (Ohlenburg, 2020<sup>[12]</sup>). L'exploitation de l'IA nécessite bien sûr des investissements complémentaires dans les données, la formation et la numérisation des processus de travail, ainsi que des changements dans les processus organisationnels.

**Compte tenu des risques et des défis à relever, les pays mettent en place les bases et les capacités nécessaires pour promouvoir en toute sécurité l'utilisation des applications d'IA.** Citons à titre d'exemple le programme « AI Lighthouse » du Royaume-Uni et le décret américain n° 139 060 de 2020 intitulé « Promoting the use of trustworthy AI in the federal government » (Promouvoir l'utilisation d'une IA fiable au sein du gouvernement fédéral). Afin d'accroître la transparence, le décret américain exige que les cas d'utilisation de l'IA par le gouvernement fédéral soient rendus publics (<sup>1</sup>(par exemple, HHS, 2025<sup>[5]</sup>). De même, la « stratégie en matière d'intelligence artificielle » de l'Écosse exigera que « chaque projet utilisant l'IA [...] soit enregistré dans le registre écossais de l'IA, une base de données accessible au public qui fournit une série d'informations sur l'utilisation de cette technologie dans les projets développés par les organismes publics » (gouvernement écossais, 2024<sup>[13]</sup>).

Les pays réfléchissent attentivement à la meilleure façon de tirer parti des nouvelles technologies et procèdent avec prudence, en mettant en œuvre et en surveillant des projets à petite échelle avant de décider de les étendre. De nombreuses utilisations des technologies de pointe dans les services sociaux, et en particulier de l'IA, restent limitées à des essais ponctuels visant à déterminer la faisabilité, la fonctionnalité et la portée de leur déploiement. Actuellement, l'utilisation de l'IA dans la protection sociale se concentre sur l'assistance à la clientèle, l'automatisation des tâches et la détection des erreurs et des fraudes, les applications plus avancées – par exemple l'utilisation de l'analyse prédictive à des fins de prévention – restant encore assez limitées.

### Encadré 1. Qu'est-ce que l'IA ? Catégories, applications et utilisations de l'intelligence artificielle

Les définitions de l'IA sont souvent larges, et il peut être tentant de classer la plupart des exemples d'analyse dans la catégorie de l'IA. Cependant, toutes les méthodes analytiques ne font pas appel à des techniques d'IA. L'OCDE décrit un système d'IA comme un système automatisé qui, à des fins explicites ou implicites, déduit, à partir des données qui lui sont fournies, comment générer des résultats tels que des prévisions, des contenus, des recommandations ou des décisions susceptibles d'influencer des environnements physiques ou virtuels. L'OCDE a cherché à fournir une description compréhensible, techniquement précise, neutre sur le plan technologique et applicable à court et à long terme.

Les différents systèmes d'IA varient en termes de niveau d'autonomie et d'adaptabilité après leur déploiement. Ils peuvent utiliser des données provenant de machines et/ou d'êtres humains pour percevoir des environnements réels et/ou virtuels, abstraire ces perceptions en modèles (de manière automatisée, par exemple grâce à l'apprentissage automatique<sup>3</sup>, ou manuellement) et utiliser l'inférence de modèles pour générer du contenu et/ou formuler des options d'information ou d'action. Les systèmes d'IA sont conçus pour fonctionner avec différents niveaux d'autonomie.

Les systèmes d'IA sont généralement construits sur un ou plusieurs modèles qui déduisent des résultats à partir d'entrées spécifiées, notamment :

Les techniques d'apprentissage automatique (ML), qui utilisent des algorithmes pour identifier et apprendre des modèles à partir de données, apprennent et améliorent leurs performances au fil du temps. Ces modèles sont « entraînés » sur un ensemble de données afin de générer des résultats tels que des prévisions ou des décisions, sans instructions explicites d'un opérateur humain, et les commentaires peuvent aider le modèle à « apprendre » et à devenir plus précis. Par exemple, les techniques d'apprentissage automatique peuvent être utilisées pour prédire les ménages exposés au risque de sans-abrisme sur la base de données historiques et de modèles associés à ce risque, tels que les impayés de factures. Les techniques d'apprentissage automatique enseignent souvent aux machines à atteindre un résultat en leur montrant de nombreux exemples de résultats corrects.

Les systèmes d'IA symboliques ou basés sur la connaissance s'appuient sur des descriptions explicites des variables et de leurs relations pour tirer des conclusions logiques et/ou probabilistes à partir des données.

Les domaines d'application des modèles d'IA sont vastes et comprennent le traitement du langage naturel, les systèmes intelligents d'aide à la décision et la vision par ordinateur (OCDE, 2024<sup>[14]</sup>). Le traitement du langage naturel (TLN), par exemple, est un sous-ensemble de l'IA qui automatise les fonctions du langage naturel telles que l'analyse, la génération ou la réponse à des textes (OCDE, 2023<sup>[15]</sup>). Cela inclut les grands modèles linguistiques (LLM) tels que ChatGPT, qui utilisent des techniques d'apprentissage automatique – entraînées sur de vastes quantités de données – pour comprendre et répondre à des textes, générant ainsi des interactions conversationnelles entre les machines et les personnes. Les systèmes d'IA générative qui produisent des « contenus » tels que des textes ou des images sont devenus un type d'IA si important et si répandu que les contenus ont été inclus dans une catégorie distincte dans la définition révisée de l'IA par l'OCDE en 2023.

Source : Adapté de l'OCDE (2024<sup>[7]</sup>), Exposé explicatif sur la définition actualisée de l'OCDE d'un système d'IA, <https://doi.org/10.1787/623da898-en>; OCDE (2023<sup>[15]</sup>), Modèles linguistiques d'IA : considérations technologiques, socio-économiques et politiques, <https://doi.org/10.1787/13d38f92-en>.

## Le soutien à la clientèle et l'automatisation au cœur des préoccupations

### Chatbots et assistance numérique basés sur l'IA

L'utilisation de chatbots et d'assistants numériques pour améliorer les services à la clientèle en ligne, disponibles 24 heures sur 24, 7 jours sur 7 et dans différentes langues, est de plus en plus courante. Ces assistants intelligents imitent le comportement humain et répondent de manière autonome aux demandes des utilisateurs. Les assistants numériques en contact avec la clientèle apparaissent comme une utilisation relativement courante de l'IA dans les pays de l'OCDE (OCDE, 2024<sup>[11]</sup>) et dans le monde entier. Dans une enquête sur l'adoption de l'IA réalisée en 2021 par l'ISSA auprès de 166 agences gouvernementales à travers le monde, les chatbots sont arrivés en tête, 26 % des répondants les ayant déjà mis en œuvre et 59 % prévoyant de le faire dans les trois ans. Et dans une étude portant sur 230 services publics utilisant l'IA dans l'UE, les chatbots sont apparus comme l'application la plus courante, représentant plus d'un cinquième des cas d'utilisation (ISSA, 2022<sup>[16]</sup>).

Les chatbots sont utilisés au moins en **Autriche**, en **Australie**, en **Belgique**, en **Finlande**, en **France**, en **Allemagne**, en **Corée**, en **Norvège**, en **Espagne** et aux **États-Unis** pour soutenir la prestation de services sociaux au niveau national ou infranational. Dans la plupart des cas, les chatbots ont été déployés pour fournir des informations aux citoyens et répondre à leurs questions, tandis qu'une poignée de pays les utilisent à des fins de prévention, par exemple pour vérifier le bien-être des personnes. L'institution d'assurance sociale (Kela) en **Finlande**, par exemple, dispose de deux chatbots, Kela-Kelpo et FPA-Folke, pour aider les clients à trouver des informations sur les prestations sur le portail web en libre-service de Kela. Les chatbots parlent deux langues, le finnois et le suédois, et comprennent également l'anglais, ce qui permet aux personnes d'obtenir des informations et de remplir des demandes de prestations. Ils fournissent également des conseils personnalisés en fonction de variables contextuelles lorsque les clients remplissent des demandes pour des programmes spécifiques, tels que les prestations parentales. En **Allemagne**, le chatbot et le robot téléphonique « Digital First Contact » sont en cours de développement afin d'améliorer l'accès des clients aux services de conseil les plus appropriés (Public Sector Tech Watch, 2025<sup>[17]</sup>). En **Corée**, un service de conversation personnalisé basé sur l'IA (qui mémorise les conversations précédentes et les utilise pour l'appel suivant) est utilisé pour vérifier le bien-être des personnes une ou deux fois par semaine, en discutant avec elles pendant environ deux minutes (OCDE, 2024<sup>[11]</sup>). Aux **États-Unis**, le ministère de la Santé et des Services sociaux a utilisé des chatbots alimentés par l'IA pour diverses tâches, notamment pour répondre aux questions des demandeurs de subventions et fournir une assistance aux utilisateurs du « Child Welfare Information Gateway » (Portail d'information sur la protection de l'enfance) (Administration américaine pour les enfants et les familles, 2024<sup>[18]</sup>).

