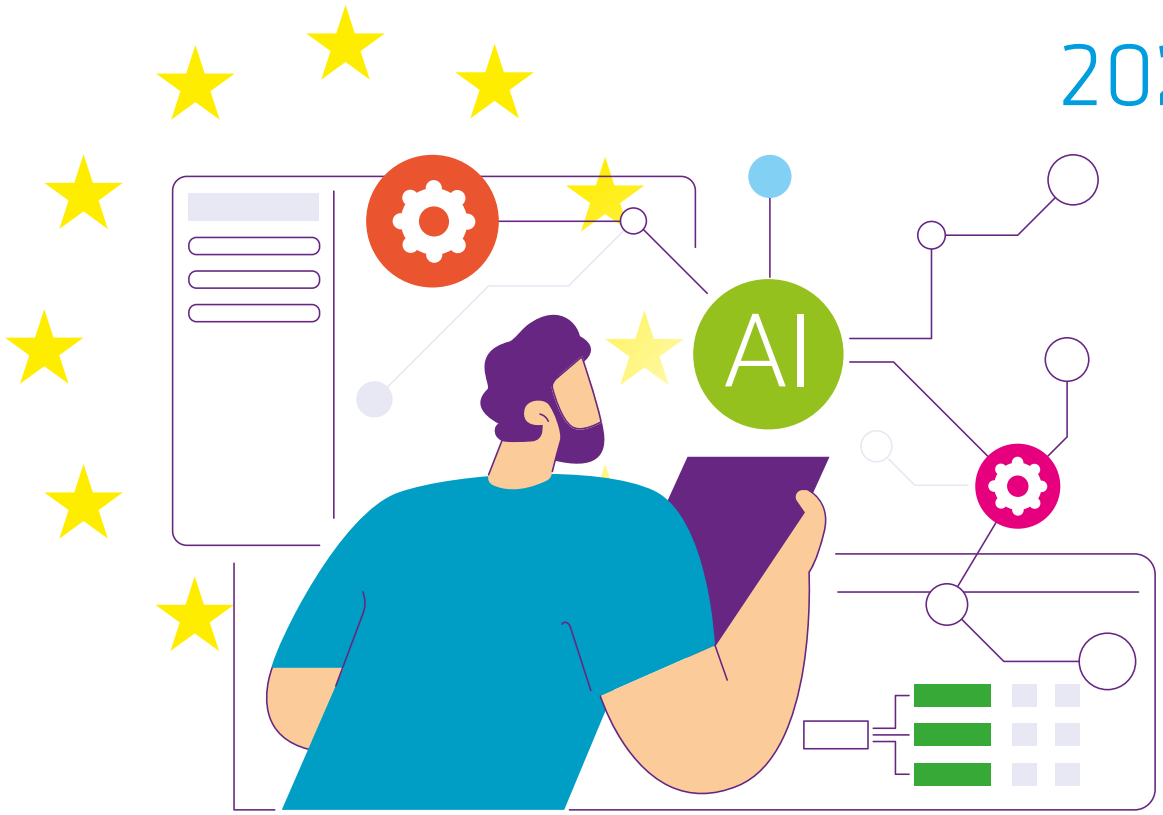
2025



MYTHES ET ENJEUX DE L'UTILISATION DE L'INTELLIGENCE ARTIFICIELLE DANS LES COLLECTIVITÉS TERRITORIALES

POINTS D'ATTENTION POUR UNE IA RESPONSABLE

Les interconnectés
RÉSEAU DES TERRITOIRES INNOVANTS

FRANCE URBAINE
MÉTROPOLES, AGGLOS ET GRANDES VILLES

INTERCOMMUNALITÉS DE FRANCE

TABLE DES MATIÈRES

INTRODUCTION

Les collectivités territoriales utilisent de plus en plus fréquemment des outils s'appuyant sur l'intelligence artificielle pour prendre des décisions ou orienter des politiques publiques, dans des domaines allant de l'urbanisme à la relation aux usagers en passant par la tarification solidaire et à l'attribution des places en crèche. Le récent engouement autour de l'intelligence artificielle générative a contribué à intensifier et populariser le phénomène. Selon le baromètre de l'Observatoire Data Publica, le nombre de projets d'intelligence artificielle est passé d'une cinquantaine, majoritairement dans des grandes collectivités, en 2023, à plusieurs centaines dans des collectivités de toutes tailles en 2024[1].
Un rapport d'élèves de l'INET a estimé qu'au moins 25% des tâches des métiers d'une collectivité témoin pouvait être effectuée au moins partiellement par l'IA générative[2].

Les techniques d'intelligence artificielle (IA) promettent d'améliorer le service public. Mais, dans ce contexte de course à l'IA, les collectivités sont souvent confrontées à un flou : comment évaluer les promesses alléchantes de certains outils ? Comment s'assurer que leur acquisition et leur usage sont responsables et maîtrisés ? Autrement dit : l'IA, oui, mais pour quoi faire ?

Les systèmes d'intelligence artificielle sont des outils de mise en œuvre des politiques publiques.
À ce titre, ils doivent répondre aux mêmes principes d'égalité de traitement, d'accessibilité et de responsabilité que le reste des actions menées par les collectivités[3].

Dans ce cadre, l'objectif de ce rapport est de donner aux agents publics les outils pour déconstruire les mythes et comprendre les grands enjeux soulevés par les algorithmes et l'intelligence artificielle pour les services publics[4], et en permettre une appropriation responsable et éclairée. Il n'est pas conçu comme un guide des «pour» et des «contre» de l'IA, mais comme une exploration des points d'attention à garder en tête avant et pendant l'utilisation de ces outils dans nos territoires.

Il s'adresse aux profils aussi bien techniques (membres des directions des systèmes d'information) que métiers (et notamment les directions générales des services). Pour se faire, il s'appuie sur la littérature scientifique et sur des études de cas existantes, en France et à l'étranger.

Bien sûr, il y a "IA" et "IA". Le terme recouvre différents types de technologies et d'usages. Comment comparer un outil pour contrôler les bénéficiaires des prestations sociales et un système pour modéliser et prévenir l'impact des phénomènes naturels sur la gestion de l'eau ? En réalité, on peut appréhender les outils d'IA dans un continuum d'usages, de conséquences et de risques. Les grands enjeux soulevés dans ce rapport seront plus ou moins applicables selon les cas d'usages.

Ce rapport né de l'initiative de la Commission numérique commune des Interconnectés, France urbaine et Intercommunalités de France vise à nourrir un des axes de travail de sa feuille de route «stratégie des intelligences associées».
La rédaction en a été confiée à Soizic Pénicaud, cofondatrice de l'Observatoire des algorithmes publics et enseignante à Sciences Po Paris.

Il commence par replacer les algorithmes et l'intelligence artificielle dans leur contexte (section 1). Il se penche ensuite sur quatre mythes à déconstruire quand on parle d'IA : l'objectivité, la performance, le contrôle humain et le potentiel bénéfique pour les agents (section 2). Il balaye enfin trois grandes catégories d'enjeux supplémentaires : les enjeux environnementaux et sociétaux, ceux de cybersécurité et de protection des données et ceux de transparence, de redevabilité et de maîtrise (section 3).

1 *Observatoire Data Publica. (2024). Baromètre de l'Observatoire Data Publica 2024, 3e édition : les collectivités territoriales et la donnée.*
2 *Demoures, C. (2024). Un outil de cartographie des métiers concernés par l'intelligence artificielle dans les collectivités. Une étude des élèves de l'INET 2023-2024.*
3 *Vie-publique.fr. (2 janvier 2025). La notion de service public.*
4 *À noter que le Défenseur des droits a publié en novembre un rapport détaillé sur les enjeux des algorithmes et de l'intelligence artificielle pour les droits des usagers. Voir Défenseur des droits. (2024). Algorithmes, systèmes d'IA et services publics : quels droits pour les usagers ? Points de vigilance et recommandations.*

1. CONTEXTE
ALGORITHMES ET INTELLIGENCE ARTIFICIELLE DANS LE SECTEUR PUBLIC, DE QUOI PARLE-T-ON ?

Le terme «algorithme» ne fait l'objet d'aucune définition légale, mais la CNIL[5] et la CADA[6] l'ont défini comme «la description d'une suite finie et non ambiguë d'étapes (ou d'instructions) permettant d'obtenir un résultat à partir d'éléments fournis en entrée»[7]. Cette description, très large, englobe de nombreux types d'automatisation dans le secteur public.

Les algorithmes ne sont pas chose nouvelle dans l'administration : la numérisation du système des impôts et de la sécurité sociale commence dès les années 1960. Mais de nouvelles techniques ont progressivement été développées, qui permettent de nouveaux usages.

A. ALGORITHMES DÉTERMINISTES ET ALGORITHMES APPRENANTS

Schématiquement, on distingue deux types de systèmes[8] :

Les systèmes déterministes : pour ces systèmes, les règles sont définies par des humains, et ensuite encodées dans des programmes informatiques. C'est le cas par exemple du calcul des impôts, ou des algorithmes de Parcoursup. Un tableur avec des règles de calcul (type Excel ou LibreOffice) peut être considéré comme un système déterministe.

Les systèmes apprenants (ou probabilistes), qui s'appuient sur des techniques d'apprentissage automatique (le fameux machine learning). Ici, les règles ne sont pas définies par des humains. Les modèles algorithmiques déduisent eux-mêmes les règles en s'appuyant sur les données. Mais les concepteurs du système jouent quand-même un rôle, en fournissant aux modèles algorithmiques des «données d'entraînement» et (dans le cas de l'apprentissage supervisé) le résultat attendu. Ces nouvelles techniques permettent de nouveaux usages, qui vont du traitement automatique du langage (par exemple, pour analyser le contenu de documents), à la vision par ordinateur (computer vision, pour analyser des images satellites ou de caméras de sécurité), à des modèles prédictifs (par exemple, pour attribuer des scores de risque ou prédire des tendances).

Au sein des systèmes probabilistes, on trouve des sous-catégories, comme **l'apprentissage profond** (deep learning) et **l'intelligence artificielle générative,** qui est issue du deep learning. Ces nouvelles techniques permettent de générer des images, du texte, ou du son.

Le terme «intelligence artificielle» ne dispose pas non plus de définition universellement arrêtée[9] (voir encadré), et peut être utilisé de manière plus ou moins large[10]. Certaines personnes restreignent l'intelligence artificielle aux systèmes apprenants, et en excluent les systèmes déterministes. Or, beaucoup des algorithmes utilisés dans les collectivités restent relativement simples techniquement et déterministes. Par ailleurs, les systèmes font souvent appel à une combinaison de techniques. C'est pour cela qu'il faut considérer les algorithmes au sens large, tout en ayant conscience des spécificités de différentes technologies.

5 Commission nationale de l'informatique et des libertés, le régulateur français des données personnelles.
6 Commission d'accès aux documents administratifs l'autorité administrative indépendante chargée de veiller à la liberté d'accès aux documents administratifs et aux archives publiques ainsi qu'à la réutilisation des informations publiques.
7 CNIL, CADA, Etalab. (2019). Guide pratique de la publication en ligne et de la réutilisation des données publiques (open data). p.26.
8 CNIL. (2017). Comment permettre à l'Homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle. p.20.
9 D'un point de vue légal et institutionnel, le règlement européen sur l'intelligence artificielle du 13 juin 2024 (AI Act) définit le terme «système d'IA» en s'appuyant sur une définition de l'OCDE.
10 Ce rapport utilise les termes «intelligence artificielle» et «algorithme» de manière interchangeable.

Intelligence artificielle : quelles définitions légales et institutionnelles ?

Le règlement européen sur l'intelligence artificielle du 13 juin 2024 (AI Act) définit le terme "système d'IA" en s'appuyant sur une *définition* de l'OCDE :
« Un système automatisé qui est conçu pour fonctionner à différents niveaux d'autonomie et peut faire preuve d'une capacité d'adaptation après son déploiement, et qui, pour des objectifs explicites ou implicites, déduit, à partir des entrées qu'il reçoit, la manière de générer des sorties telles que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels » (Article 3(1))

Les débats restent ouverts sur le périmètre exact. La Commission européenne a publié une première version de *lignes directrices* sur l'interprétation de la définition, mais seules des décisions de justice permettront de trancher ce qui relève de cette définition ou non.

Par ailleurs, la loi n'aborde pas les algorithmes seulement sous le prisme de l'IA. Le Code des relations entre le public et l'administration parle notamment de « traitement algorithmique » et le droit de la protection des données à caractère personnel de « traitement de données à caractère personnel », dont les « traitements automatisés ».