Les chatbots se sont révélés particulièrement utiles pour fournir des informations pendant la pandémie de COVID-19, démontrant ainsi leur potentiel considérable pour élargir l'accès à l'aide et libérer du temps aux travailleurs sociaux afin qu'ils puissent se consacrer à des tâches plus complexes. Entre mars et mai 2020, l'assistant conversationnel intelligent de l'administration **norvégienne** chargée de l'emploi et de la protection sociale (NAV) a répondu à plus de 8 000 demandes par jour, contre 2 000 avant la COVID-19.

Une évaluation a révélé que les principaux facteurs de réussite étaient la formation de l'assistant à partir d'une base de connaissances mise à jour quotidiennement et l'existence d'un lien permanent entre l'assistant et un expert humain. De nouveaux thèmes ont été ajoutés, notamment pour aider les employeurs et les travailleurs indépendants (AISS, 2020<sup>[19]</sup>). **La Belgique et la Finlande** ont également déployé des chatbots pour alléger la charge des centres d'appel pendant la COVID-19 (ibid.).

Cependant, tous les chatbots ne sont pas développés à l'aide de la technologie de l'IA. Certains chatbots se contentent de reproduire les actions et les tâches humaines dans des systèmes numériques, généralement des tâches répétitives et basées sur des règles, c'est-à-dire qu'ils ne réfléchissent pas et n'apprennent pas. **L'Agence du revenu du Canada** (ARC), par exemple, développe actuellement des solutions de chatbot et de chat en ligne qui reposent sur différents niveaux de maturité de l'IA. Le chatbot existant est basé sur des règles et utilise un modèle de questions-réponses. Il ne peut répondre qu'à partir de sa base de connaissances avec des réponses correspondant à un ensemble spécifique de sujets.

### **Automatisation des processus administratifs**

Les gouvernements utilisent l'IA pour automatiser les processus administratifs et réduire le temps consacré par les fonctionnaires aux tâches administratives. Il s'agit par exemple du traitement de grandes quantités de données provenant de bases de données traditionnelles et de textes et images non structurés issus de documents papier numérisés. Cela présente l'avantage potentiel de libérer du temps pour le personnel, qui peut alors se consacrer aux bénéficiaires qui ont besoin d'aide.

**L'Autriche, le Canada, la Finlande et les États-Unis** utilisent actuellement des techniques d'IA pour automatiser les processus administratifs dans le domaine de la protection sociale. Emploi et Développement social Canada (EDSC), par exemple, a exploité le traitement du langage naturel (NLP) pour automatiser l'examen des commentaires en texte libre reçus sur les dossiers d'emploi, contribuant ainsi à réduire la charge de travail manuelle des agents de Service Canada (OCDE, 2024<sup>[11]</sup>).

L'institution finlandaise d'assurance sociale (*Kansaneläkelaitos*) utilise la reconnaissance d'images par IA pour automatiser les processus administratifs grâce à la reconnaissance de documents, tandis que l'agence autrichienne d'assurance sociale (*Dachverband der österreichischen Sozialversicherungsträger*) a recours à un système de reconnaissance vocale pour soutenir les services des centres d'appel en transférant automatiquement les demandes aux services compétents, et a utilisé l'IA pour distribuer automatiquement les courriels aux services concernés, avec un taux de précision d'environ 93 %. En outre, un projet est en cours pour mettre en place un processus de remboursement médical semi-automatisé basé sur l'IA, dans lequel l'IA est utilisée pour automatiser plusieurs tâches telles que la reconnaissance des documents soumis, la classification des diagnostics selon les codes CIM-10 et l'extraction des données nécessaires au remboursement (par exemple, le montant de la facture et l'IBAN). Le traitement semi-automatisé accélère le processus de remboursement et aide le personnel concerné (ISSA, 2020<sup>[19]</sup>). Au Brésil, l'Institut national de sécurité sociale utilise l'IA pour accélérer l'identification des bénéficiaires décédés afin d'éviter les paiements indus, qui dans de nombreux pays sont traités manuellement (et lentement) (ISSA, 2020<sup>[19]</sup>).

L'IA est également utilisée pour rationaliser le traitement des demandes d'indemnisation. La COVID-19 a entraîné un volume sans précédent de demandes d'aide et de prestations. Au Canada, l'accent mis pendant la pandémie sur la mise en œuvre de la Prestation d'urgence pour la réponse à la COVID-19 (PURC) et la simplification des demandes d'assurance-emploi (AE) a contribué à l'accumulation d'un arriéré de demandes à examiner lorsque le traitement normal a repris. La mise en œuvre d'un modèle d'apprentissage automatique permettant de prédire le résultat du recalcul de l'AE avant la PRERB visait à réduire au minimum le nombre de demandes anciennes (antérieures à mars 2020) devant être examinées par un agent. Ce modèle a été utilisé pour prédire le résultat le plus probable de chaque recalcul et trier les tâches associées en conséquence, en tenant compte de l'impact probable des recalculs sur les demandeurs (OCDE, 2024<sup>[11]</sup>).

Aux États-Unis, le ministère de la Santé et des Services sociaux a déployé l'IA pour toute une série de tâches administratives « essentielles à la mission », notamment la synthèse de documents, la rédaction et la révision de contenus, l'analyse qualitative de données, l'aide à la revue de la littérature, la conception de l'aménagement intérieur des équipements publics (par exemple, pour les logements sociaux) et d'autres tâches encore. Sur les 272 cas répertoriés pour 2024, 133 sont classés comme « facilitant la mission » (HHS, 2025<sup>[5]</sup>). Le Japon note également que l'IA est utilisée dans le traitement des demandes afin d'identifier les cas qui nécessitent un examen humain, et que ce taux d'examen a diminué au cours des deux dernières années.

### **Détection des erreurs et des fraudes par les fournisseurs et retour d'information par les clients**

L'IA offre également un potentiel considérable tant pour les pouvoirs publics que pour les clients en termes d'amélioration de la qualité et de la précision des demandes de remboursement et de la prestation des services. Certains pays utilisent déjà des techniques d'IA pour détecter les erreurs et les fraudes. Le **Royaume-Uni**, par exemple, utilise l'IA pour aider à détecter les fraudes dans les demandes de prestations sociales (Marr, 2018<sup>[20]</sup>). **La Corée** a réalisé des progrès considérables dans l'utilisation des mégadonnées pour la détection des erreurs et des fraudes dans les demandes de remboursement d'assurance maladie. Le Service national d'assurance maladie (NHIS) de Corée dispose de mégadonnées sur un large éventail de variables socio-économiques, de comportements en matière de santé, d'utilisation des soins de santé et de soins de longue durée, auxquelles sont appliqués des algorithmes d'audit intelligents afin de prédire les établissements de santé présentant un risque élevé de demandes frauduleuses, ce qui permet d'aider les enquêteurs de manière préventive (ISSA, 2022<sup>[21]</sup>).

La détection des fraudes est un domaine dans lequel l'utilisation des techniques d'IA devrait se développer rapidement. Tout comme l'IA devrait entraîner une augmentation du volume et de la sophistication des fraudes et des escroqueries, elle fournira également des solutions de détection et de prévention. Au Portugal, par exemple, la reconnaissance faciale ou vocale basée sur l'IA est utilisée pour vérifier que les fonctionnaires vivant à l'étranger sont bien en vie afin qu'ils puissent continuer à percevoir leur pension.

La détection de la fraude est également un domaine où les questions éthiques occupent une place importante. Aux Pays-Bas, entre 2005 et 2019, de nombreuses familles ont été accusées à tort de fraude en raison d'algorithmes discriminatoires (voir ci-dessous). Alors que les gouvernements cherchent à développer l'utilisation de l'IA dans ce domaine, il sera particulièrement important de mettre en place des garanties appropriées contre les biais.

Une autre utilisation prometteuse de l'IA pour promouvoir la responsabilité des pouvoirs publics et l'amélioration de la qualité, qui n'est pas encore couramment déployée, est le suivi des plaintes. Les techniques d'apprentissage automatique ont été utilisées pour identifier les facteurs qui augmentent la probabilité d'une plainte, ce qui pourrait aider les décideurs politiques et les praticiens à identifier les lacunes potentielles dans la prestation des services et à hiérarchiser les domaines à améliorer (Ohlenburg, 2020<sup>[12]</sup>).

### **Analyse prédictive, aide à la décision et interventions sur mesure : l'avenir de l'IA dans la protection sociale ?**

L'un des cas d'utilisation les plus prometteurs de l'IA dans le domaine de la protection sociale réside dans l'analyse prédictive. Celle-ci pourrait jouer un rôle important dans la prévention et la gestion des chocs ou des crises, l'identification des personnes à risque et le ciblage des mesures préventives, l'identification des groupes de bénéficiaires (potentiels) (par exemple, par région ou selon des caractéristiques sociodémographiques) afin de mieux adapter les interventions, ainsi que l'amélioration de la portée des programmes sociaux et la réduction du non-recours à ceux-ci.