B. LES ALGORITHMES DANS L'ADMINISTRATION

L'administration utilise les algorithmes pour différents usages, en mobilisant différentes techniques. Ces usages se sont déployés par phase.

USAGE / PHASE	EXEMPLES
1. Calculer Dans cette première phase, il s'agit d'effectuer les mêmes tâches que celles qui peuvent être effectuées à la main, mais avec moins d'erreurs et plus rapidement.	Calcul des impôts, des prestations sociales, tarification solidaire des transports
2. Apparier une « offre » et une « demande » Dès 1999, le ministère de l'Éducation nationale développe un algorithme pour gérer l'allocation des postes des enseignants. Dans ce cas, il ne s'agit pas uniquement de faire passer à l'échelle des calculs, mais de permettre des tâches d'optimisation difficilement réalisables par les humains.	Gestion de la mobilité des agents, accès à l'enseignement supérieur (Parcoursup), attribution de greffons cardiaques (Score Cœur)
3. Prédire Les techniques d'apprentissage automatique susmentionnées et la possibilité de traiter des données à grande échelle marquent un tournant pour les prédictions.	Gestion de l'eau, de l'électricité, cibler les actions de lutte contre la fraude, identifier des situations sur des images de vidéosurveillance
4. Générer des contenus L'intelligence artificielle générative étend les usages possibles des algorithmes dans l'administration.	Robots conversationnels, synthèse d'informations
5. Aider à la décision des usagers Un ensemble de techniques est utilisé pour construire des outils de simulation ou de prédiction.	Aider les demandeurs d'emploi à cibler leurs candidatures spontanées (La Bonne Boîte) Simuler le coût d'une embauche

Tableau adapté du Guide d'Etalab

2. MYTHES DE L'INTELLIGENCE ARTIFICIELLE

Une fois ce contexte posé, prenons le temps de nous pencher sur quatre mythes qui entourent l'utilisation de systèmes d'intelligence artificielle.

A. L'IA, NEUTRE ET OBJECTIVE ? IA, BIAIS ET CHOIX

On pourrait avoir l'impression que les outils d'intelligence artificielle sont neutres et objectifs, parce qu'ils sont ancrés dans l'informatique, les statistiques, et les mathématiques. En réalité, la subjectivité humaine, voire les biais[11], sont toujours présents, mais sont déplacés.

1. Les biais dans les algorithmes apprenants

Comme dit précédemment (voir la section «Contexte»), les algorithmes apprenants (apprentissage automatique, apprentissage profond, IA générative) sont entraînés sur des jeux de données pour repérer des motifs récurrents, appelés patterns en anglais. La qualité des données d'entraînement va avoir des conséquences importantes sur celle des résultats. C'est le fameux adage «garbage in, garbage out»[12]. Mais il ne suffit pas que les données soient de bonne qualité pour que les algorithmes ne soient pas biaisés. Le choix des éléments mesurés et la manière de les mesurer sont aussi déterminants.

Il existe de nombreux types de biais qui peuvent s'introduire sous le capot des algorithmes. **Nous en citerons ici trois**[13].

Tout d'abord, les biais historiques, qui reflètent les biais de la société. En 2020, un algorithme utilisé au Royaume-Uni pour prédire les notes du baccalauréat s'appuyait sur les résultats passés des lycées[14]. Ce faisant, il a pénalisé les bons étudiants venant de lycées historiquement moins performants (et donc souvent des étudiants de milieux socio-économiques défavorisés), dont la note était plafonnée par celle des meilleurs étudiants des années précédentes.

Mais les données peuvent aussi représenter le réel de manière inexacte, en surreprésentant ou sous-représentant certaines populations (biais par omission)[15]. Par exemple, en 2016, une équipe de recherche menée par la chercheuse Joy Buolamwini a montré que les outils de reconnaissance faciale étaient moins performants sur les visages des femmes noires que sur ceux des personnes blanches, du fait du manque de données dans les jeux d'entraînement[16]. À l'inverse, la chercheuse Virginia Eubanks a souligné qu'un logiciel utilisé par les services de protection de l'enfance d'un état américain avait tendance à surcontrôler les populations précaires, sur lesquelles l'administration détenait beaucoup plus de données que les populations plus aisées, chez lesquelles la maltraitance infantile était donc plus difficile à déceler[17].

Les biais peuvent également se situer dans la manière dont un phénomène est mesuré (biais de proxy). Par exemple, pour prédire le nombre de crimes dans une zone donnée, les algorithmes de police prédictive vont en réalité se baser sur celui des arrestations, qui peuvent elles-mêmes être biaisées : certains crimes ne seront jamais découverts (par exemple : dans certaines zones moins contrôlées), et donc jamais enregistrés comme tels, ce qui pourra mener à leur sous-détection par un algorithme.

Les affaires d'algorithmes publics ont montré que les biais ont tendance à pénaliser les populations qui souffrent déjà de discriminations : femmes, personnes à faibles revenus, minorités[18].

11 Le terme biais peut porter à confusion, car c'est aussi un terme statistique (un biais statistique se réfère à une déviation par rapport à un standard). Ici, on utilise son acception «grand public», pour désigner les effets potentiellement discriminatoires ou délétères des algorithmes.
12 «Des déchets en entrée, des déchets en sortie».
13 Voir Leufer, D. (2020). Myth: AI can be objective or unbiased. AI Myths. pour un article plus détaillé.
14 Biret, V. (18 août 2020). Mal notés au bac par un algorithme, les lycéens britanniques crient au scandale. Ouest-France.
15 Philippe Besse parle de «taux d'erreur de prévision». Voir Vallet, F. (2 juin 2020). Philippe Besse : «Les décisions algorithmiques ne sont pas plus objectives que les décisions humaines». Linc - CNIL.
16 Gender Shades. How well do IBM, Microsoft, and Face++ AI services guess the gender of a face?
17 Eubanks, V. (16 février 2018). A response to Allegheny County DHS. Pour une chronique en français de l'ouvrage Automating Inequality, voir InternetActu. (19 janvier 2018). Automatiser les inégalités.
18 Voir les nombreux exemples en matière de protection sociale, au Danemark, Pays-Bas, et en France.

La difficulté est que, dans les algorithmes apprenants, ces biais peuvent être difficiles à déceler et corriger, car les paramètres utilisés par les algorithmes pour générer leurs résultats restent opaques et sont donc difficilement auditables.

Focus sur... L'analyse des CV

Un des exemples les plus connus d'algorithmes biais est celui du projet de logiciel de recrutement d'Amazon : en 2014, une équipe d'Amazon développe un algorithme pour automatiser le recrutement dans l'entreprise, en analysant les CV. L'outil est entraîné sur les CV des personnes qu'Amazon a déjà recrutées. Un an plus tard, le projet est mis à la poubelle : l'algorithme se révèle systématiquement discriminant envers les femmes pour des positions techniques (ingénieures en informatique, par exemple). En effet, l'entreprise avait recruté beaucoup d'hommes dans le passé, et le logiciel avait donc déterminé que les hommes étaient meilleurs... Ici, l'algorithme n'est que le reflet des pratiques de recrutement de l'entreprise.
Cette question de l'analyse des CV est renouvelée avec l'IA générative. Une étude récente de l'UNESCO a montré que les larges modèles de langage les plus connus, tels que GPT, contiennent des biais «indéniables» contre les femmes et les personnes homosexuelles[19]. Les journalistes de Bloomberg ont également montré que ChatGPT analysait les CV avec des biais racistes[20].

2. Les choix subjectifs dans les algorithmes déterministes

Pas de données d'entraînement, pas de problème ? Comme expliqué précédemment, les algorithmes déterministes ne sont pas entraînés sur des données : ils exécutent des règles et des calculs décidés directement par des humains, parfois à l'aide de techniques statistiques. Cela n'empêche pas ces algorithmes d'encoder des choix.

Focus sur... Parcoursup

Dans le cadre de la procédure Parcoursup d'accès à l'enseignement supérieur, les établissements d'enseignement supérieur utilisent souvent un algorithme pour classer les dossiers de candidature. Ces algorithmes ne sont pas entraînés sur des données, mais paramétrés par des humains, dans la limite des critères permis par la loi. Le ministère de l'Éducation nationale met à la disposition des établissements un algorithme (OAD, outil d'aide à la décision), qui se fonde sur les notes des candidats. Les établissements étant autorisés, dans une certaine mesure, à utiliser d'autres critères, Sciences Po Bordeaux a développé son propre algorithme de sélection des dossiers qui se fonde sur l'écart entre les notes de l'élève et la moyenne de la classe, et est paramétré pour favoriser les élèves boursiers et ceux issus des lycées participant aux «cordées de la réussite». Les deux systèmes mènent à des résultats drastiquement différents : en 2023, seulement 3% des élèves venant d'une cordée de la réussite sont considérés comme admissibles par l'OAD du ministère, contre 19% quand c'est l'algorithme de Sciences Po Bordeaux qui est utilisé. Comme le souligne le Défenseur des droits, «il y a derrière cet algorithme comme pour d'autres un choix politique (et non pas seulement technique) qui est fait par des humains, des décisions qui sont prises, voire des effets qui sont anticipés»[21].

Tout algorithme est ainsi le produit de décisions. Il est donc impossible de «débiaiser» les algorithmes, c'est-à-dire de les rendre objectifs et neutres. Comme le dit la chercheuse Cathy O'Neil : «les algorithmes sont des opinions intégrées dans du code»[22]. La question est plutôt : comment identifier là où les choix sont faits et quelles conséquences peuvent en découler, notamment en termes de discrimination et d'égalité d'accès aux services publics ?

19 *UNESCO, IRCAI. (2024). Challenging systematic prejudices: an Investigation into Gender Bias in Large. Language Models.*
20 *Walker, B. (2024, 8 mars). OpenAI's GPT Shows Bias in Résumé Ranking Experiment. Bloomberg.*
21 *Défenseur des droits. (2024). Algorithmes, systèmes d'IA et services publics : quels droits pour les usagers ? Points de vigilance et recommandations. pp.16-17*
22 *O'Neil, C. (2018). Algorithmes : la bombe à retardement. Les Arènes.*

3. « Les humains aussi sont biaisés ! » Oui, mais…

Reconnaître que les algorithmes sont biaisés ne veut pas dire qu'il faut idéaliser les décisions humaines. Celles-ci sont également marquées par des subjectivités et des biais discriminatoires.

La différence est que, quand il s'agit d'algorithmes, les biais sont centralisés et systématisés. On passe d'un pouvoir discrétionnaire au niveau des agents à un pouvoir discrétionnaire au niveau du système, qui défavorise ou favorise systématiquement les mêmes situations[23]. Cela a plusieurs conséquences. D'une part, alors que les agents prennent leurs décisions en engageant leur responsabilité et avec une vision du service public, la responsabilité devient plus floue quand c'est en apparence le système qui décide. D'autre part, la centralisation des critères de décision fait que ce sont toujours les mêmes populations qui vont être affectées. Elle peut aussi avoir des conséquences systémiques, notamment sur les politiques de la ville. Par exemple, l'algorithme de GPS de Waze peut troubler la gestion de la circulation dans les villes[24]. Par ailleurs, les agents ont de moins en moins d'espace pour exercer leur pouvoir d'appréciation[25], ce qui rend de plus en plus difficile la prise en compte ou la réactivité à des situations qui ne rentrent pas dans les cases.

Enfin, comme nous le verrons plus loin, les biais humains peuvent persister même dans le cas où les décisions sont assistées par un algorithme (voir section C : « Le contrôle humain, une garantie ? »).

4. Les outils de détection de biais sont insuffisants et pas standardisés

De plus en plus d'outils sont développés pour identifier et corriger les biais techniques des algorithmes. Cependant, ils souffrent encore d'un manque de standardisation. La chercheuse en droit Sandra Wachter avance que, parmi les 20 mesures de biais utilisés en science des données, 13 ne permettent pas de respecter les standards du droit de la non-discrimination européenne, notamment en matière de discrimination indirecte[26].

Une autre étude par le World Privacy Project a montré que les outils de débiaisage et d'explicabilité des données les plus fréquemment utilisés et recommandés s'appuient sur des mesures non reconnues par le monde scientifique, notamment la règle des 4/5e, ancrée dans un contexte et une culture étasuniens, qui s'exporte souvent dans des outils ensuite utilisés en Europe (notamment dans le recrutement)[27], ou bien les outils tels que LIME et SHAP[28]. Ces derniers sont des outils de data science fréquemment utilisés pour expliquer des modèles complexes. Cependant, la communauté scientifique alerte sur le fait qu'ils sont parfois utilisés en-dehors de leur périmètre initial, et leur efficacité est contestable sur des systèmes complexes.