#### **L'analyse prédictive pour prévoir la demande et les chocs**

L'analyse prédictive est déjà utilisée pour identifier les personnes susceptibles de subir un choc ou un événement indésirable, afin de permettre la prise de mesures pour éviter cet événement ou en atténuer l'impact lorsqu'il ne peut être évité. Des outils d'IA sont utilisés pour prévoir la probabilité d'accidents du travail en **Allemagne**, par exemple, et contribuent à prévenir les accidents en identifiant les entreprises qui ont le plus besoin d'accompagnement et en affectant des inspecteurs sur place en conséquence. Le modèle agrège les données provenant des chantiers, des travailleurs et des employeurs afin d'identifier les risques sur les chantiers, mettant ainsi en œuvre une approche fondée sur les risques pour allouer des ressources limitées dans des circonstances où les inspecteurs ne peuvent atteindre que 1 % des entreprises (GIZ, 2024<sup>[22]</sup>). Les inspecteurs établissent des rapports sous forme de notes vocales qui sont ensuite transcrites, et les résultats sont utilisés pour améliorer le modèle d'apprentissage automatique. Le projet semble porter ses fruits, les recommandations automatisées libérant du temps que les travailleurs peuvent consacrer aux inspections. Il aurait déjà touché plus de 4 000 sites qui n'avaient jamais été inspectés auparavant et généré des économies estimées à environ 250 millions d'euros en indemnités et en coûts de réhabilitation sur une période de cinq ans (GIZ, 2024<sup>[22]</sup>). L'IA pourrait éventuellement être utilisée à l'avenir pour des analyses prédictives similaires par les inspections du travail, par exemple pour prédire la probabilité de violations du droit du travail et estimer la demande d'inspections. Ce type de système d'IA pourrait également être

utilisées, par exemple, pour déterminer les services de garde d'enfants qui méritent le plus d'être inspectés lorsque les ressources sont limitées.

L'analyse prédictive a également été utilisée pour atténuer et aider à amortir l'impact des chocs lorsqu'ils ne peuvent être évités, notamment au Togo, où elle a été utilisée pour prévoir les inondations, avertir les citoyens afin qu'ils prennent des mesures pour en atténuer l'impact et mettre à disposition des financements avant que les inondations ne surviennent (Ohlenburg, 2020<sup>[12]</sup>). Des applications similaires ont été explorées en **Europe**, où des chercheurs ont collaboré avec l'Agence spatiale européenne pour étudier l'utilisation de l'IA afin de développer un système de prévision et d'alerte précoce des inondations, et en Irlande, où l'organisation à but non lucratif « CeADAR », soutenue par le gouvernement, a utilisé l'IA pour prévoir les inondations futures dans les zones inondables (Université de Lancaster, 2021<sup>[23]</sup>) ; (CeADAR, n.d.<sup>[24]</sup>).

### **Analyse prédictive pour l'identification et l'intervention précoce**

Les nouvelles technologies offrent des possibilités considérables pour identifier et inscrire les bénéficiaires potentiels dans les programmes sociaux, même s'il convient de noter que des progrès importants pourraient également être réalisés en utilisant les outils existants et des données personnelles interconnectées de bonne qualité (Frey et Hye, 2024<sup>[25]</sup>). Si l'on se projette dans l'avenir et que l'on envisage les cas d'utilisation actuels et potentiels de l'IA, il existe des possibilités tant pour identifier et inscrire les personnes qui n'ont pas été en contact avec les organismes de protection sociale que pour identifier les personnes qui sont déjà intégrées dans certains programmes sociaux et qui sont susceptibles d'être éligibles à d'autres.

L'analyse prédictive pourrait également contribuer à améliorer les résultats en aidant les prestataires à prendre des décisions. En traitant et en synthétisant rapidement et systématiquement des ensembles de données volumineux, complexes et non structurés, l'IA pourrait aider les décideurs politiques et les praticiens à identifier et hiérarchiser les risques et à élaborer des interventions ciblées et personnalisées.

Compte tenu du retour sur investissement substantiel de certaines interventions préventives dans les services sociaux, les avantages sociaux et économiques de l'analyse prédictive pourraient être considérables. Un domaine dans lequel cela est déjà mis en œuvre, à l'aide de modèles prédictifs, est la prévention et la réduction du sans-abrisme. À Los Angeles, **aux États-Unis**, des chercheurs universitaires et les autorités locales ont mis en place un programme visant à identifier et à venir en aide aux habitants les plus exposés au risque de se retrouver sans domicile pour la première fois (California Policy Lab, 2024<sup>[26]</sup>). Le risque individuel est prédit à partir de données sur les facteurs de risque chez les adultes célibataires bénéficiant déjà des services sociaux « traditionnels ». Le modèle prédictif suggère que l'utilisation élevée (multiple), croissante et ponctuelle des services<sup>(4)</sup> constitue un signe avant-coureur d'un risque élevé de sans-abrisme chez les personnes en situation d'extrême pauvreté (California Policy Lab, 2020<sup>[27]</sup>). Une initiative similaire a été mise en œuvre au Royaume-Uni, où le conseil municipal de Maidstone et des partenaires externes ont développé « OneView », un outil de données et d'analyse prédictive qui agrège les données de différents organismes afin de permettre aux agents de première ligne d'identifier et de contacter les résidents exposés au risque de sans-abrisme (Šuica et al., 2024<sup>[28]</sup>). La plateforme envoie des alertes aux agents de première ligne en cas d'événements susceptibles d'indiquer un risque de sans-abrisme, tels que le non-paiement des factures de services publics (Šuica et al., 2024<sup>[28]</sup>), avec des résultats prometteurs (Maidstone Borough Council, 2022<sup>[29]</sup>) (Ernst & Young, 2022<sup>[30]</sup>).

Les techniques d'apprentissage automatique ont également été utilisées pour **améliorer les prévisions de récurrence de violence domestique**, améliorant ainsi les évaluations des risques existantes. En exploitant les données de la police au **Royaume-Uni**, des chercheurs ont découvert que les algorithmes d'apprentissage automatique qui tiennent compte des antécédents criminels sont plus efficaces que les outils traditionnels d'évaluation des risques pour prédire la récurrence violente dans les cas de violence domestique, en partant du principe qu'il vaut mieux surestimer que sous-estimer (Grogger, Ivandic et Kirchmaier, 2020<sup>[31]</sup>). L'erreur de prédiction négative de l'algorithme, c'est-à-dire la proportion de cas qui auraient été prédits à tort comme ne présentant pas de violence, est relativement faible, à 6,3 %, contre 11,5 % pour l'outil de dépistage traditionnel (Centre for Economic Performance, 2020<sup>[32]</sup>). Les chercheurs suggèrent que l'apprentissage automatique pourrait contribuer à pallier la faible capacité prédictive et les incohérences considérables dans l'application des outils de dépistage existants, et aider à hiérarchiser les appels signalant des violences domestiques en fonction du risque, afin de garantir que les appels les plus urgents reçoivent les réponses les plus rapides. Les chercheurs appellent néanmoins à la mise au point d'un instrument de dépistage plus sensible afin de remédier aux faux positifs qui apparaissent lors de ce type de dépistage, ce qui pourrait s'avérer particulièrement important. La tolérance à l'égard des faux positifs dépendra probablement aussi des mesures déclenchées par un signalement de risque :



La tolérance aux faux positifs serait probablement plus élevée si la prédiction était utilisée pour trier les appels plutôt que pour prendre des mesures plus sévères.

**Les soins et l'aide au revenu pour les personnes âgées** constituent un autre domaine présentant un énorme potentiel en matière de prévention et d'intervention précoce. Des pays tels que **la Suède** et le **Royaume-Uni** utilisent l'IA pour identifier les besoins en matière de réadaptation et les risques de blessures, respectivement, ce qui permet d'adapter les interventions précoces. En Suède, un modèle d'IA a été utilisé pour prévoir à la fois les besoins et le potentiel des services préventifs, dans le but d'améliorer l'allocation des ressources grâce à un soutien préventif ciblé, tandis qu'au Royaume-Uni (Angleterre), un outil d'IA est en cours de déploiement dans le cadre des visites à domicile, avec une précision de 97 % dans la prédiction du risque de chute des patients, l'une des principales causes d'hospitalisation d'urgence chez les personnes âgées (Šuica et al., 2024<sup>[28]</sup>) ; (NHS England, 2025<sup>[33]</sup>). Il s'agit d'un domaine où la prévention et l'allocation efficace des ressources deviennent de plus en plus importantes, les travaux récents de l'OCDE montrant que le vieillissement de la population et la baisse de la fécondité vont considérablement accroître la demande de soins de santé et d'aide sociale dans les années à venir, tout en réduisant à la fois la main-d'œuvre et les fonds disponibles pour les soins (OCDE, 2024<sup>[2]</sup>). **L'IA est également utilisée pour identifier de manière préventive les risques de pauvreté et d'endettement** grâce à l'analyse de données provenant de différentes agences dans différentes municipalités aux **Pays-Bas** (Public Sector Tech Watch, 2025<sup>[17]</sup>).