L'utilisation d'outils de détection de biais est donc utile, mais ne peut pas être considérée comme une garantie suffisante à l'absence de caractère discriminatoire des algorithmes.

5. Statistiquement correct, donc acceptable ?

La focalisation à outrance sur les biais entraîne un autre effet : celui de se concentrer sur les aspects techniques du modèle, et non les politiques publiques en jeu. Comme le disent Julia Powles et Helen Nissenbaum : « les biais sont une diversion séduisante »[29].

Reprenons l'exemple de la reconnaissance faciale : suite aux études Gender Shades, les entreprises ont considérablement amélioré leurs algorithmes de reconnaissance faciale, qui reconnaissent désormais mieux les visages noirs ou féminins. Ces algorithmes sont désormais moins biaisés techniquement. Mais l'amélioration des

23 Bovens, M. et Zouridis, S. (2002). From Street-Level to System-Level Bureaucracies: How Information and Communication Technology Is Transforming Administrative Discretion and Constitutional Control. Public Administration Review.
24 Courmont, A. (2018). Plateforme, big data et recomposition du gouvernement urbain : les effets de Waze sur les politiques de régulation du trafic. Revue française de sociologie, 59-3. Pp. 423-449.
25 Vie-publique.fr. (25 juillet 2024). Compétence liée, pouvoir discrétionnaire ou d'appréciation, comment agit l'administration ?
26 Wachter, S., Mittelstadt, B., Russel, C. (2021). Bias Preservation in Machine Learning: The Legality of Fairness Metrics under EU Non-Discrimination Law. West Virginia Law Review, Vol. 123, No. 3.
27 Sánchez-Monedero, J., Dencik, L., & Edwards, L. (2020). What does it mean to 'solve' the problem of discrimination in hiring? social, technical and legal perspectives from the UK on automated hiring systems. Dans Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 458–468). Association for Computing Machinery.
28 Kaye, K. and Dixon, P. (2023). Risky Analysis: Assessing and Improving AI Governance Tools. World Privacy Forum.
29 Powles, J., Nissenbaum, H. (2018). The Seductive Diversion of 'Solving' Bias in Artificial Intelligence. OneZero.

logiciels occulte une question qui mérite un débat démocratique : un logiciel de reconnaissance faciale non biaisé est-il un horizon souhaitable pour nos villes ?

Rappelons enfin que, d'après les droits de la non-discrimination français et européen, certains critères ne peuvent pas légalement être utilisés, ou bien à des conditions strictes, même s'ils correspondent à une réalité statistique (voir section «Que dit le droit ?»).

Que dit le droit ?

Le droit de la non discrimination

Le droit de la non discrimination prévoit deux types de discrimination : la discrimination directe, où une personne est traitée de manière moins favorable dans une situation comparable, sur le fondement d'un critère protégé (par exemple, le sexe, la situation économique, l'orientation sexuelle).

Mais la discrimination peut également être indirecte : quand bien même aucun critère protégé ne serait utilisé, et qu'il n'y aurait aucune intention de traiter une personne de manière moins favorable que l'autre, il peut y avoir discrimination à certaines conditions si certaines personnes sont désavantagées par rapport à d'autres.
Pour un aperçu accessible et plus complet de ce cadre légal, voir Service-public.fr. (7 juin 2024). Qu'est-ce que la discrimination ?.

Le règlement européen sur l'intelligence artificielle

Le règlement européen sur l'intelligence artificielle (RIA)[30] introduit de nouvelles obligations pour les systèmes d'IA considérés comme «à haut risque». Notamment, il exige que les concepteurs de ces systèmes prennent les mesures nécessaires pour débiaiser leurs jeux de données. Cependant, comme le remarque la chercheuse en droit Sandra Wachter, le règlement reste laconique sur ce que cela implique[31]. La Commission européenne travaille actuellement à l'élaboration de standards pour rendre plus concrètes ces obligations.
Le règlement prévoit également que les utilisateurs de systèmes d'IA à haut risque devront désormais conduire des analyses d'impact sur les droits fondamentaux pour les systèmes à haut risque. Ces analyses d'impact devraient permettre d'identifier les potentielles discriminations pouvant survenir du fait des biais et des subjectivités contenues dans les algorithmes.
Enfin, le règlement européen sur l'intelligence artificielle reconnaît que certains systèmes sont trop problématiques pour être autorisés, et prévoit des lignes rouges sur certaines utilisations, en cas d'impact sur les droits fondamentaux.
Pour un aperçu plus complet du RIA, voir le guide des Interconnectés. (2024). *Data et IA : les nouvelles règles du jeu en Europe.*

B. L'IA, PERFORMANTE ? IA ET ERREURS

1. Le leurre de l'efficacité

Les débats sur les dangers de l'intelligence artificielle se concentrent beaucoup sur les biais. Ce que cela occulte, c'est que bon nombre de logiciels ne fonctionnent surtout pas si bien que cela. En 2018, Julia Dressel et Hany Farid se penchent sur le logiciel propriétaire COMPAS, utilisé par les juges américains pour prédire le risque de récidive des personnes qui comparaissent. Ils démontrent que des personnes sans expérience en matière de justice pénale, recrutées sur le site de microtravail à la tâche Amazon Mechanical Turk, prédisent le risque de récidive aussi bien que le logiciel COMPAS (67% d'exactitude contre 65,2% pour COMPAS), et ce avec seulement 7 facteurs, alors que COMPAS se targue d'en utiliser plus d'une centaine[32]. Ce qui peut remettre en question les dépenses faites pour acquérir le logiciel…

30 Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle et modifiant les règlements (CE) no 300/2008, (UE) no 167/2013, (UE) no 168/2013, (UE) 2018/858, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828 (règlement sur l'intelligence artificielle).
31 Wachter, S. (2024). Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond. Yale Journal of Law & Technology, Volume 26, Issue 3.
32 Dressel, J. et Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science Advances.

Cette histoire illustre ce que la chercheuse Inioluwa Deborah Raji appelle le « leurre de la fonctionnalité »[33] : on a tendance à croire que les outils prédictifs d'intelligence artificielle fonctionnent et sont plus efficaces que les humains, alors que ce n'est souvent pas le cas. On ne compte plus les cas d'algorithmes qui se trompent ou qui ne sont pas si efficients que leurs promoteurs l'affirment. Le Data Justice Lab de l'université de Cardiff s'est penché sur les raisons qui mènent les administrations à abandonner les projets d'intelligence artificielle. Dans la moitié des cas (31 systèmes sur 61 étudiés), le manque d'efficacité des systèmes a joué un rôle dans leur abandon[34].

2. Les fausses promesses des IA génératives

Les outils d'IA générative ne sont pas en reste. Les outils permettant de générer du texte, type chatbots (agents conversationnels), ou de faire des retranscriptions, sont sujets à des « hallucinations », c'est-à-dire que le modèle va produire des informations fausses ou inexistantes.

En 2024, un chatbot mis en place par la ville de New York pour offrir des conseils aux dirigeant·es d'entreprise en se basant sur la législation existante les a enjoint à ne pas respecter la loi, en leur donnant des informations fausses sur le droit du travail[35].

Or, ce chatbot citait ses sources ! En réalité, le fait de citer ses sources n'est pas un gage d'exactitude. En 2023, une étude des chercheurs en informatique Nelson F. Liu et al. a montré que les moteurs de recherche d'IA générative qui citent leurs sources apportent des réponses qui, bien qu'elles soient fluides et semblent pertinentes, contiennent souvent des affirmations non sourcées et des citations inventées ou inexactes. Seulement 51.5% des phrases générées par les moteurs de recherche étaient appuyées par des sources, et seulement 74.5% des sources corroboraient les phrases auxquelles elles étaient associées[36].

L'entreprise Vectara maintient un classement du taux d'hallucination des 25 LLMs les plus importants du marché[37]. Elle montre qu'en novembre 2024, le taux d'hallucination de ces modèles varie entre 1,3% et 4,1%… Le doute est donc toujours permis, et implique une vérification humaine systématique qui peut remettre en question l'argument du « gain de temps » avancé pour mettre en place des agents conversationnels d'aide aux agents.

Soulignons également que les données peuvent être erronées, par exemple dans le cas d'une IA générative entraînée sur des informations obsolètes ou en dehors du contexte du territoire. D'où l'importance de se pencher sur la qualité des données d'entraînement.

Focus sur… L'IA générative pour résumer et retranscrire

En 2024, l'Autorité des Marchés Financiers australienne a testé différents modèles d'IA générative pour résumer des demandes parlementaires, en les comparant à des résumés faits par des agents de l'autorité, à différents niveaux de séniorité : le test a montré que le modèle était bien moins performant que les humains, et manquaient de priorisation, de contexte, et de nuance[38]. Des chercheurs ont également montré que les logiciels de retranscription basés sur l'intelligence artificielle, comme le modèle Whisper d'OpenAI, inventent des choses qui n'ont pas été dites[39]. De quoi interroger à propos de la retranscription de réunions via IA !
Cela ne veut pas dire qu'il faut renoncer aux outils d'IA générative ; seulement que ce qu'ils promettent doit être évalué avec attention et précision.

3. Un algorithme qui dysfonctionne peut coûter cher

33 *Raji, D. I., Kumar, E., Horowitz, A. and Selbst, A. (2022). The Fallacy of AI Functionality. In 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea.*
34 *Redden, J., Brand, J., Sander, I. and Warne, H. (2022). Automating Public Services: Learning from Cancelled Systems. Data Justice Lab.*
35 *Lecher, C. (29 mars 2024). NYC's AI Chatbot Tells Businesses to Break the Law. The Markup.*
36 *Liu, N. F., Zhang, T., Liang, P. (2023). Evaluating Verifiability in Generative Search Engines. Findings of the Association for Computational Linguistics: EMNLP 2023.*
37 *Vectara Leaderboard.*
38 *Wilson, C. AI worse than humans in every way at summarising information, government trial finds. Crikey.*
39 *Burke, G., Schellmann, H. (26 octobre 2024). Researchers say an AI-powered transcription tool used in hospitals invents things no one ever said. AP News.*

L'utilisation d'outils d'automatisation est souvent avancée comme permettant de réduire les coûts, en optimisant le travail des agents. C'est plus compliqué que cela. D'une part, même quand les outils fonctionnent, le coût réel de leur utilisation est rarement évalué : en octobre 2024, la Cour des Comptes a montré que les administrations des ministères économiques et financiers prennent rarement en compte le coût de maintenance des outils dans leurs calculs[40].

Au-delà de ça, la note peut être salée quand l'outil dysfonctionne. En Australie, l'algorithme Robodebt, utilisé entre 2016 et 2019, a généré à tort 470 000 notifications de dettes chez les bénéficiaires de prestations sociales (le taux d'inexactitude de l'outil est estimé à 80%). Une action de groupe a ensuite condamné l'administration australienne à rembourser 721 millions de dollars australiens (plus de 750 millions d'euros) de dettes indues, et à dédommager les parties prenantes d'une action de groupe de plus d'un milliard d'euros[41].

4. L'évaluation des systèmes d'intelligence artificielle : qui fixe les objectifs de l'IA ?

Le sujet de l'efficience pose aussi la question de l'évaluation des systèmes d'IA : qui décide de ce qui constitue le succès d'un système ?

Cette évaluation peut se situer à plusieurs niveaux :

1. Sur la qualité de l'outil en tant que tel, via des métriques d'évaluation techniques (qui ne sont pas encore standardisées)[42], en matière de cybersécurité (voir la section E. «Cybersécurité et protection des données») ou des métriques sur l'impact environnemental (voir la section D. «Conséquences sur l'environnement») ;

2. Au niveau de la politique publique. Au-delà des qualités techniques de l'outil, quels sont les résultats par rapport aux objectifs de politique publique poursuivis ? Cela inclut l'impact social et sur les personnes (notamment par les prismes des droits fondamentaux et de la protection des données personnelles), l'impact organisationnel (voir la section C.) et l'impact économique (retour sur investissement, enjeux de vendor lock-in, c'est-à-dire d'enfermement dans un outil propriétaire).

Ce deuxième niveau pose la question : qui fixe les objectifs de l'outil ? Par exemple, dans le cas d'un outil de prédiction destiné à optimiser le travail des agents : les agents ont-ils le droit de s'exprimer sur les tâches qu'ils souhaitent optimiser ? Par ailleurs, les citoyens et citoyennes ont-ils une voix pour fixer les objectifs d'une politique publique ?

Que dit le droit ?

Le règlement européen sur l'intelligence artificielle précité prévoit de nouvelles obligations spécifiques en matière de qualité des données et des modèles, d'anticipation des risques, de robustesse, d'exactitude, et de documentation, pour les systèmes d'IA à hauts risques.
Les fournisseurs de modèles d'IA à usage général sont également soumis à des obligations de documentation et d'évaluation. Les fournisseurs de modèles d'IA à usage général qui présentent des risques systémiques doivent respecter des obligations supplémentaires.

C. LE CONTRÔLE HUMAIN, UNE GARANTIE ? IA ET SUPERVISION HUMAINE

Face aux limites des outils d'intelligence artificielle précitées, un garde-fou souvent avancé est leur contrôle par les humains. Le règlement européen sur la protection des données (RGPD), par exemple, restreint les décisions qui peuvent être prises de manière complètement automatisée[43]. En 2018, le Conseil Constitutionnel a également jugé que «des algorithmes susceptibles de réviser eux-mêmes les règles qu'ils appliquent, sans le contrôle et la

40 Cour des Comptes. (2024). L'intelligence artificielle dans les politiques publiques : l'exemple du ministère de l'économie et des finances.
41 Redden, J., op. cit.
42 L'élaboration de standards est en cours, notamment dans le cadre du règlement européen sur l'intelligence artificielle.
43 Voir Article 22 du Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données (RGPD) et l'article 47 de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

validation du responsable du traitement» ne pouvaient pas être utilisés pour prendre des décisions administratives individuelles[44]. Le droit prévoit enfin qu'il est interdit d'utiliser des données personnelles dites «sensibles» pour fonder des décisions individuelles entièrement automatisées.

Avec le contrôle humain, la machine proposerait (par exemple, de cibler telle ou telle entreprise pour un contrôle, ou un message type à envoyer à un administré), l'agent disposerait, corrigeant ainsi les erreurs et les biais de l'algorithme. Cependant, ce contrôle est un garde-fou moins efficace qu'il pourrait y paraître.

1. Le contrôle humain peut ne pas être significatif

Parfois, le contrôle humain se limite à un «coup de tampon» (en anglais, rubber stamping). En décembre 2023, la Cour de Justice de l'Union européenne a statué sur le cas de l'agence de crédit allemande SCHUFA, qui produit pour les banques un «score de crédit» qui correspond à une estimation de la solvabilité d'un client. Elle a considéré que cet outil constituait une décision entièrement automatisée, car les employés de la banque qui décidaient d'octroyer ou non le crédit se fiaient systématiquement au score attribué par l'outil[45].

Même quand de réels efforts sont mis en œuvre pour éviter que les humains ne fassent qu'entériner les résultats des algorithmes, des obstacles persistent. Le recours au contrôle humain implique que les personnes qui utilisent les algorithmes sont capables de les contrôler efficacement. Pour ce faire, elles doivent avoir à la fois l'autorité (se voir reconnaître une qualité pour agir) et la compétence (définie par le Défenseur des droits comme «avoir les moyens intellectuels et pratiques pour pouvoir exercer le contrôle»)[46] pour le faire.

Or, les travaux académiques[47] et études sur le terrain montrent une réalité différente.

2. Les biais des humains face aux algorithmes

Dans certains cas, les utilisateurs de ces outils sont sujets au «biais [cognitif] d'automatisation»[48], c'est-à-dire la tendance à faire trop confiance aux résultats de l'algorithme. Dans une étude menée au Canada, des radiologues devaient analyser des mammographies avec ou sans IA. L'outil était paramétré pour proposer de faux résultats pour une partie des images. Quand l'IA proposait un résultat incorrect, les radiologues expérimentés passaient de 80% d'exactitude à 45%, et les débutants à une moyenne de 80 à 22%[49].

Des études ont montré que le biais d'automatisation est susceptible de persister même après des formations et des instructions explicites de vérifier les résultats d'un système automatisé[50]. Le biais d'automatisation peut aussi être encouragé par les pratiques managériales : par exemple, dans le cas d'outils où les agents doivent se justifier uniquement s'ils ne suivent pas les propositions de l'algorithme[51].

Dans d'autres cas, les agents font preuve d'«aversion aux algorithmes» : en réalité, ils n'utilisent pas l'outil, notamment quand ils ne le trouvent pas utile[52] ou bien qu'il va à l'encontre des conclusions auxquelles ils seraient arrivés[53].

44 Cons. Const., 12 juin 2018, n° 2018-765 DC.
45 Ministère de l'Économie, des Finances, et de la Souveraineté industrielle et numérique. (21 décembre 2023). Lettre de la DAJ – Notion de décision individuelle automatisée : la CJUE se prononce sur l'établissement automatisé de la capacité à rembourser un crédit.
46 Op. cit., p.22
47 Voir notamment Liane Huttner, La décision de l'algorithme, étude de droit privé sur les relations entre l'humain et la machine, Dalloz, 2024.
48 Pour un aperçu complet et accessible des biais dans la prise de décision, voir Barot, C. (16 octobre 2024). Croire ou douter : la question des biais de confiance dans la prise de décision. Linc - CNIL.
49 Dratsch, T., Chen, X., Mehrizi, M. R., Kloeckner, R., Mähringer-Kunz, A., Püsken, M., Baeßler, B., Sauer, S., Maintz, D., & Santos, D. P. D. (2023). Automation Bias in Mammography : The Impact of Artificial Intelligence BI-RADS Suggestions on Reader Performance. Radiology, 307(4).
50 Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: an attentional integration. Human factors, 52(3), 381–410.
51 Le juriste Winston Maxwell parle des biais induits par le régime de responsabilité. Voir Maxwell, W. (2022). «Le contrôle humain des systèmes algorithmiques – un regard critique sur l'exigence d'un «humain dans la boucle», mémoire original pour présenter l'habilitation à diriger des recherches, droit. Université Paris 1 Panthéon Sorbonne.
52 Pruss, D. (2023). Ghosting the Machine : Judicial Resistance to a Recidivism Risk Assessment Instrument. 2022 ACM Conference On Fairness, Accountability, And Transparency, 312-323.
53 Voir aussi les travaux de la sociologue Angèle Christin, par exemple Courmont, A. (9 juillet 2020). Angèle Christin : «Les méthodes ethnographiques nuancent l'idée d'une justice prédictive et entièrement automatisée». Linc - CNIL.

Enfin, les agents publics peuvent conserver leurs biais malgré les outils d'aide à la décision. Par exemple, des études ont montré qu'aux États-Unis, des juges utilisant un outil de justice prédictive prenaient des décisions plus sévères à l'encontre des accusés noirs que des accusés blancs, à score de risque égal[54]. Ici, le contrôle humain va à l'encontre de l'objectif premier de l'utilisation des outils algorithmiques, qui est de prévenir les biais humains.

Ainsi, avoir un « humain dans la boucle » ne suffit pas : le contrôle humain est un garde-fou imparfait aux limites des algorithmes. Rappelons d'ailleurs que, dans beaucoup des cas qui ont fait scandale, les algorithmes n'étaient pas complètement automatisés[55].

Un décalage entre le droit et les pratiques : l'exemple d'AFFELNET

Dans son rapport Algorithmes, systèmes d'IA et services publics : quels droits pour les usagers ? de novembre 2024, le Défenseur des droits souligne un « décalage entre le droit et les pratiques ». Il donne l'exemple d'AFFELNET, l'algorithme utilisé pour faciliter la gestion de l'affectation des élèves et des apprentis dans les classes de seconde et de première. Chaque élève se voit attribuer des points sur la base du classement des établissements établi par la famille, des évaluations de l'élève (notamment ses résultats scolaires) et de la capacité d'accueil des établissements. Légalement, l'algorithme est censé simplement participer à la prise de décision, rendue par une commission. Mais, lors de l'instruction d'une réclamation, le Défenseur des droits a pu constater que la décision d'affectation d'une élève avait été prise de façon entièrement automatisée : à cause d'une erreur dans son dossier, son barème de points s'élevait à 0, sans que la commission ne relève cette anomalie. Par la suite, aucun élément sur l'intervention de la commission n'a été communiqué par l'académie aux services du Défenseur.

3. Entre paradoxe de la performance et amortisseurs moraux : un risque d'accroissement de la responsabilité sur les agents

Ce que l'introduction d'outils d'IA change aussi, c'est le rôle des agents et la manière dont leurs décisions vont être perçues. Dans son rapport sur l'IA au travail, le LaborIA parle du « paradoxe de la performance », une situation où « plus les résultats sont bons, moins [l'agent] y gagne » car les améliorations sont mises au crédit de l'outil et non du travail humain. L'agent « ne ferait que bénéficier des apports de l'IA qui augmenteraient ses capacités »[56].

À l'inverse, quand l'outil ne fonctionne pas, la faute a tendance à être rejetée sur l'humain. La chercheuse Madeleine C. Elish, via une observation des systèmes automatisés utilisés dans les avions, a développé le concept d'« amortisseurs moraux » (moral crumple zones) : les pilotes sont considérés responsables des dysfonctionnements de l'outil de pilotage automatique[57].

Ainsi, ce que le contrôle humain mal fait risque de provoquer, c'est un faux sentiment de sécurité qui permet de féliciter l'algorithme quand il fonctionne et de rejeter la faute sur l'humain quand il ne fonctionne pas.

54 *Voir les travaux cités par Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. Computer Law & Security Review, 45, 105681.*
55 *Ben Green.*
56 *Borel, S. (2024). Étude des impacts de l'IA sur le travail. Rapport d'enquête du LaborIA Explorer. p.27*
57 *Elish, M. C. (2019). Moral Crumple Zones : Cautionary Tales in Human-Robot Interaction. Engaging Science Technology And Society, 5, 40-60.*

Focus sur... Les agents conversationnels

Les agents conversationnels sont de plus en plus introduits par les collectivités, dans le but d'améliorer le travail des agents et les réponses aux usagers. Ces usages gagneront à être évalués précisément et en contexte.
Les agents conversationnels sont parfois mis en place pour être utilisés par les agents publics eux-mêmes, comme aide à la recherche ou à la rédaction. Le but : leur faire gagner du temps et améliorer leur travail. Mais, comme vu précédemment, les agents conversationnels, même en citant leurs sources, peuvent être sujets à des hallucinations. Cela invite les agents publics à systématiquement vérifier l'information présentée. Dès lors, le gain de temps initialement souhaité peut être questionné et, dans tous les cas, doit être évalué avec rigueur.
Un autre risque est que l'outil d'IA soit considéré comme responsable d'une éventuelle amélioration de la qualité de service, mais qu'en cas d'erreur (un agent public laisserait passer une erreur dans une réponse), ce soit l'humain qui se voit blâmer, dans la logique de l'amortisseur moral vu précédemment.

Par ailleurs, l'introduction de ces outils pose la question de l'objectif de politique publique poursuivi et de qui a voix au chapitre. D'une part, il est souvent question de "personnaliser les réponses" grâce à l'IA. Or, le récent baromètre Data et Territoire de Data Publica nous apprend que 80% des Français «préfèrent avoir accès à un humain capable de [leur] apporter plus d'explications, même si ça ne peut se faire qu'aux horaires de guichet», plutôt qu'à «une messagerie instantanée avec une intelligence artificielle, qui peut répondre plus vite et à n'importe quelle heure»[(58)]. D'autre part, qui décide des tâches qui sont rébarbatives dans le travail d'un agent public ? Les agents eux-mêmes ont-ils le droit de s'exprimer sur la question ? L'OCDE rappelle ainsi que l'automatisation de toutes les tâches simples des agents les privera du «répit que pouvaient [leur] procurer l'alternance» entre tâches faciles et missions complexes»[(59)].

Que dit le droit ?

Sur le contrôle humain

Le règlement européen sur l'intelligence artificielle introduit de nouvelles dispositions sur le «contrôle humain» pour les systèmes d'IA à haut risque (article 14). Il prévoit que les systèmes doivent être soumis à «un contrôle effectif par des personnes physiques», notamment via les interfaces de ces systèmes. Ce contrôle effectif implique que les humains chargés de la supervision des systèmes doivent comprendre les capacités et les limites du système, soient conscients des biais d'automatisation, soient capables d'interpréter les résultats du système, puissent ne pas suivre les recommandations du système ou l'interrompre. Le règlement prévoit aussi que les agents amenés à utiliser des systèmes d'intelligence artificielle doivent être dotés d'une littératie en IA suffisante.
Les dispositions du règlement doivent encore être précisées via des normes harmonisées (standards), qui sont en cours d'élaboration par le comité technique JTC-21.
Dans tous les cas, une intervention humaine est requise en cas de recours administratif contre une décision individuelle (article 47 LIL).