Des systèmes d'aide à la décision basés sur des algorithmes ont également été développés ou testés pour **aider les travailleurs sociaux à identifier les cas possibles de maltraitance et/ou d'abus envers les enfants** dans des pays tels que **l'Australie, les Pays-Bas, la Nouvelle-Zélande et les États-Unis**, bien que certains chercheurs suggèrent que des travaux supplémentaires sont nécessaires pour garantir que ces outils sont adaptés à leur objectif (Gillingham, 2019<sup>[34]</sup>) (Hall et al., 2024<sup>[35]</sup>). Des techniques d'apprentissage automatique ont été testées pour identifier et signaler les cas possibles de maltraitance infantile aux Pays-Bas, par exemple, où des chercheurs ont mis au point un système d'aide à la décision combinant des données sur la taille et le poids des enfants avec des commentaires libres rédigés par des professionnels de la santé, afin d'alerter les travailleurs sociaux sur les cas possibles de maltraitance infantile (Amrit et al., 2017<sup>[36]</sup>), cité dans (Robila et Robila, 2020<sup>[37]</sup>). De telles approches offrent la possibilité d'améliorer l'identification et l'intervention sans avoir à collecter de données supplémentaires (Robila et Robila, 2020<sup>[37]</sup>), et elles pourraient avoir de nombreuses applications dans les services sociaux pour compléter et/ou améliorer les méthodes d'évaluation des risques existantes.

**Le risque de faux positifs nécessite la prudence et l'atténuation des risques** de la part des décideurs politiques et des prestataires de services. Signaler une famille comme présentant un risque élevé de conséquences négatives a des implications graves, notamment pour les relations de la famille avec l'État et pour le bien-être de la famille. L'utilisation signalée de logiciels pour surveiller les appareils scolaires afin de prévenir l'automutilation chez les élèves aux États-Unis, par exemple, a suscité des réactions mitigées. Dans certains cas, la détection a permis d'intervenir rapidement auprès d'enfants et de jeunes présentant un risque d'automutilation ou de suicide, mais de nombreux défenseurs et parents s'inquiètent des faux positifs (par exemple, signaler un élève comme suicidaire alors qu'il ne l'est pas) et de la surveillance continue des jeunes (Barry, 2024<sup>[38]</sup>) ; (Académie américaine de pédiatrie, 2025<sup>[39]</sup>).

### **Améliorer la sensibilisation et réduire le non-recours aux programmes**

Outre l'amélioration de l'identification et de la prévention des risques chez les personnes et les ménages déjà en contact avec les programmes sociaux, l'IA pourrait aider à identifier de manière proactive les bénéficiaires *potentiels* des programmes sociaux et à les atteindre afin de remédier à la non-participation des personnes susceptibles d'y avoir droit.

La non-participation aux programmes sociaux reste un défi persistant dans tous les pays de l'OCDE (Frey et Hyee, 2024<sup>[25]</sup>). Certaines populations qui ont le plus besoin des prestations et des services sociaux peuvent être difficiles à atteindre et peu enclines à rechercher activement de l'aide. Par exemple, les jeunes qui ne sont ni scolarisés ni en formation bénéficient énormément d'interventions précoces visant à prévenir le chômage de longue durée, mais il peut être difficile de les inscrire à des programmes sociaux car ils ne connaissent généralement pas les organismes compétents ou ne les contactent pas (OCDE, 2024<sup>[40]</sup>). Certains pays s'efforcent activement de contacter ces jeunes, mais cela nécessite des recherches approfondies et de nombreux jeunes sans emploi passent de plus en plus de temps en ligne plutôt que dans les espaces publics. Dans quelques cas, l'IA est utilisée pour aider à identifier les personnes qui ont besoin d'aide et leur offrir une assistance proactive.



En **Espagne**, le département des services sociaux de la mairie de Madrid utilise un assistant virtuel alimenté par l'IA pour **appeler les personnes âgées de plus de 75 ans afin d'évaluer si elles risquent de souffrir de solitude**. L'outil alimenté par l'IA pose une série de questions structurées afin de déterminer si ces personnes se sentent seules, ont de la famille, des amis ou quelqu'un vers qui se tourner, ou souhaitent bénéficier d'un suivi de la part des services municipaux. Grâce à ce dispositif, plus de 600 personnes ont été identifiées et contactées par des travailleurs sociaux du département des services sociaux (Lara-Montero, 2024<sup>[41]</sup>). **La Grèce** a également mis en place des interventions personnalisées qui contribuent à améliorer la sécurité, à réduire le sentiment de solitude et à renforcer l'autonomie (Public Sector Tech Watch, 2025<sup>[17]</sup>).

Bien entendu, l'IA n'est pas un outil *indispensable* pour identifier les groupes de personnes dans le besoin. Comme dans de nombreux autres exemples présentés ici, des méthodes d'analyse de données plus traditionnelles ont été utilisées pour identifier des groupes d'individus ou de ménages ayant des besoins similaires, souvent à partir de données d'enquête permettant d'établir des estimations probabilistes qui peuvent être ventilées selon différents profils sociodémographiques (Frey, Hyee et Minondo Canto, 2024<sup>[42]</sup>).

### **Adapter les interventions aux besoins des clients**

Une autre utilisation potentielle de l'IA dans les programmes sociaux consiste à mieux adapter les interventions aux besoins des utilisateurs. L'IA pourrait favoriser une plus grande personnalisation au niveau de la population – par exemple, en regroupant les bénéficiaires de programmes sociaux afin d'adapter les services proposés à leurs besoins (Ohlenburg, 2020<sup>[12]</sup>) – ou au niveau individuel. En **Corée**, le Service d'indemnisation et de protection sociale des travailleurs (COMWEL) a mis au point le Système intelligent de recommandation en matière de réadaptation (IRRS) afin d'améliorer son soutien aux travailleurs accidentés. Si le COMWEL met en œuvre depuis 2011 des plans de réadaptation personnalisés pour les travailleurs accidentés, ce processus s'appuie sur des informations limitées et sur l'expérience des responsables, ce qui se traduit par une qualité et une rapidité variables des services. L'IRRS, un système basé sur l'IA, a été développé pour sélectionner les travailleurs accidentés susceptibles d'être actifs et concevoir des services de réadaptation scientifiquement adaptés à leurs besoins. L'IRRS calcule un indice de vulnérabilité à partir des données administratives accumulées depuis 2011 sur 98 millions de travailleurs, comprenant des informations détaillées sur les indemnités d'accident du travail, l'assurance chômage et la gestion des cas de réadaptation, à l'aide d'un filtrage basé sur des règles et d'une méthodologie de raisonnement par cas.

L'IRRS coréen propose également un plan de réadaptation basé sur le modèle d'IA. Les travailleurs sélectionnés pour bénéficier d'une réadaptation et d'un retour à l'emploi sont consultés par les experts en réadaptation de COMWEL avant que les plans générés par l'IA ne soient finalisés. L'IRRS aurait aidé COMWEL à améliorer la cohérence de la qualité des services à l'échelle nationale tout en garantissant des interventions opportunes et appropriées afin d'améliorer le taux de retour à l'emploi (AISS, 2022<sup>[21]</sup>).

**L'Agence danoise pour le marché du travail et le recrutement (STAR)** a mis au point un modèle de profilage utilisant des techniques d'apprentissage automatique qui prédisent la probabilité qu'une personne devienne chômeur de longue durée (plus de 26 semaines). Ce modèle combine des données issues de registres administratifs et d'une enquête en ligne qui recueille des informations sur le comportement. En collaboration avec l'université de Copenhague, un nouvel outil d'enquête est actuellement en cours d'élaboration afin de saisir les traits de personnalité structurels tels que les préférences en matière de temps et de risque. L'utilisation du système est facultative pour les demandeurs d'emploi, mais ceux qui y ont recours ont pleinement accès aux résultats du modèle. Le système ne dirige pas automatiquement les demandeurs d'emploi vers des programmes actifs du marché du travail (PAMT), mais soutient les travailleurs sociaux qui conservent toute leur responsabilité discrétionnaire (Desiere, Langenbucher et Struyven, 2019<sup>[43]</sup>).

**L'utilisation de l'IA dans les services publics de l'emploi (SPE) est très répandue.** Une étude de l'OCDE sur l'utilisation de l'IA dans les pays de l'OCDE révèle que près de la moitié des SPE utilisent l'IA pour améliorer leurs services, le plus souvent pour mettre en relation les demandeurs d'emploi et les offres d'emploi et pour identifier les besoins des demandeurs d'emploi en matière d'aide à l'aide d'outils de profilage (Brioscú et al., 2024<sup>[44]</sup>). Aux **Pays-Bas**, le programme « CompetentNL » utilise l'IA pour faciliter l'adéquation des compétences et des emplois (CompetentNL, 2025<sup>[45]</sup>), tandis qu'au **Portugal**, l'agence pour l'emploi utilise l'IA pour mieux identifier les personnes exposées à un risque élevé de chômage de longue durée et soutenir une allocation plus efficace des ressources de l'agence afin de répondre aux besoins des personnes sans emploi (Public Sector Tech Watch, 2025<sup>[17]</sup>). Pour un examen complet, voir (Brioscú et al., 2024<sup>[44]</sup>).