D. L'IA, UN ATOUT POUR LES AGENTS ? IA ET CONDITIONS DE TRAVAIL

1. Une augmentation du travail

Les outils d'IA sont souvent mis en avant comme une solution pour optimiser le travail des agents et leur permettre de se consacrer à des tâches plus importantes. On pense ainsi que l'IA va mécaniquement alléger ou améliorer le travail des agents. En pratique, c'est plus nuancé. Une enquête portant sur le secteur privé aux États-Unis a montré que 77% des salariés considéraient que les outils d'IA avaient en réalité augmenté leur charge de travail[(60)].

58 *Op. cit.*
59 *OCDE. (2023). Intelligence artificielle, qualité des emplois et inclusivité. in Perspectives de l'emploi de l'OCDE 2023. Éditions OCDE.*
60 *Upwork. (23 juillet 2024). Upwork Study Finds Employee Workloads Rising Despite Increased C-Suite Investment in Artificial Intelligence.*

Dans son rapport sur l'utilisation de l'intelligence artificielle par les ministères économiques et financiers, la Cour des Comptes explique : « ces nouveaux outils peuvent ainsi augmenter la charge de travail des services du fait de leurs performances, en détectant par exemple d'avantage d'irrégularités à traiter, ou du fait de leurs limites, en générant par exemple des réponses erronées qu'il revient aux humains de corriger »[61].

Ce constat est corroboré par le LaborIA, qui rappelle que l'utilisation de nouveaux systèmes d'IA s'accompagne également souvent de temps d'apprentissage et de formation qui s'additionnent aux temps et méthodes de travail habituels des agents[62].

Soulignons également que, dans le secteur privé, les outils numériques peuvent être utilisés à des fins de surveillance et d'optimisation excessive de la cadence de travail[63]. Un appel à redoubler de vigilance pour des utilisations similaires dans le secteur public.

2. Des risques pour les emplois

Dans son rapport cité plus haut, la Cour souligne un paradoxe : même quand les outils sont introduits pour réduire les coûts via une diminution des emplois, cet objectif est souvent peu explicité. Elle indique : « les directions sont amenées à présenter les projets de systèmes d'IA exclusivement sous l'angle de l'amélioration de la qualité de service et de vie au travail. Or cette stratégie oblitère une partie de l'information nécessaire pour évaluer la pertinence des projets et décider de l'affectation des gains de productivité, sans écarter pour autant la défiance des partenaires sociaux. Elle conduit par ailleurs les porteurs de projets à insister sur des améliorations qualitatives qu'ils ne sont pas toujours en capacité de prouver ».

Bien que ce constat porte sur les ministères économiques et financiers, il nous invite à nous interroger sur des dynamiques similaires dans d'autres administrations. De surcroît, même si le but initial n'est pas de diminuer les emplois, la tentation peut être grande de le faire une fois les outils introduits.

3. Quels enjeux pour l'expertise métier ?

Perte d'expertise métier

L'introduction des systèmes algorithmiques risque d'entraîner une perte de l'expertise métier. On sait d'une part que la tendance à se fier aux algorithmes dépend du niveau d'expertise dans la tâche donnée[64]. D'autre part, quand les résultats de l'algorithme sont opaques et mal expliqués, cela empêche les agents de les superviser. L'ouvrage L'IA aux impôts du syndicat Solidaires Finances Publiques décrit comment les agents s'interrogent sur les directives de l'algorithme TAAP, qui sert à déterminer les dépenses de l'État à contrôler, via des critères opaques : « Pourquoi certaines dépenses insignifiantes qui n'étaient jamais contrôlées auparavant, le sont à présent, alors que d'autres sont validées en masse quand elles étaient scrupuleusement regardées avant l'IA ? »[65]. Conséquence : les nouveaux agents n'auront pas nécessairement la latitude ou l'initiative d'aller chercher les dossiers non suggérés par l'algorithme, menant à un appauvrissement de l'expertise.

France Stratégie parle ainsi de « prolétarisation des savoirs »[66]. C'est aussi ce qu'esquisse l'étude du LaborIA sur l'impact de l'IA au travail, qui alerte sur le risque de perte de la vigilance, entraînant une dépendance à l'IA[67]. Elle donne l'exemple de son étude de terrain dans une entreprise du secteur aérien, où la délégation à l'IA des opérations de détection des défauts moteur fait courir le risque que les techniciens, notamment ceux sans expérience métier, « [perdent] en capacité d'exercer la responsabilité sur [leur] travail et de pouvoir en rendre compte sur le fond »[68]. Cependant, ladite étude propose aussi des pistes de configurations dites « encapacitantes » entre l'humain et les systèmes d'IA.

61 *Cour des Comptes, op. cit.*
62 *Borel, S. (2024). Étude des impacts de l'IA sur le travail. Rapport d'enquête du LaborIA Explorer. p.28*
63 *Voir par exemple Dumont, A. (2019, 11 décembre). « C'est inhumain, l'aliénation maximum » : dans les entrepôts logistiques de la région lyonnaise. Médiacités.*
64 *Barot, C. op. cit.*
65 *Guillaud, H. (13 novembre 2024). IA aux impôts : vers un « service public artificiel » ? Dans les algorithmes.*
66 *France Stratégie. (2016). Anticiper les impacts économiques et sociaux de l'Intelligence Artificielle. Annexe 1 : L'intelligence Artificielle en quête d'acceptabilité et de confort.*
67 *Borel, S. (2024). Étude des impacts de l'IA sur le travail. Rapport d'enquête du LaborIA Explorer.*
68 *Borel, S. (2024). Étude des impacts de l'IA sur le travail. Rapport d'enquête du LaborIA Explorer. p.44*

Réduction de la capacité d'agir

Dans L'IA aux impôts, les agents de Solidaires Finances Publiques expliquent qu'ils n'ont plus la possibilité de prendre en compte les fraudes complexes, l'algorithme poussant pour la détection des fraudes simples. Le système les force par exemple à contrôler les dépenses d'essence des ambassades, jusqu'à présent vues commes des dépenses à faibles enjeux. Leur expérience fait écho à l'analyse de Ferguson et Pecoste, qui expliquent que le déploiement des systèmes d'IA au travail peut réduire « la capacité d'agir des individus » qui doivent « suivre des instructions sans aucune autre finalité » et sans pouvoir les adapter avec leur savoir-faire propre[69]. C'est l'un des effets de la centralisation des critères de décision par les algorithmes mentionnés dans la section A sur les choix et les biais.

Cela a aussi des conséquences potentielles sur le bien-être au travail. Le LaborIA explique que « le déploiement d'un [système d'IA] au sein d'une équipe de travail peut entraîner une dévaluation de la reconnaissance, [qui] concerne à la fois la singularité de l'individu en question, ainsi que ses pratiques, les efforts qu'il déploie et les résultats qu'il obtient ». Un cadre de direction d'une administration publique utilisant un système de détection d'anomalies détaille ainsi : « pour certains agents, ça serait un retour en arrière de ne faire que de la vérification. Ils font comme tout le monde alors qu'avant ils avaient des tâches bien précises qui n'incombaient qu'à eux. Ils étaient les seuls à faire ce travail »[70]. Le CESE alerte aussi sur les risques sur la santé entraînés par ces transformations et la perte de sens qui peut en résulter[71].

4. Des enjeux à étudier sur le terrain

Au-delà des espoirs et des craintes, les effets de l'IA et leur perception par les métiers sont difficilement prévisibles avant leur utilisation en conditions réelles. C'est ce que révèle par exemple l'étude du LaborIA sur un système de détection des irrégularités dans la rédaction d'éléments administratifs. L'enquête terrain menée par les sociologues du LaborIA révèle que les alertes générées par le système, « loin d'être discutées comme économes en temps ou chronophages, formatrices ou prescriptrices de nouvelles tâches », sont en réalité peu nombreuses et n'ont pas de réelle incidence le travail des agents, qui n'y pensent même pas[72].

Cela plaide pour une étude en contexte des usages, qui peuvent parfois être moins clivants qu'anticipé, mais aussi pour une mise en place progressive des systèmes, afin d'identifier des effets inattendus. Par exemple, dans leur étude sur l'intégration d'un outil d'IA de prédiction des risques de septicémie dans un hôpital, Madeleine Clare Elish et Elizabeth Anne Watkins mettent en lumière les effets non anticipés de l'outil sur le travail des infirmières, qui réorganisent leur travail pour que l'introduction d'un nouvel outil n'impacte pas les médecins et soit mieux acceptés[73].

69 *Ferguson, Y., Pecoste. C.. (2022). L'IA au travail: propositions pour outiller la confiance. Conférence Nationale sur les Applications Pratiques de l'Intelligence Artificielle (APIA-PFIA).*
70 *Borel, S. (2024). Étude des impacts de l'IA sur le travail. Rapport d'enquête du LaborIA Explorer. p.36*
71 *Conseil économique, social et environnemental. (2025). Analyse de controverses : intelligence artificielle, travail et emploi.*
72 *Borel, S. (2024). Étude des impacts de l'IA sur le travail. Rapport d'enquête du LaborIA Explorer.*
73 *Elish, M.C. et Watkins, E. A. (2020). Repairing Innovation: A Study of Integrating AI in Clinical Care. Data & Society.*

3. ENJEUX POUR UNE IA RESPONSABLE

Une fois ces mythes déconstruits, penchons-nous maintenant sur trois autres enjeux à prendre en compte avant de mettre en place des systèmes d'IA.

A. CONSÉQUENCES SUR L'ENVIRONNEMENT ET LA SOCIÉTÉ

Si l'utilisation du numérique (et donc des algorithmes) par l'administration pose dans tous les cas des enjeux pour l'environnement[74], ceux-ci sont intensifiés par le développement des outils d'IA générative.

Pour être entraînés et ensuite pour répondre aux différentes requêtes, les outils d'intelligence artificielle reposent sur une infrastructure matérielle (hardware) : des serveurs (aussi appelés GPU, pour graphics processing units) qui sont stockés dans des data centers un peu partout dans le monde. L'intelligence artificielle générative implique des infrastructures bien plus importantes et gourmandes en ressources que les technologies précédentes, notamment car elle repose sur beaucoup plus de données et s'appuie sur des modèles avec encore plus de paramètres.

1. Des systèmes gourmands en ressources

On manque encore de données standardisées sur les impacts environnementaux de l'intelligence artificielle générative[75]. Cependant, les recherches et chiffres existantes montrent que celle-ci est gourmande en ressources naturelles (eau, minerais, électricité) et productrice de gaz à effets de serre (du fait de l'électricité mais aussi des matériaux nécessaires à la construction des data centers tels que le béton) tout au long de son cycle de vie, de la fabrication des infrastructures à l'entraînement des modèles en passant par l'utilisation des outils.

Pour ne citer que quelques chiffres :

- Une recherche effectuée sur un moteur de recherche utilisant l'IA générative utilise 6 à 10 fois plus d'énergie qu'un moteur de recherche classique[76].
- Entraîner GPT-3 a nécessité 5.4 millions de litres d'eau, et GPT-3 a besoin d'une bouteille d'eau de 50cl toutes les 10 à 50 requêtes[77].
- Les baies de GPU consacrés à l'IA nécessitent six à neuf fois plus de puissance électrique qu'une baie de serveurs classiques[78].
- En 2024, RTE a annoncé que la puissance électrique demandée par l'ensemble des projets data centers en Île-de-France atteint 7,5 GW. C'est le niveau de puissance nécessaire à l'alimentation de tous les habitants, services et entreprises sur la Métropole du Grand Paris à la pointe de l'hiver. Elle estime aussi que la consommation des data centers pourrait atteindre 80 TWh/an, dépassant toute la consommation électrique d'Île de France (60 TWh/an)[79]. Aux États-Unis, Microsoft a sollicité le gestionnaire d'une centrale nucléaire fermée pour la rouvrir répondre à ses besoins futurs[80].

La prévision en consommation d'énergie est telle qu'une étude du cabinet Gartner prédit que d'ici 2027, 40% des data centers dans le monde seront contraints par des pénuries d'énergie[81].

74 *Voir ADEME Magazine. (Avril 2022). Numérique : quel impact pour l'environnement ?, et l'étude ADEME - Arcep publiée en 2025 et téléchargeable ici*

75 *Voir les remarques de Luccioni, S., Jernite, Y., & Strubell, E. (2024, June). Power Hungry Processing: Watts driving the cost of AI deployment?. In Proceedings of 2024 ACM FAccT Conference. pp. 85-99 ainsi que Roussilhe, G. (2023). Territorialiser les systèmes numériques, l'exemple des centres de données [4/4]. Gauthier Roussilhe.*

76 *Luccioni, S. et al, op. cit.*

77 *Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). Making AI Less «Thirsty» : Uncovering and Addressing the Secret Water Footprint of AI Models. arXiv (Cornell University).*

78 *Voir notamment Piquard, A., Pinaud, O. et Pécout, A. (14 juin 2024). Derrière l'IA, la déferlante des data centers. Le Monde.*

79 *Diguet, C. (2024, 8 juillet). «Les data centers s'implantent de manière totalement opportuniste». La Croix.*

80 *Blandin, N. (2024, 21 septembre). Aux États-Unis, Microsoft veut relancer une centrale nucléaire arrêtée. Ouest-France.*

81 *Gartner. (2024). Gartner Predicts Power Shortages Will Restrict 40% of AI Data Centers By 2027.*

2. Conséquences sur les territoires et les populations

L'empreinte environnementale de l'IA générative ne se limite pas à sa consommation de ressources. L'accroissement de la demande en data centers a également des conséquences sur les territoires, notamment en mobilisant le foncier existant sans qu'un débat démocratique autour de son utilisation ne soit instauré, ou que l'impact de ces implantations sur la biodiversité ne soit vraiment mesurée[82].

D'autre part, les minerais nécessaires à la fabrication des composantes s'inscrivent dans des filières d'extraction minières violentes, notamment en République Démocratique du Congo[83].

Rappelons enfin que la production des données nécessaires au fonctionnement des outils d'intelligence artificielle générative repose souvent sur des travailleurs sous-payés dans les pays du Sud. Par exemple, OpenAI a fait appel à des travailleurs au Kenya payés 2 dollars de l'heure[84]. Les données francophones sont très souvent annotées par des travailleurs à Madagascar qui travaillent dans des conditions similaires.

Un mouvement vers l'IA frugale

Pour répondre à ces enjeux, le ministère Aménagement du territoire Transition écologique a été à l'initiative du référentiel AFNOR Spec 2314 « Référentiel général pour l'IA frugale ». Publié en juin 2024, il est issu d'un groupe de travail piloté par l'AFNOR et composé d'une quarantaine d'acteurs emmenés par le Commissariat général au développement durable ».

L'IA frugale implique :

- De démontrer la nécessité de recourir à un système d'IA plutôt qu'à une autre solution moins consommatrice pour répondre au même objectif ;
- D'adopter de bonnes pratiques pour diminuer les impacts environnementaux du service ;
- Que les usages et les besoins ont été préalablement questionnés, et qu'ils visent à rester dans les limites planétaires.

Que dit le droit ?

Au niveau européen, les questions environnementales sont prises en compte dans plusieurs textes, notamment dans le cadre du ***pacte vert pour l'Europe.*** La directive sur l'efficacité énergétique vise à développer des scores sur la soutenabilité des data centers en France. Les mesures environnementales font également partie intégrante du règlement européen sur l'IA.
La loi Réduction de l'Empreinte Environnementale du Numérique (REEN), votée en novembre 2021, prévoit notamment que les communes et EPCI de plus de 50 000 habitants doivent s'être dotés d'une stratégie numérique responsable à partir du 1er janvier 2025. Celle-ci doit indiquer les objectifs de réduction de l'empreinte environnementale du numérique et les moyens pour les atteindre. L'INR et les Interconnectés, avec le concours de la Banque des Territoires, ont mis au point un outil d'auto-évaluation de l'impact environnemental du numérique pour les collectivités et un guide méthodologique disponibles ***ici.***

82 Balova, A. et Kolbas, N. (2023). Biodiversity and Data Centers: What's the connection?. Ramboll Group. ; pour une analyse détaillée, voir Roussilhe, G. (2023), op. cit.
83 Pauron, M. (2025, 27 janvier). « La République démocratique du Congo est une réserve pour les dominants ». Médiapart. Voir également les travaux de l'association Génération Lumière.
84 Perrigo, B. (2023, 18 janvier). Exclusive : OpenAI Used Kenyan Workers on Less Than $ 2 Per Hour to Make ChatGPT Less Toxic. TIME.

B. CYBERSÉCURITÉ ET PROTECTION DES DONNÉES PERSONNELLES

On l'a montré précédemment : la qualité des données est essentielle, à la fois pour se prémunir de la reproduction de stéréotypes ou d'inégalités contenues dans les jeux de données d'entraînement, mais aussi pour s'assurer que les résultats des outils seront de qualité. À ces enjeux s'ajoutent ceux de la cybersécurité et de la protection des données personnelles.

1.Traitement de données et protection des données personnelles

Plus de données, plus de risques de les voir fuiter

L'utilisation d'outils d'intelligence artificielle pour prendre des décisions ou réaliser des prédictions vis-à-vis d'individus implique de collecter un nombre toujours plus important de données, et donc d'accroître le risque qu'elles soient piratées et exploitées pour des usages malveillants. Par ailleurs, les administrations sont susceptibles de confier ces données à des prestataires privés, augmentant d'autant plus le risque. En mars 2024, France Travail et son prestataire Cap Emploi ont été victimes d'une cyberattaque ayant conduit à une fuite de données touchant 43 millions de personnes[85]. La CNAF a vu ses données fuiter à deux reprises en 2024[86]. D'autres outils sont également concernés par les risques de protection des données, par exemple ceux qui traitent des données de la voirie qui captent des images sur la voie publique, dont des visages de personnes.

Ces considérations sont d'autant plus importantes que les collectivités sont vulnérables aux cyberattaques. Entre janvier 2022 et juin 2023, l'Agence de sécurité des systèmes d'information (ANSSI) a enregistré 187 cyberattaques visant les collectivités territoriales[87]. En 2024, 26% des collectivités de plus de 3500 habitants ayant répondu au baromètre Data & Territoires ont déclaré avoir subi une cyberattaque sévère[88].

Exploitation des données versus respect du droit à la vie privée

Par ailleurs, le traitement de données personnelles est à mettre en balance avec le respect de certains droits fondamentaux. Par exemple, aux Pays-Bas, la justice néerlandaise a considéré que l'outil de lutte contre la fraude aux prestations sociales SyRI était contraire à l'article 8 de la Convention européenne des droits de l'Homme, car son usage des données était disproportionné par rapport au droit au respect de la vie privée, pour l'objectif poursuivi[89].

Ce qui est en jeu ici, c'est la confiance des citoyens et citoyennes en la capacité de la collectivité à protéger les données essentielles à l'exercice de ses missions de service public.

2.Enjeux de cybersécurité spécifiques aux outils d'IA générative

Les outils d'IA générative exacerbent les enjeux de cybersécurité et de protection des données.

Par exemple, les grands modèles de langage sont entraînés sur de larges quantités de données. S'ils ne sont pas sécurisés correctement, ils peuvent divulguer des données personnelles provenant des données d'entraînement, notamment par « régurgitation »[90].

De surcroît, quand un utilisateur utilise un agent conversationnel, il y a un risque que l'éditeur du logiciel ait accès aux données fournies. Cela pose des enjeux à la fois de respect de la vie privée (si les données concernent des individus) et de sécurité et de propriété intellectuelle. En 2023, des employés de l'entreprise Samsung ont été sanctionnés pour avoir fourni à l'entreprise OpenAI, qui développe ChatGPT, des informations sur des discussions internes à l'entreprise ainsi que des codes sources, via des requêtes au chatbot[91].

85 *France Travail. (13 mars 2024). Communiqué : France Travail et Cap emploi victimes d'une cyberattaque.*
86 *CNAF. (23 février 2024). Communiqué : Alerte aux mots de passe volés : la Cnaf renforce la sécurité des comptes des allocataires.*
87 *Vie publique. (3 novembre 2023). Cyberattaques : de fortes menaces sur les collectivités territoriales.*
88 *Observatoire Data Publica, op. cit.*
89 *OpenGlobalRights. (19 mars 2020). Aux Pays-Bas, une décision judiciaire historique sur les États-providences numériques et les droits humains.*
90 *Leautier, A. (2024). Incursion dans la caisse à outils de la sécurité des grands modèles de langage (1/2) : risques visés et sécurité dès la conception. Linc CNIL. Voir aussi Weidinger, L. and others. (2022). Taxonomy of Risks Posed by Language Models. In 2022 ACM Conference on Fairness, Accountability, and Transparency.*
91 *Kachalova, E. (2023, 12 avril). Can't take it back : Samsung ChatGPT fiasco highlights generative AI privacy woes. AdGuard Blog.*

Les risques se posent donc à la fois pour les outils achetés et utilisés de manière officielle par les collectivités territoriales, mais aussi en cas d'utilisation informelle d'outils propriétaires gratuits par les agents, comme ChatGPT ou Claude (ce qu'on appelle le shadow IT, en français « informatique de l'ombre »).

Par ailleurs, les outils d'intelligence artificielle générative sont sujets à des cyberattaques particulières, tels que les « prompts adversariaux » qui permettent de détourner des informations sans que l'utilisateur ne s'en aperçoive[92]. Dans ses Recommandations de sécurité pour un système d'IA générative d'avril 2024, l'Anssi détaille de manière claire et concise les cyberattaques susceptibles de toucher les outils d'IA générative[93].

Que dit le droit ?

En France, le droit des données personnelles est régi par la loi dite « informatique et libertés » (LIL) et le règlement européen sur la protection des données. Entre autres, l'article 4 de la LIL (et l'article 5 du RGPD) listent les principes relatifs au traitement des données à caractère personnel, dont la proportionnalité. La loi prévoit également la réalisation d'une analyse d'impact sur la protection des données (AIPD) dans certains cas de traitements de données.
Le règlement européen sur l'intelligence artificielle introduit des obligations de garanties en termes de cybersécurité pour les systèmes considérés comme à hauts risques.
En matière de cybersécurité, la directive européenne NIS 2 est en cours de transposition en droit français via un projet de loi. Pour plus d'informations sur la directive, voir ANSSI. (2023). ***La directive NIS 2,*** et le ***Guide Data IA des Interconnectés,*** p.17.

C. TRANSPARENCE, REDEVABILITÉ ET MAÎTRISE

La transparence et la redevabilité sont des principes phares de l'action publique. L'article 15 de la déclaration de l'homme et du citoyen consacre le principe de redevabilité de l'administration (« la société a le droit de demander compte à tout agent public de son administration ») et le droit d'accès des citoyens aux documents administratifs, instauré par la loi « CADA »[94] dès 1978, est un droit constitutionnel[95].

La transparence ne concerne pas seulement les systèmes algorithmiques, mais bien toute l'action publique. Cependant, l'acquisition et l'utilisation de systèmes d'intelligence artificielle renouvellent les questions de transparence et de redevabilité des administrations.

1. La transparence et la redevabilité : comment, pourquoi ?

La transparence, un pilier de la redevabilité

On distingue :

- **La transparence individuelle,** d'une décision (par exemple : un administré soumis à une décision se voit notifier qu'un algorithme a été utilisé pour calculer ses prestations sociales et le rôle qu'il a joué) ou d'un résultat (par exemple : un agent public qui utilise un logiciel pour prédire les entreprises en difficulté comprend les raisons d'attribution d'un score à un dossier).
- **La transparence générale,** qui porte sur le système en lui-même. Elle comprend non seulement des informations techniques sur l'algorithme, mais aussi sur la manière dont celui-ci s'insère dans le processus de décision, ses conditions de production (qui a construit l'outil ? est-ce que toutes les parties prenantes ont été impliquées ? combien cela a-t-il coûté ?), les choix effectués (pourquoi l'algorithme a-t-il été mis en place par rapport à un autre système ? quels ont été les compromis entre précision et sécurité, ou précision et transparence ?), ou les impacts de l'outil (sur les politiques publiques, les agents, l'environnement, les citoyens et citoyennes, etc.). Cette

92 *Burgess, M. (2024, 8 août). Microsoft's AI Can Be Turned Into an Automated Phishing Machine. WIRED.*
93 *ANSSI. (2024). Recommandations de sécurité pour un système d'IA générative.*
94 *Loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal.*
95 *Décision n°2020-834 QPC du 3 avril 2020.*

transparence peut prendre diverses formes : inventaires d'algorithmes publics[96], documentation des systèmes[97], formats multimédia[98], ou ouverture du code source[99], des jeux de données et des modèles afin de les rendre consultables et auditables par des tiers.

La transparence est un continuum, qui va de l'information (transmettre des informations), à l'explication (rendre compréhensible), à la justification (donner des raisons en faisant appel à des éléments extérieurs tels que des textes de loi et des objectifs de politiques publiques)[100].

Si elle présente des limites[101], la transparence est donc à la fois un principe en soi et une première étape pour comprendre, débattre, voire contester les politiques publiques outillées par des algorithmes. Elle peut ainsi permettre d'améliorer les systèmes, d'éviter des atteintes aux droits fondamentaux, et de construire la confiance pour s'assurer que les systèmes sont acceptés par les utilisateurs et les usagers.