Les gouvernements commencent également à mettre en œuvre l'IA pour aider les travailleurs sociaux dans leur travail quotidien, notamment en les aidant à prendre des décisions. À **Barcelone, en Espagne**, par exemple, un algorithme formé à partir de 300 000 entretiens propose des ressources et des services aux 700 employés qui servent environ 50 000 clients par an, même si la décision finale revient au membre du personnel (Public Sector Tech Watch, 2025<sup>[17]</sup>). Plusieurs conseils locaux au Royaume-Uni ont testé un outil de prise de notes basé sur l'IA qui permet de saisir, de transcrire et de formuler des recommandations à partir des réunions avec les clients, ce qui libère du temps que les travailleurs peuvent consacrer aux soins et à l'aide de première ligne. Cet outil peut générer une liste d'actions proposées que les employés peuvent examiner et approuver. Le conseil municipal de Camden, à Londres (**Royaume-Uni**), a par exemple testé cet outil pour créer un résumé des conversations avec les habitants, les aidants et d'autres prestataires de services, et pour générer une liste d'actions suggérées (Camden Care Choices, 2024<sup>[46]</sup>). Au conseil municipal d'Ealing Borough, à Londres, en Angleterre (**Royaume-Uni**), les travailleurs sociaux utilisent cet outil pour transcrire les réunions avec les clients, avec leur consentement, ce qui aurait permis de réduire d'au moins 40 % le temps consacré par les travailleurs sociaux à des tâches administratives (Local Government Association, 2025<sup>[47]</sup>) (OCDE, 2025<sup>[48]</sup>).

### ***Lutter contre la discrimination humaine (réelle ou perçue) dans le processus de traitement des réclamations***

Le fait que l'IA puisse être sujette à des préjugés raciaux, ethniques ou sexistes fait l'objet de nombreux débats (Parikh, Teeple et Navathe, 2019<sup>[49]</sup>) (Ntoutsis et al., 2020<sup>[50]</sup>) (Booth, 2024<sup>[51]</sup>). Bien entendu, la prise de décision humaine n'est pas non plus exempte de préjugés, et l'IA pourrait en fait être utilisée pour minimiser les préjugés ou leur apparence, même si cette utilisation n'a pas encore été signalée dans la pratique par les ministères des pays de l'OCDE.

De nombreux programmes sociaux sont assortis de règles d'éligibilité claires et objectives, telles que des plafonds de revenus ou de patrimoine, ou encore des conditions de cotisation, mais d'autres règles peuvent être sujettes à interprétation. Par exemple, la détermination des efforts suffisants en matière de recherche d'emploi pour les bénéficiaires d'allocations chômage comporte une part de subjectivité. De nombreuses prestations sont assorties de « conditions de comportement » que les bénéficiaires doivent respecter, telles que la participation à des réunions avec des assistants sociaux ou la communication en temps voulu d'informations actualisées. Le non-respect de ces conditions entraîne des sanctions, telles que la suppression partielle ou totale des prestations (Hyee et al., 2024<sup>[52]</sup>). Étant donné qu'il existe une marge d'appréciation pour déterminer si les exigences comportementales sont respectées ou non, il existe un risque de discrimination fondée sur la race, l'origine ethnique, la religion ou le sexe dans les sanctions liées aux prestations.

La discrimination dans l'octroi ou le retrait des prestations sociales peut être due à des préjugés (explicites ou implicites) de la part des travailleurs sociaux. Cela est plus probable lorsqu'il s'agit de déterminer si un demandeur a droit à la prestation ou si la situation est légitimement complexe et ambiguë (voir, par exemple, (Assouline, Gilad et Ben-Nun Bloom, 2021<sup>[53]</sup>) et (Emeriau, 2022<sup>[54]</sup>).

**L'IA pourrait aider à détecter et à signaler les cas de discrimination dans les décisions d'octroi de prestations**, en particulier lorsque ces décisions sont prises dans des circonstances ambiguës (telles que le handicap, mais aussi l'intensité de la recherche d'emploi ou d'autres facteurs similaires) ou par des agents chargés des dossiers moins qualifiés ou moins expérimentés.

Selon la documentation limitée disponible, cette discrimination est principalement statistique et non fondée sur les préférences (voir, par exemple,

Bell et Jilke, 2024<sup>[55]</sup>), ce qui signifie que la détection et le signalement pourraient améliorer la précision des décisions. Par exemple, lorsque l'accès à une prestation ou à un service est limité (par exemple, un programme est sursouscrit), les agents chargés des dossiers donnent souvent la priorité aux clients qui ont les meilleures chances de succès. Les agents des services publics de l'emploi peuvent orienter les demandeurs d'emploi qu'ils jugent les plus susceptibles de trouver un emploi vers des programmes de formation afin d'augmenter les chances de les retirer de leurs fichiers, un phénomène connu sous le nom de « écrémage » (Carter et Whitworth, 2014<sup>[56]</sup>). Si les véritables facteurs déterminants de la réussite d'une intervention donnée sont inconnus, les agents chargés des dossiers peuvent recourir à des moyennes au niveau du groupe, et donc à une discrimination statistique.

Même s'il n'y a pas de discrimination réelle dans le processus de demande, les demandeurs potentiels issus de groupes défavorisés peuvent hésiter à solliciter des prestations sociales par crainte d'être victimes de discrimination. Dans l'enquête « Risques importants » de l'OCDE de 2024, les personnes s'identifiant comme appartenant à une minorité en raison de leur origine ethnique ou de la couleur de leur peau étaient en moyenne 3 points de pourcentage plus susceptibles de penser qu'elles ne seraient pas

traités équitablement par l'administration chargée de traiter leur demande d'aide sociale ou de services publics, par rapport aux personnes qui ne se sont pas identifiées comme appartenant à une minorité en raison de leur origine ethnique ou de la couleur de leur peau.<sup>5</sup>Cela correspond aux conclusions de l'enquête de l'OCDE sur la confiance, qui montre qu'en moyenne dans les 29 pays de l'OCDE, il existe un écart de 14 points de pourcentage dans la confiance accordée aux institutions gouvernementales entre les personnes qui s'identifient comme appartenant à un groupe défavorisé (29,6 %) et celles qui ne le font pas (43,3 %) (OCDE, 2024<sup>[57]</sup>).

Le fait que des biais dans le processus de traitement des réclamations soient susceptibles d'apparaître dans des situations ambiguës et où l'information est limitée montre que l'IA peut améliorer l'équité dans ce processus, à condition que des mesures de contrôle et de protection appropriées soient mises en place.

## Mettre en place des mesures de protection pour garantir une utilisation sûre et fiable de l'IA

Si les technologies de pointe peuvent jouer un rôle important dans l'amélioration de la conception, de la mise en œuvre et de la couverture des programmes sociaux, elles posent également des défis complexes aux gouvernements, et les risques liés à l'adoption de nouvelles technologies numériques et de nouvelles technologies de données peuvent être considérables (Verhagen, 2024<sup>[58]</sup>). Les gouvernements doivent mettre en place des cadres et des procédures de responsabilisation appropriés, sans lesquels les innovations fondées sur la technologie et les données risquent de priver les citoyens de leur pouvoir d'action, de les désengager et d'éroder la confiance du public. Le principe 1.5 des principes de l'OCDE sur l'IA précise que les acteurs de l'IA doivent être responsables du bon fonctionnement des systèmes d'IA, en fonction de leurs rôles, du contexte et conformément à l'état de l'art (OCDE, 2019<sup>[59]</sup>).

**Des problèmes de mise en œuvre très médiatisés ont attiré l'attention du public sur les risques liés à l'IA,** notamment en Australie et aux Pays-Bas, où des systèmes automatisés de détection des fraudes et de recouvrement de créances ont eu des conséquences imprévues et importantes. Aux Pays-Bas, près de 26 000 familles ont été accusées à tort de fraude entre 2005 et 2019 par les autorités fiscales néerlandaises en raison d'algorithmes discriminatoires. Des profils de risque ont été créés pour les personnes demandant des allocations pour garde d'enfants, dans lesquels les « noms à consonance étrangère » et la « double nationalité » étaient utilisés comme indicateurs de fraude potentielle. Des milliers de familles à faibles et moyens revenus ont ensuite fait l'objet d'un contrôle approfondi, ont été accusées à tort de fraude et ont été invitées à rembourser les allocations qu'elles avaient obtenues légalement, qui s'élevaient dans certains cas à plusieurs milliers d'euros. En conséquence, beaucoup se sont endettés, certains se sont retrouvés dans la pauvreté et d'autres ont perdu leur logement et/ou leur emploi. Plus de 1 000 enfants ont été placés sous la tutelle de l'État (Parlement européen, 2022<sup>[60]</sup>). Lorsque les erreurs ont été révélées, le gouvernement néerlandais a démissionné et l'évaluation ultérieure du fonctionnement du programme de détection des fraudes a abouti à la conclusion « douloureuse » en 2022 qu'il y avait eu racisme institutionnel (Gouvernement des Pays-Bas, 2022<sup>[61]</sup>). Aujourd'hui, les Pays-Bas surveillent de près l'utilisation de l'IA dans l'administration publique ; l'Agence pour l'assurance des travailleurs, par exemple, a mis en place un comité consultatif sur l'éthique de l'IA qui donne des conseils sur tous les projets liés à l'IA.

Le programme australien Robodebt, mis en place en 2015 pour évaluer les droits aux prestations, met en évidence les défis liés à la détection d'un problème systémique dans un système automatisé (même s'il ne s'agit pas d'une IA). De 2015 à 2019, le ministère des Services sociaux a mis en œuvre un programme de recouvrement de créances, Robodebt, afin de récupérer les trop-perçus versés aux bénéficiaires de prestations sociales depuis 2010-2011. Pour calculer les trop-perçus, les données relatives aux prestations sociales ont été croisées avec les données sur les revenus annuels de l'administration fiscale australienne et un processus appelé « lissage des revenus » a été utilisé pour évaluer les revenus et les droits aux prestations. Des avis de dette étaient ensuite envoyés aux bénéficiaires de l'aide sociale concernés, qui devaient prouver qu'ils n'avaient pas de dette. Ce processus a donné des résultats inexacts et n'était pas conforme aux dispositions de la loi de 1991 sur la sécurité sociale relatives au calcul des revenus. Malgré les critiques du public et les conclusions défavorables du tribunal administratif d'appel dans certains cas individuels, la nature systémique du problème n'a pas été immédiatement identifiée et le programme a continué à fonctionner jusqu'en 2019. Un recours collectif a été réglé en juin 2020 et le Premier ministre de l'époque a présenté des excuses, suivies d'une commission royale d'enquête sur le programme Robodebt en 2022, qui a formulé 57 recommandations.

Les affaires Robodebt et des allocations familiales aux Pays-Bas soulignent l'importance cruciale de la transparence et de l'explicabilité, ainsi que la nécessité d'une intervention humaine significative, en particulier lorsque des décisions automatisées peuvent avoir un impact potentiel et significatif sur la vie des personnes. Le manque de transparence a contribué au scandale néerlandais

, en raison d'une implication humaine insuffisante et d'un contrôle insuffisant des décisions automatisées (Błażej Kuźniacki, 2023<sup>[62]</sup>).

Bien que le déploiement de toute technologie de pointe comporte des risques, certains ont fait valoir que l'utilisation de techniques d'IA de type « boîte noire » devrait être évitée dans des domaines tels que la protection sociale, ce qui a conduit certains gouvernements à mettre en place des mesures supplémentaires pour garantir une utilisation avancée et sûre de l'IA. Le ministère du Travail et des Retraites (DWP) du Royaume-Uni, par exemple, a créé un programme phare sur l'intelligence artificielle (IA) afin d'explorer en toute sécurité l'utilisation des technologies émergentes d'IA générative. Conscient des opportunités et des risques, le DWP a mis en place un cadre et un processus pour explorer ces technologies de manière sûre, éthique et transparente (Adams, 2024<sup>[8]</sup>).

**La transparence et l'explicabilité constituent un principe clé de l'IA fiable dans les principes de l'OCDE relatifs à l'IA** (encadré 2), qui suggèrent qu'il devrait y avoir un certain degré d'intervention humaine dans la prise de décision automatisée, proportionnel à l'impact potentiel des résultats générés. Le principe 1.2(b) des principes de l'OCDE en matière d'IA précise que les acteurs de l'IA devraient mettre en œuvre des mécanismes et des garanties, tels que la capacité de détermination humaine, qui soient adaptés au contexte et conformes à l'état de l'art (OCDE, 2019<sup>[59]</sup>). Les réglementations internationales ont également cherché à établir des garde-fous pour une utilisation sûre et fiable de l'IA et des données à caractère personnel, en imposant des restrictions et des responsabilités à certaines utilisations (et certainement à haut risque) de l'IA dans les services publics et sociaux. C'est le cas notamment de la loi européenne sur l'IA, qui classe explicitement comme « à haut risque » les systèmes d'IA destinés à évaluer, accorder, réduire, révoquer ou récupérer l'éligibilité ou l'accès à des prestations et services d'aide publique essentiels (y compris les prestations de sécurité sociale), et qui définit une série d'exigences pour le déploiement autorisé de ces utilisations à haut risque, couvrant à la fois les systèmes d'IA et les organisations qui les déploient. Cela pourrait avoir une incidence sur la manière dont certaines applications de l'IA dans le domaine de la protection sociale peuvent être conçues et utilisées.

#### Encadré 2. Transparence et explicabilité : principes fondamentaux d'une IA fiable

Les Principes de l'OCDE relatifs à l'IA, première norme intergouvernementale en matière d'IA, ont été adoptés par les pays membres de l'OCDE et d'autres adhérents afin de promouvoir une utilisation innovante et fiable de l'IA qui respecte les droits de l'homme et les valeurs démocratiques. La transparence et l'explicabilité sont essentielles à une IA fiable, comme le soulignent les Principes, au même titre que la croissance inclusive, le développement durable et le bien-être ; les droits de l'homme et les valeurs démocratiques, y compris l'équité et la vie privée ; la robustesse, la sécurité et la sûreté ; et la responsabilité.

La transparence implique de divulguer quand des systèmes automatisés sont utilisés, par exemple pour faire une prédiction, une recommandation ou prendre une décision, la divulgation étant proportionnée à l'importance de l'interaction. La transparence consiste également à être en mesure de fournir des informations qui permettent aux personnes concernées par un système d'IA de comprendre le résultat, y compris des informations sur la manière dont un système automatisé a été développé et déployé, quelles informations ont été fournies et pourquoi. Un autre aspect de la transparence consiste à faciliter, si nécessaire, la participation du public et des différentes parties prenantes afin de favoriser la sensibilisation et la compréhension générales des systèmes automatisés et d'accroître leur acceptation et la confiance à leur égard (OCDE, 2022<sup>[63]</sup>).

L'explicabilité est le concept selon lequel le résultat d'un système ou d'un algorithme automatisé peut être expliqué d'une manière qui « a du sens » pour les personnes, ce qui permet à celles qui sont concernées par un résultat de le comprendre et de le contester. Cela implique notamment de fournir, en termes clairs et simples, et en fonction du contexte, les principaux facteurs pris en compte dans une décision, les facteurs déterminants et les données, la logique ou l'algorithme utilisés pour parvenir à cette décision (OCDE, 2022<sup>[63]</sup>). Certains algorithmes sont plus faciles à expliquer, mais potentiellement moins précis (et vice versa). Si l'exigence d'explicabilité peut avoir un impact négatif sur les performances d'un algorithme, elle peut dans certains cas constituer un facteur prépondérant.

Les principes de l'OCDE relatifs à l'IA et les réglementations spécifiques en matière de protection des données soulignent également la nécessité de mettre en œuvre des garanties appropriées pour assurer la responsabilité et l'équité, en particulier en ce qui concerne la prise de décision automatisée.

la prise de décision automatisée. Dans l'UE, le règlement général sur la protection des données (RGPD) consacre le droit des personnes physiques de ne pas être soumises à des décisions (uniquement) automatisées, à quelques exceptions près, par exemple lorsque leur utilisation est légalement autorisée *et que* l'État membre prévoit des « mesures appropriées » pour garantir les droits, libertés et intérêts légitimes de la personne concernée. Le droit de réexaminer une décision ou un résultat automatisé est une caractéristique importante d'un cadre de responsabilité. Les personnes affectées négativement par une décision automatisée devraient pouvoir faire appel d'une décision et savoir comment le faire. Comme le précise la recommandation de l'OCDE de 2019 sur l'IA, les personnes lésées par un système d'IA devraient pouvoir contester le ou les résultats du système sur la base d'informations faciles à comprendre concernant les facteurs qui ont servi de base à la prédiction, à la recommandation ou à la décision (OCDE, 2019<sup>[59]</sup>).

Les plaintes et les enquêtes doivent être prises au sérieux et rendues publiques, ainsi que leurs résultats, afin que des enseignements puissent être tirés et partagés avec d'autres personnes effectuant des travaux similaires. Cependant, si l'intégration de l'humain dans la boucle des systèmes automatisés est un moyen essentiel de garantir la responsabilité et la surveillance, elle n'est pas sans poser de défis. Par exemple, quel niveau de surveillance, de responsabilité et d'obligation de rendre compte s'applique aux décisions prises par l'humain ? Quelles qualifications et/ou compétences sont nécessaires pour contester une décision automatisée ? Certains groupes peuvent ne pas savoir qu'ils ont été négligés ou ne pas disposer des ressources nécessaires pour traiter les problèmes (Lokshin et Umapathi, 2022<sup>[64]</sup> ; Barca et Chirchir, 2019<sup>[65]</sup>). Les procédures de plainte doivent tenir compte de cette situation et les organismes publics doivent veiller à ce que les populations marginalisées et exclues bénéficient d'un soutien pour introduire une demande de révision d'une décision.

Outre les garanties appropriées visant à assurer la responsabilité et l'équité, les principes de l'OCDE relatifs à l'IA mettent également l'accent sur les points suivants :

- La nécessité de garantir la sûreté et la sécurité en mettant en place des systèmes robustes capables de résister à des conditions défavorables telles que les risques liés à la sécurité numérique, et en veillant à ce que les systèmes d'IA puissent être remplacés ou mis hors service s'ils causent un préjudice injustifié ou présentent un comportement indésirable ; et
- La nécessité de renforcer les capacités des marchés du travail afin qu'ils puissent interagir avec les systèmes d'IA et s'adapter aux changements induits par l'IA sur les marchés du travail.