Par ailleurs, la transparence et la capacité à expliquer un modèle d'intelligence artificielle sont des facteurs importants d'appropriation des outils par les agents publics[102]. À l'inverse, les chercheurs du LaborIA soulignent que, dans le cas d'un système d'intelligence artificielle utilisé dans un établissement public, le manque de formation et l'opacité du système d'IA le rendait difficilement supervisable et donc exploitable[103].

La redevabilité : de grands pouvoirs, de grandes responsabilités

La redevabilité (en anglais, accountability) peut être définie simplement comme le fait qu'une entité ou une personne en charge d'un outil d'intelligence artificielle soit responsable («rende des comptes») devant un public de la prévention et de la réparation des éventuels effets néfastes de l'utilisation de l'outil[104]. Cela implique, entre autres, de s'assurer que le système fonctionne, de mettre en place des voies de recours, des processus et outils pour faire remonter les erreurs ou les soucis des agents publics ou des usagers, de prendre en compte des personnes concernées dans la conception et l'évaluation du système, et d'instaurer un processus de suivi et d'audit tout au long du cycle de vie du système[105].

TRANSPARENCE		REDEVABILITÉ
Générale	De l'insertion du système dans le processus de décision	Identification des personnes responsables de système
	Des détails techniques	Implication des parties prenantes
	Des choix de conception	Formation des utilisateurs et des concepteurs
	Des performances du système	Maintenance
	De le conformité	Évaluation, traçabilité et monitoring
	Des impacts	Réversibilité
Individuelle	Des l'utilisation d'un algorithme	
	Explication d'un résultat	

Les facettes de la transparence et de la redevabilité. Source : autrice.

96 *Comme à Nantes ou à Antibes.*
97 *Voir par exemple le Guide du Score Cœur, qui explique le fonctionnement de l'algorithme d'allocation des greffons cardiaques.*
98 *Explication vidéo de La Bonne Boîte.*
99 *Voir l'inventaire des codes sources du secteur public de la DINUM, qui contient des algorithmes.*
100 *Vallet, F. (6 janvier 2021). Entretien avec Clément Henin, Daniel Le Métayer : «Fournir des explications du fonctionnement des algorithmes compréhensibles par des profanes». Linc - CNIL.*
101 *Voir notamment : Ananny, M. & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society. et Vuarin, L., & Steyer, V. (2023). L'injonction à la transparence : un levier réglementaire à double tranchant pour les organisations. Bulletin de l'Association Française pour l'Intelligence Artificielle.*
102 *Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. International Journal of Human-Computer Studies 146: 102551., cité dans Borel, S. (op. cit.)*
103 *Borel, S. (op. cit.), p.17.*
104 *Wieringa, M. (2020). What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). Association for Computing Machinery, New York, NY, USA, 1–18.*
105 *Pour une liste de questions exhaustives afférentes à la transparence et la redevabilité, voir le questionnaire d'évaluation du groupe d'experts de haut niveau sur l'intelligence artificielle de la Commission européenne, publié en juillet 2020 et toujours pertinent.*

2. Des obligations légales peu respectées

Or, la manière dont les outils algorithmiques ont été rendus transparents et redevables jusqu'ici peut jeter le doute sur la capacité des administrations publiques à utiliser ces outils de manière véritablement responsable.

Un cadre juridique restreint

La France jouit d'une législation nationale ambitieuse, qui impose des obligations de transparence aux administrations utilisant des algorithmes pour prendre des décisions partiellement ou complètement automatisées (voir encadré «Que dit le droit ?»).

Cependant, celle-ci comporte des limites :

- **En termes de périmètre :** elle ne concerne que les systèmes impliqués dans les décisions administratives individuelles. Les systèmes de gestion urbaine par exemple ne sont donc pas concernés.
- **Elle comporte des exceptions liées à la protection de certains secrets,** notamment «la sécurité publique, la sécurité des personnes», «la sécurité des systèmes d'information des administrations» ou «la recherche et (...) la prévention (...) d'infractions de toute nature» (par exemple : la lutte contre la fraude)[106]. Ainsi, certains algorithmes critiques échappent à la transparence. Par exemple, la CADA a considéré que l'algorithme utilisé par la CNAF pour attribuer un score de risque aux allocataires n'était pas communicable, car la communication des critères de risque était susceptible d'augmenter le risque de fraude[107][108].
- **Enfin, la transparence doit être mise en balance avec d'autres droits et enjeux,** tels que le respect de la vie privée, la propriété intellectuelle et la cybersécurité. Si cela est justifié dans bien des cas, il en résulte que de nombreuses informations restent inaccessibles.

En pratique, un respect limité des obligations

Ce que l'on constate sur le terrain, c'est aussi que très peu d'administrations respectent leurs obligations de transparence.

Ainsi, seule une poignée de collectivités respectent l'obligation d'information générale impliquant de publier en ligne un inventaire de leurs algorithmes publics[109].

Plusieurs facteurs peuvent l'expliquer, dont l'absence de sanction en cas de non respect[110], mais aussi par un manque de compétences et de ressources au sein des administrations[111].

En conséquence, difficile pour les citoyens et citoyennes de savoir que les outils existent, mais aussi qu'une décision a été prise et les tenants de la décision, et donc de faire valoir leurs droits. En 2019, le rapporteur spécial à l'ONU sur l'extrême pauvreté et aux droits humains Philip Alston avait également souligné que cette opacité entraîne une inversion de la charge de la preuve : c'est désormais aux citoyens et citoyennes de prouver que les algorithmes dysfonctionnent, et pas l'inverse[112].

106 2° de l'article L.311-5 du CRPA.
107 Commission d'accès aux documents administratifs. Rapport d'activité 2022-2023.
108 D'autres exceptions sont inscrites dans des lois spécifiques, comme le secret des délibérations pédagogiques pour le cas de Parcoursup.
109 Parmi les collectivités territoriales qui le font, mentionnons la Région Île de France, la Ville de Paris, la Métropole Européenne de Lille, le Département du Val d'Oise, le Département d'Ille et Vilaine, la Ville d'Antibes, Nantes Métropole et la Ville de Nantes.
110 Avec une exception : depuis 1er juillet 2020, tout traitement automatisé doit comporter, à peine de nullité, l'obligation de mention explicite. Voir Etalab. (2021). Les algorithmes publics : pourquoi et comment les expliquer ?
111 ENA. Promotion 2018-2019 «MOLIÈRE». (2019). Rapport collectif sur commande d'une administration centrale Ethique et responsabilité des algorithmes publics Groupe n° 12. École nationale d'administration. Voir aussi le rapport du Public Law Project sur cinq cadres légaux sur la transparence algorithmique : Public Law Project. (2024). Securing meaningful transparency of public sector use of AI: Comparative approaches across five jurisdictions.
112 Citer source

3. Des expérimentations opaques

Au-delà des obligations légales, on constate également que les expérimentations des outils algorithmiques se font souvent dans une certaine opacité. Leur existence est souvent connue au lancement de l'expérimentation, mais les résultats des expérimentations et leurs conséquences sont rarement publiés. Pour un exemple à l'étranger : le journal The Guardian révélait récemment que le ministère des affaires sociales avait renoncé à plusieurs expérimentations testées en 2024, du fait de résultats insuffisants. Ces informations n'ont été obtenues qu'en faisant des demandes CADA au ministère concerné[113].

À notre connaissance, il n'existe pas d'expérimentation en France qui ait clairement indiqué les conditions de sa réversibilité.

La publication des résultats des expérimentations n'est pas une obligation légale. Cependant, elle soulève des questions démocratiques : dans ces conditions, il est difficile de s'assurer que le terme « expérimentation » n'est pas dévoyé et utilisé pour introduire des systèmes dont il sera ensuite impossible de se passer.

Introduire des outils d'IA dans un service public sans participation, transparence ou mesures de redevabilité peut éroder la confiance apportée par les citoyens et citoyennes aux services publics[114]. À noter que le baromètre de l'observatoire Data Publica note une augmentation de la défiance des citoyen·nes à l'égard de l'IA (52% de sentiments négatifs en 2024 contre 51% en 2022)[115].

4. Un besoin de formation et de dialogue social

La Cour des Comptes remarque par ailleurs que cette modification du travail, et celle de la répartition des métiers et des compétences, ne sont ni documentées ni intégrées dans le dialogue social. Or, l'absence de discussion autour des outils, ou de suivi et de mesure des usages, risque de contribuer à la mise en œuvre d'outils mal adaptés ou mal évalués. Le LaborIA explique ainsi que, dans leur terrain sur une entreprise du secteur aérien, les techniciens interviewés étaient « demandeurs de participer à la compréhension et à la dénition du « curseur de sensibilité » » utilisé pour détecter les défauts, et/ou de « modalités d'analyse comparative des défauts détectés (…) au risque que les techniciens ne prennent pas en compte les détections de l'IA (« je laisse l'IA de côté ») »[116].

5. Des enjeux de souveraineté et de maîtrise

Comme esquissé précédemment, la transparence est non seulement importante pour les citoyens et citoyennes, mais aussi pour l'administration elle-même, afin que les agents publics (et en particulier les services métier) aient connaissance, comprennent, et ainsi gardent la main sur les outils qu'ils utilisent.

Au-delà de la transparence se posent des questions de souveraineté et de maîtrise. Le recours à des logiciels développés par des éditeurs peut priver les administrations de la maîtrise de ces systèmes : phénomène de vendor lock-in (en français : dépendance à un logiciel propriétaire), impossibilité de transférer ou d'interconnecter les données d'un outil vers un autre, etc. Les outils d'IA générative exacerbent ces enjeux, car ils reposent sur un ensemble de briques développées par différentes entités. Même des outils développés « en interne » sont susceptibles de reposer sur des données d'entraînement qui ne correspondent pas aux contextes de nos territoires, ou des modèles propriétaires.

113 Booth, R. (2025, 27 janvier). AI prototypes for UK welfare system dropped as officials lament 'false starts'. The Guardian.
114 Dent, A. (2024). Automating Public Services: A Careful Approach. Promising Trouble.
115 Observatoire Data Publica. (2024). Baromètre de l'Observatoire Data Publica 2024, 3e édition : les collectivités territoriales et la donnée.
116 Borel, S. (2024). Étude des impacts de l'IA sur le travail. Rapport d'enquête du LaborIA Explorer. p.19

Que dit le droit ?

Le cadre national sur la transparence

Les algorithmes publics sont encadrés par trois cadres juridiques :

- Depuis la loi pour une République numérique de 2016[117], les administrations qui utilisent des algorithmes publics pour prendre des décisions administratives individuelles entièrement ou partiellement automatisées sont soumises à des obligations de transparence prévues par le Code des relations entre le public et l'administration : fournir une information générale (article L.312-1-3), faire figurer une mention explicite (article L.311-3-1) et fournir une information individuelle à la demande de l'intéressé (article R.311-3-1-2). À noter que ces décisions peuvent concerner des personnes morales (entreprises, associations, collectivités…).
- Les algorithmes qui impliquent un traitement de données personnelles sont, de surcroît, concernés par le droit de la protection des données personnelles. L'article 47 de la loi informatique et libertés encadre notamment l'utilisation d'algorithmes prenant des décisions automatisées avec des conditions de transparence et d'information et d'un droit de recours des administrés.
- Les codes sources sont des documents administratifs communicables[118] et relèvent donc du cadre du droit d'accès aux documents administratifs.

Pour un tableau récapitulatif du cadre légal, voir le rapport Algorithmes, systèmes d'IA et services publics : quels droits pour les usagers ? Points de vigilance et recommandations du Défenseur des droits, pp. 33-34.

Le droit d'accès aux documents administratifs

Outre les dispositions spécifiques aux algorithmes, le cadre juridique du droit d'accès aux documents administratifs prévoit que les collectivités doivent publier par défaut leur répertoire d'informations publiques et les documents qu'il contient (article L. 322-6 du CRPA). Cela pourrait concerner les évaluations des expérimentations de systèmes d'intelligence artificielle.
Par ailleurs, les administrations[119] qui communiquent des documents suite à une demande CADA sont tenues de les mettre en ligne (article L. 312-1-1 du CRPA).

Le règlement européen sur l'intelligence artificielle

Le règlement introduit de nouvelles dispositions pour les systèmes considérés comme à haut risque, notamment en matière de transparence individuelle (exemple : droit à l'explication des décisions individuelles, information des personnes qu'elles interagissent avec un système d'intelligence artificielle), de transparence des systèmes (notamment : constitution d'une base de données recensant les systèmes considérés comme à haut risque) et d'évaluation (réalisation d'études d'impacts sur les droits fondamentaux et obligation de maîtriser les risques des systèmes d'IA tout au long de leur cycle de vie).