Le besoin de formation et de développement des compétences va de pair avec la nécessité de garantir la transparence et l'explicabilité. Le personnel qui est en contact avec les demandeurs de prestations sociales doit être en mesure d'expliquer comment une décision a été prise et de fournir des informations sur la manière dont cette décision peut être réexaminée. Cela nécessite que le personnel soit correctement formé et que des procédures de plainte suffisantes soient mises en place. En outre, les organismes publics devraient envisager de développer des algorithmes en interne en faisant appel à des experts internes et/ou de comprendre et être en mesure d'expliquer les algorithmes développés par des partenaires externes (OCDE, 2019<sup>[66]</sup>), et il peut également être nécessaire que les avocats, les juges ou autres arbitres reçoivent une formation sur le fonctionnement et la fiabilité des algorithmes afin de pouvoir répondre de manière appropriée à toute réclamation liée à des résultats défavorables (Citron, 2007<sup>[67]</sup> ; Gilman, 2020<sup>[68]</sup>).

Grâce à des politiques et à une gouvernance efficaces, l'IA pourrait contribuer à garantir que les systèmes de protection sociale soient bien préparés pour relever les défis d'aujourd'hui et de demain. Les mégatendances économiques et sociodémographiques vont amplifier et modifier les besoins en matière de soutien dans les années à venir, car le changement climatique, la numérisation et le vieillissement de la population vont accroître ces besoins tout en limitant les ressources disponibles pour y répondre (Frey et Hyee, 2024<sup>[69]</sup>). En contribuant à améliorer le ciblage, la fourniture et l'efficacité des services, l'IA a le potentiel d'apporter un soutien meilleur et plus rapide aux personnes qui en ont le plus besoin.

## Conclusions

L'utilisation de l'IA dans les systèmes de protection sociale peut améliorer la prestation des services, combler les lacunes en matière de couverture et libérer du temps pour les fonctionnaires, mais cette technologie présente également des risques importants, notamment en matière de confidentialité des données et de prise de décision automatisée. Les gouvernements de l'OCDE avancent prudemment dans ce domaine, compte tenu des enjeux importants, et les applications actuelles de l'IA se concentrent sur l'aide à la clientèle, l'automatisation des processus administratifs et la détection des fraudes.

Il existe des cas d'utilisation prometteurs. Il s'agit notamment d'utiliser l'analyse prédictive pour prévoir la demande et les chocs sociaux et identifier plus tôt les clients vulnérables, de réduire le non-recours grâce à des actions de sensibilisation ciblées et d'adapter les interventions aux besoins spécifiques de chaque individu afin d'atteindre les objectifs du programme. L'IA pourrait également contribuer à répondre aux préoccupations relatives aux biais ou aux incohérences dans les processus décisionnels (humains). Toutefois, pour tirer pleinement parti de ces avantages, il faudra investir dans les infrastructures numériques et évaluer avec soin quand et comment l'IA apporte une valeur ajoutée par rapport à d'autres outils.

Pour avancer de manière responsable, les gouvernements devraient continuer à tester des applications d'IA et à identifier les domaines dans lesquels l'IA est l'outil le plus approprié, car d'autres méthodes permettent encore d'obtenir des gains importants (par exemple, faciliter l'inscription grâce à la mise en relation de données administratives). Les gouvernements devraient établir une justification claire de l'utilisation de l'IA dans chaque contexte, fondée sur des objectifs politiques. Ils doivent également associer le public dès le début et de manière transparente. La confiance du public reste fragile : seuls 40 % des personnes interrogées dans 27 pays de l'OCDE estiment que le traitement des demandes à l'aide de l'IA est bénéfique pour les utilisateurs. Un engagement significatif, la sécurité des données, la transparence et l'accent mis sur l'équité aideront les gouvernements à renforcer la confiance dans les systèmes de protection sociale soutenus par l'IA à l'avenir.

## Références

- 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]
- Académie américaine de pédiatrie (2025), *Logiciel de surveillance du risque de suicide dans les écoles*, [https://www.aap.org/en/patient-care/media-and-children/center-of-excellence-on-social-media-and-youth-mental-health/ga-portal/ga-portal-library/ga-portal-library-questions/monitoring-software-for-suicide-risk-in-schools/?srsltid=AfmBOoo4I2IMc19OIP9npFh-Q7AVUI\\_3zCvaBdAxfNBmB\\_Pb9eIJ2d19](https://www.aap.org/en/patient-care/media-and-children/center-of-excellence-on-social-media-and-youth-mental-health/ga-portal/ga-portal-library/ga-portal-library-questions/monitoring-software-for-suicide-risk-in-schools/?srsltid=AfmBOoo4I2IMc19OIP9npFh-Q7AVUI_3zCvaBdAxfNBmB_Pb9eIJ2d19) (consulté le 19 juin 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 et P. Ben-Nun Bloom 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, p. 75-96, <https://doi.org/10.1093/jopart/muab016>. [53]
- Barca, V. et 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), *En espionnant les appareils des élèves, les écoles cherchent à intercepter les comportements autodestructeurs avant qu'ils ne se produisent*, <https://www.nytimes.com/2024/12/09/health/suicide-monitoring-software-schools.html>. [38]
- Bell, E. et 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), *Le scandale des allocations familiales aux Pays-Bas montre que nous avons besoin de règles explicables en matière d'IA*, Université d'Amsterdam, <https://www.uva.nl/en/shared-content/faculteiten/en/faculteit-der-rechtsgeleerdheid/news/2023/02/childcare-benefit-scandal-transparency.html?cb> (consulté en novembre 2023). [62]
- Booth, R. (2024), « Révélation : des biais ont été découverts dans le système d'IA utilisé pour détecter les fraudes aux prestations sociales au Royaume-Uni », *The Guardian*, <https://www.theguardian.com/society/2024/dec/06/revealed-bias-found-in-ai-system-used-to-detect-uk-benefits> (consulté le 13 décembre 2024). [51]
- Brioscu, A. et al. (2024), « Une nouvelle ère pour les services publics de l'emploi : la prestation de services à l'ère de l'intelligence artificielle », OCDE, 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/> (consulté le 8 janvier 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/> (consulté le 19 juin 2025). [46]



- Carter, E. et 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 – Plateforme en ligne de données d'observation de la Terre*, <https://ceadar.ie/blog/cameo-online-platform-of-earth-observation-data/> (consulté le 1er avril 2025). [24]
- Centre for Economic Performance (2020), *En bref... Lutter contre la violence domestique grâce à l'apprentissage automatique*, <https://cep.lse.ac.uk/pubs/download/cp579.pdf> (consulté le 13 décembre 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), *site web CompetentNL*, <https://competentnl.nl/>. [45]
- Desiere, S., K. Langenbucher et L. Struyven (2019), « Statistical profiling in public employment services: An international comparison », *Documents de travail de l'OCDE sur les affaires sociales, l'emploi et les migrations*, n° 224, Éditions OCDE, 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), *Comment les données peuvent-elles empêcher le sans-abrisme avant qu'il ne commence ?*, Ernst & Young, [https://www.ey.com/en\\_uk/insights/government-public-sector/how-can-data-stop-homelessness-before-it-starts](https://www.ey.com/en_uk/insights/government-public-sector/how-can-data-stop-homelessness-before-it-starts) (consulté le 13 décembre 2024). [30]
- Bureau exécutif du président (2020), *Décret 13960 : Promouvoir l'utilisation d'une intelligence artificielle fiable au sein du gouvernement fédéral*, <https://www.federalregister.gov/d/2020-27065>. [6]
- Frey, V. et R. Hyee (2024), *Non-recours et transformation numérique des programmes sociaux dans les pays de l'OCDE*, Éditions OCDE, <https://doi.org/10.1787/af31746d-en>. [25]
- Frey, V. et R. Hyee (2024), *Aperçu : mégatendances et avenir de la protection sociale*, OCDE, <https://doi.org/10.1787/6c9202e8-en>. [69]
- Frey, V., R. Hyee et P. Minondo Canto (2024), *Cadres nationaux pour identifier les bénéficiaires potentiels et les intégrer dans la protection sociale*, Éditions OCDE, <https://doi.org/10.1787/af31746d-en>. [42]
- Gillingham, P. (2019), « Les algorithmes prédictifs peuvent-ils aider à la prise de décision dans le domaine du travail social auprès des enfants et des familles ? », *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), *Prévenir les accidents grâce à l'IA*. [22]
- Gouvernement écossais (2024), *Accroître la transparence de l'IA*, <https://www.gov.scot/news/increasing-ai-transparency/>. [13]



- Gouvernement des Pays-Bas (2022), *Lettre au Parlement sur le dispositif de détection des fraudes et la question du racisme institutionnel*, <https://www.rijksoverheid.nl/documenten/kamerstukken/2022/05/30/kamerbrief-reactie-op-verzoeken-over-fraudesignaleringsvoorziening> (consulté le 8 janvier 2025). [61]
- Grogger, J., R. Ivandic et 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> (consulté le 20 décembre 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 et T. Jackson (2024), « Everyone must understand the environmental costs of AI » (Tout le monde doit comprendre les coûts environnementaux de l'IA), <https://oecd.ai/en/work/understand-environmental-costs> (consulté le 20 janvier 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 », *document de travail IZA n° 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), *L'intelligence artificielle dans les institutions de sécurité sociale : le cas des chatbots intelligents*. [16]
- AISS (2022), *L'innovation fondée sur les données dans la sécurité sociale : bonnes pratiques en Asie et dans le Pacifique*. [21]
- ISSA (2020), *L'intelligence artificielle dans la sécurité sociale : contexte et expériences*. [19]
- Université de Lancaster (2021), *Développement d'un nouveau système d'alerte précoce basé sur l'IA pour les inondations*. [23]
- Lara-Montero, A. (2024), *Le potentiel de transformation de l'IA dans les services sociaux*, European Public Mosaic, <https://eapc.gencat.cat/ca/publicacions/epum>. [
- Association des collectivités locales (2025), *Rapport d'étape sur le défi entre pairs : arrondissement londonien d'Ealing*, [https://www.ealing.gov.uk/download/downloads/id/20338/review\\_of\\_ealings\\_adult\\_social\\_care\\_-\\_january\\_2025.pdf](https://www.ealing.gov.uk/download/downloads/id/20338/review_of_ealings_adult_social_care_-_january_2025.pdf) (consulté le 19 juin 2025). [47]
- Lokshin, M. et 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), *L'équipe chargée du logement du MBC remporte un prix national pour la prévention sans-abrisme*, <https://news.maidstone.gov.uk/home/news/2022/november/mbc-housing-team-wins-national-award-for-preventing-homelessness> (consulté le 13 décembre 2024). [29]

- Marr, B. (2018), *Comment le gouvernement britannique utilise l'intelligence artificielle pour détecter les fraudes aux prestations sociales et aux aides de l'État*, Forbes. [20]
- NHS England (2025), *Déploiement national d'un outil d'intelligence artificielle qui prédit les chutes et les virus*, <https://www.england.nhs.uk/2025/03/nationwide-roll-out-of-artificial-intelligence-tool-that-predicts-falls-and-viruses/> (consulté le 1er avril 2025). [1]
- Ntoutsis, 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]
- OCDE (2025), *Une protection sociale plus efficace pour une croissance économique plus forte : principales conclusions de l'enquête 2024 de l'OCDE sur les risques qui comptent*, Éditions OCDE, Paris, <https://doi.org/10.1787/3947946a-en>. [10]
- OCDE (2025), *Gagner du temps, améliorer l'accompagnement : comment l'IA aide les travailleurs sociaux à consacrer plus de temps aux personnes dont ils s'occupent*, OCDE, <https://www.oecd.org/en/about/projects/ai-and-the-future-of-social-protection.html>. [48]
- OCDE (2024), *L'IA dans le domaine de la santé : un potentiel énorme, des risques considérables*, OCDE, [https://www.oecd.org/en/publications/2024/01/ai-in-health-huge-potential-huge-risks\\_ff823a24.html](https://www.oecd.org/en/publications/2024/01/ai-in-health-huge-potential-huge-risks_ff823a24.html). [3]
- OCDE (2024), « Note explicative sur la définition actualisée de l'OCDE d'un système d'IA », *Documents de l'OCDE sur l'intelligence artificielle*, n° 8, Éditions OCDE, Paris, <https://doi.org/10.1787/623da898-en>. [7]
- OCDE (2024), « Mémoire explicatif sur la définition actualisée de l'OCDE d'un système d'IA », *Documents de l'OCDE sur l'IA*, 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](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]
- OCDE (2024), *Investir dans la jeunesse : région du Danemark du Nord*, Investir dans la jeunesse, Éditions OCDE, Paris, <https://doi.org/10.1787/0856ce12-en>. [40]
- OCDE (2024), *Mégatendances et avenir de la protection sociale*, Éditions OCDE, Paris, <https://doi.org/10.1787/6c9202e8-en>. [1]
- OCDE (2024), *Moderniser l'accès à la protection sociale : stratégies, technologies et progrès en matière de données dans les pays de l'OCDE*, OCDE, <https://doi.org/10.1787/af31746d-en>. [1]
- OCDE (2024), *Enquête de l'OCDE sur les facteurs de confiance dans les institutions publiques – Résultats 2024 : Renforcer la confiance dans un environnement politique complexe*, Éditions OCDE, Paris, <https://doi.org/10.1787/9a20554b-en>. [57]
- OCDE (2023), « Modèles linguistiques basés sur l'IA : considérations technologiques, socio-économiques et politiques », *Documents de l'OCDE sur l'économie numérique*, n° 352, Éditions OCDE, Paris, <https://doi.org/10.1787/13d38f92-en>. [15]
- OCDE (2023), « L'état de la mise en œuvre des principes de l'OCDE sur l'IA quatre ans après », *Documents de l'OCDE sur l'intelligence artificielle*, n° 3, Éditions OCDE, Paris, <https://doi.org/10.1787/835641c9-en>. [4]

- OCDE (2022), « Cadre de l'OCDE pour la classification des systèmes d'IA », *Documents de l'OCDE sur l'économie numérique*, n° 323, Éditions OCDE, Paris, <https://doi.org/10.1787/cb6d9eca-en>. [63]
- OCDE (2019), *Exploiter les nouvelles données sociales pour des politiques sociales et des services efficaces*, OCDE, Paris, <https://www.oecd.org/els/soc/Workshop-NewSocialData-16Oct2019-BackgroundNote.pdf>. [66]
- OCDE (2019), *Recommandation du Conseil sur l'intelligence artificielle*, OCDE, Paris, <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>. [59]
- Ohlenburg, T. (2020), *L'IA dans la protection sociale - Explorer les opportunités et atténuer les risques*, GIZ ; ADB, Bonn, <https://www.adb.org/publications/ai-social-protection-exploring-opportunities-mitigating-risks> (consulté le 13 décembre 2024). [12]
- Parikh, R., S. Teeple et 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), *site web Public Sector Tech Watch*, <https://interoperable-europe.ec.europa.eu/collection/public-sector-tech-watch>. [17]
- Robila, M. et S. Robila (2020), « Applications of Artificial Intelligence Methodologies to Behavioral and Social Sciences », *Journal of Child and Family Studies*, vol. 29/10, p. 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\(ca|en\)](https://eapc.gencat.cat/ca/detalls/Noticia/EPum_23_noticia#googtrans(ca|en)) (consulté le 13 décembre 2024). [28]
- Parlement européen (2022), *Le scandale des allocations familiales aux Pays-Bas, racisme institutionnel et algorithmes*, [https://www.europarl.europa.eu/doceo/document/O-9-2022-000028\\_EN.html](https://www.europarl.europa.eu/doceo/document/O-9-2022-000028_EN.html) (consulté en novembre 2023). [60]
- Administration américaine pour les enfants et les familles (2024), *Portail d'information sur la protection de l'enfance*, <https://www.childwelfare.gov/stay-connected/contact/>. [18]
- Verhagen, A. (2024), « Utiliser l'IA pour gérer les prestations de revenu minimum et l'aide au chômage : opportunités, risques et orientations politiques possibles », *Documents de l'OCDE sur l'intelligence artificielle*, n° 21, Éditions OCDE, Paris, <https://doi.org/10.1787/718c93a1-en>. [58]

## Notes

<sup>1</sup> La section 5(e) précise : « Dans les 120 jours suivant l'achèvement de leurs inventaires, les agences mettent ceux-ci à la disposition du public, dans la mesure du possible et conformément à la législation et aux politiques applicables, y compris celles concernant la protection de la vie privée et des informations sensibles relatives à l'application de la loi, à la sécurité nationale et à d'autres informations protégées » (Bureau exécutif du président, 2020[6]).

<sup>2</sup>L'OCDE a récemment clarifié la définition d'un système d'IA contenue dans la Recommandation de l'OCDE de 2019 sur l'IA (les « Principes de l'IA ») afin de garantir sa pertinence. La définition mise à jour suivante a été adoptée par le Conseil de l'OCDE le 8 novembre 2023 : « Un système d'IA est un système mécanique qui, pour un ensemble donné d'objectifs explicites ou implicites définis par l'homme, est capable de déduire, à partir des données qui lui sont fournies, comment générer des résultats tels que des prévisions, des contenus, des recommandations ou des décisions susceptibles d'influencer des environnements physiques réels ou virtuels. Les différents systèmes d'IA sont conçus pour fonctionner avec des niveaux d'autonomie et d'adaptabilité variables après leur déploiement. »

<sup>3</sup>L'apprentissage automatique (ou apprentissage automatique) désigne l'ensemble des techniques permettant aux machines d'apprendre de manière automatisée à partir de modèles et de déductions plutôt que d'instructions explicites données par un être humain. Les approches d'apprentissage automatique consistent souvent à enseigner aux machines comment atteindre un résultat en leur montrant de nombreux exemples de résultats corrects. L'apprentissage automatique comprend de nombreuses techniques utilisées depuis des décennies par les économistes, les chercheurs et les technologues. Celles-ci vont des régressions linéaires et logistiques, des arbres de décision et de l'analyse en composantes principales aux réseaux neuronaux profonds (OCDE, 2022[63]).

<sup>4</sup>Les signes avant-coureurs du sans-abrisme chez les personnes vivant dans une grande pauvreté comprenaient « une forte augmentation des demandes de services  
utilisation, recours de plus en plus fréquent aux services et réception de multiples services auprès d'un seul organisme.

<sup>5</sup>Moyenne de 13 pays. Seuls les pays comptant un nombre suffisamment important de répondants s'identifiant comme appartenant à une minorité en raison de leur origine ethnique ou de la couleur de leur peau ont été pris en compte dans la moyenne (taille des cellules supérieure à 10 observations pour les numérateurs, les dénominateurs et les résidus).