Par ailleurs, l'article 50 du RIA introduit des obligations de transparence pour les systèmes d'IA qui ne sont pas considérés à haut risque.

117 *Loi n°2016-1321 du 7 octobre 2016 pour une République numérique.*
118 *Avis CADA n°20144578.*
119 *Plus spécifiquement, les administrations qui emploient plus de cinquante personnes en équivalents temps plein, à l'exclusion des collectivités de moins de 3 500 habitants.*

CONCLUSION

Ce panorama des mythes et des enjeux révèle une chose : si l'engouement autour de l'IA est grand, les bénéfices de l'IA n'ont rien d'une évidence. Cela nous invite à considérer les outils d'intelligence artificielle non pas comme de simples outils techniques, mais bien dans le contexte des politiques publiques dans lesquelles ils s'inscrivent. C'est aussi un appel à bien anticiper les enjeux, à évaluer les usages de manière continue, et à développer des systèmes de gouvernance robustes au sein des collectivités, qui impliquent à la fois des expertises technique et métier, et prennent en compte la voix des citoyens et citoyennes.

Les collectivités s'emparent progressivement du sujet, en France et à l'étranger. La seconde partie de ce rapport visera à mettre en avant des bonnes pratiques, afin d'outiller concrètement les collectivités pour faire face aux enjeux identifiés.

Bibliographie

- *ADEME Magazine. (Avril 2022). Numérique : quel impact pour l'environnement ?*
- *Ananny, M. et Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society.*
- *ANSSI. (2024). Recommandations de sécurité pour un système d'IA générative.*
- *Barot, C. (16 octobre 2024). Croire ou douter : la question des biais de confiance dans la prise de décision. Linc - CNIL.*
- *Balova, A. et Kolbas, N. (2023). Biodiversity and Data Centers: What's the connection?. Ramboll Group.*
- *Biret, V. (18 août 2020). Mal notés au bac par un algorithme, les lycéens britanniques crient au scandale. Ouest-France.*
- *Blandin, N. (2024, 21 septembre). Aux États-Unis, Microsoft veut relancer une centrale nucléaire arrêtée. Ouest-France.*
- *Borel, S. (2024). Étude des impacts de l'IA sur le travail. Rapport d'enquête du LaborIA Explorer.*
- *Booth, R. (2025, 27 janvier). AI prototypes for UK welfare system dropped as officials lament 'false starts'. The Guardian.*
- *Bovens, M. et Zouridis, S. (2002). From Street-Level to System-Level Bureaucracies: How Information and Communication Technology Is Transforming Administrative Discretion and Constitutional Control. Public Administration Review.*
- *Burgess, M. (2024, 8 août). Microsoft's AI Can Be Turned Into an Automated Phishing Machine. WIRED.*
- *Burke, G., Schellmann, H. (26 octobre 2024). Researchers say an AI-powered transcription tool used in hospitals invents things no one ever said. AP News.*
- *CNIL. (2017). Comment permettre à l'Homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle. p.20*
- *CNIL, CADA, Etalab. (2019). Guide pratique de la publication en ligne et de la réutilisation des données publiques (open data).*
- *Commission d'accès aux documents administratifs. Rapport d'activité 2022-2023.*
- *Conseil économique, social et environnemental. (2025). Analyse de controverses : intelligence artificielle, travail et emploi.*
- *Courmont, A. (2018). Plateforme, big data et recomposition du gouvernement urbain : les effets de Waze sur les politiques de régulation du trafic. Revue française de sociologie, 59-3. Pp. 423-449.*
- *Courmont, A. (9 juillet 2020). Angèle Christin : «Les méthodes ethnographiques nuancent l'idée d'une justice prédictive et entièrement automatisée». Linc - CNIL.*
- *Défenseur des droits. (2024). Algorithmes, systèmes d'IA et services publics : quels droits pour les usagers ? Points de vigilance et recommandations.*
- *Demoures, C. (2024). Un outil de cartographie des métiers concernés par l'intelligence artificielle dans les collectivités. Une étude des élèves de l'INET 2023-2024.*
- *Dent, A. (2024). Automating Public Services: A Careful Approach. Promising Trouble.*
- *Diguet, C. (2024, 8 juillet). «Les data centers s'implantent de manière totalement opportuniste». La Croix.*
- *Dratsch, T., Chen, X., Mehrizi, M. R., Kloeckner, R., Mähringer-Kunz, A., Püsken, M., Baeßler, B., Sauer, S., Maintz, D., & Santos, D. P. D. (2023). Automation Bias in Mammography : The Impact of Artificial Intelligence BI-RADS Suggestions on Reader Performance. Radiology, 307(4).*
- *Dressel, J. and Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science Advances.*
- *Dumont, A. (2019, 11 décembre). « C'est inhumain, l'aliénation maximum » : dans les entrepôts logistiques de la région lyonnaise. Médiacités.*
- *Elish, M. C. (2019). Moral Crumple Zones : Cautionary Tales in Human-Robot Interaction. Engaging Science Technology And Society, 5, 40-60.*
- *Elish, M.C. et Watkins, E. A. (2020). Repairing Innovation: A Study of Integrating AI in Clinical Care. Data & Society.*
- *ENA. Promotion 2018-2019 «MOLIÈRE». (2019). Rapport collectif sur commande d'une administration centrale Ethique et responsabilité des algorithmes publics Groupe n° 12. École nationale d'administration.*
- *Etalab. (2021). Les algorithmes publics : pourquoi et comment les expliquer ?*
- *Eubanks, V. (16 février 2018). A response to Allegheny County DHS.*
- *Ferguson, Y., Pecoste. C.. (2022). L'IA au travail: propositions pour outiller la confiance. Conférence*
- *Nationale sur les Applications Pratiques de l'Intelligence Artificielle (APIA-PFIA).*
- *France Stratégie. (2016). Anticiper les impacts économiques et sociaux de l'Intelligence Artificielle. Annexe 1 : L'intelligence Artificielle en quête d'acceptabilité et de confort.*
- *Gender Shades. How well do IBM, Microsoft, and Face++ AI services guess the gender of a face?.*
- *Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. Computer Law & Security Review, 45, 105681.*
- *Guillaud, H. (13 novembre 2024). IA aux impôts : vers un «service public artificiel» ?. Dans les algorithmes.*
- *Huttner, L.. (2024). La décision de l'algorithme, étude de droit privé sur les relations entre l'humain et la machine. Dalloz.*
- *Interconnectés. (2024). Data et IA : les nouvelles règles du jeu en Europe.*

- *Interconnectés, Institut du Numérique Responsable, Banque des Territoires. (2023). Guide méthodologique : La stratégie numérique responsable de la collectivité en 10 étapes.*
- *InternetActu. (19 janvier 2018). Automatiser les inégalités.*
- *Kachalova, E. (2023, 12 avril). Can't take it back : Samsung ChatGPT fiasco highlights generative AI privacy woes. AdGuard Blog.*
- *Kaye, K. and Dixon, P. (2023). Risky Analysis: Assessing and Improving AI Governance Tools. World Privacy Forum.*
- *Leautier, A. (2024). Incursion dans la caisse à outils de la sécurité des grands modèles de langage (1/2) : risques visés et sécurité dès la conception. Linc CNIL.*
- *Lecher, C. (29 mars 2024). NYC's AI Chatbot Tells Businesses to Break the Law. The Markup.*
- *Leufer, D. (2020). Myth: AI can be objective or unbiased. AI Myths.*
- *Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). Making AI Less «Thirsty» : Uncovering and Addressing the Secret Water Footprint of AI Models. arXiv (Cornell University).*
- *Liu, N. F., Zhang, T., Liang, P. (2023). Evaluating Verifiability in Generative Search Engines. Findings of the Association for Computational Linguistics: EMNLP 2023.*
- *Luccioni, S., Jernite, Y., & Strubell, E. (2024, June). Power Hungry Processing: Watts driving the cost of AI deployment? In the proceedings of 2024 ACM FAccT Conference. pp. 85-99.*
- *Maxwell, W. (2022). «Le contrôle humain des systèmes algorithmiques – un regard critique sur l'exigence d'un «humain dans la boucle», mémoire original pour présenter l'habilitation à diriger des recherches, droit. Université Paris 1 Panthéon Sorbonne.*
- *Ministère de l'Économie, des Finances, et de la Souveraineté industrielle et numérique. (21 décembre 2023). Lettre de la DAJ – Notion de décision individuelle automatisée : la CJUE se prononce sur l'établissement automatisé de la capacité à rembourser un crédit.*
- *Observatoire Data Publica. (2024). Baromètre de l'Observatoire Data Publica 2024, 3e édition : les collectivités territoriales et la donnée.*
- *OCDE. (2023). Intelligence artificielle, qualité des emplois et inclusivité. in Perspectives de l'emploi de l'OCDE 2023. Éditions OCDE.*
- *O'Neil, C. (2018). Algorithmes : la bombe à retardement. Les Arènes.*
- *OpenGlobalRights. (19 mars 2020). Aux Pays-Bas, une décision judiciaire historique sur les États-providences numériques et les droits humains.*
- *Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: an attentional integration. Human factors, 52(3), 381–410.*
- *Pauron, M. (2025, 27 janvier). «La République démocratique du Congo est une réserve pour les dominants». Médiapart.*
- *Perrigo, B. (2023, 18 janvier). Exclusive : OpenAI Used Kenyan Workers on Less Than $ 2 Per Hour to Make ChatGPT Less Toxic. TIME.*
- *Piquard, A., Pinaud, O. et Pécout, A. (14 juin 2024). Derrière l'IA, la déferlante des data centers. Le Monde.*
- *Powles, J., Nissenbaum, H. (2018). The Seductive Diversion of 'Solving' Bias in Artificial Intelligence. OneZero.*
- *Pruss, D. (2023). Ghosting the Machine : Judicial Resistance to a Recidivism Risk Assessment Instrument. 2022 ACM Conference On Fairness, Accountability, And Transparency, 312-323.*
- *Public Law Project. (2024). Securing meaningful transparency of public sector use of AI: Comparative approaches across five jurisdictions.*
- *Raji, D. I., Kumar, E., Horowitz, A. and Selbst, A. (2022). The Fallacy of AI Functionality. In 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea.*
- *Redden, J., Brand, J., Sander, I. and Warne, H. (2022). Automating Public Services: Learning from Cancelled Systems. Data Justice Lab.*
- *Roussilhe, G. (2023). Territorialiser les systèmes numériques, l'exemple des centres de données [4/4]. Gauthier Roussilhe.*
- *Sánchez-Monedero, J., Dencik, L., & Edwards, L. (2020). What does it mean to 'solve' the problem of discrimination in hiring? social, technical and legal perspectives from the UK on automated hiring systems. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 458–468). Association for Computing Machinery.*
- *Upwork. (23 juillet 2024). Upwork Study Finds Employee Workloads Rising Despite Increased C-Suite Investment in Artificial Intelligence.*
- *UNESCO, IRCAI. (2024). Challenging systematic prejudices: an Investigation into Gender Bias in Large*
- *Language Models.*
- *Vallet, F. (2 juin 2020). Philippe Besse : «Les décisions algorithmiques ne sont pas plus objectives que les décisions humaines». Linc - CNIL.*
- *Vie-publique.fr. (2 janvier 2025). La notion de service public.*
- *Vie-publique.fr. (25 juillet 2024). Compétence liée, pouvoir discrétionnaire ou d'appréciation, comment agit l'administration ?*
- *Vallet, F. (6 janvier 2021). Entretien avec Clément Henin, Daniel Le Métayer : «Fournir des explications du fonctionnement des algorithmes compréhensibles par des profanes». Linc - CNIL.*
- *Vuarin, L., & Steyer, V. (2023). L'injonction à la transparence : un levier réglementaire à double tranchant pour les organisations. Bulletin de l'Association Française pour l'Intelligence Artificielle.*

- *Wachter, S. (2024). Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond. Yale Journal of Law & Technology, Volume 26, Issue 3.*
- *Wachter, S., Mittelstadt, B., Russel, C. (2021). Bias Preservation in Machine Learning: The Legality of Fairness Metrics under EU Non-Discrimination Law. West Virginia Law Review, Vol. 123, No. 3.*
- *Walker, B. (2024, 8 mars). OpenAI's GPT Shows Bias in Résumé Ranking Experiment. Bloomberg.*
- *Weidinger, L. and others. (2022). Taxonomy of Risks Posed by Language Models. In 2022 ACM Conference on Fairness, Accountability, and Transparency.*
- *Wieringa, M. (2020). What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). Association for Computing Machinery, New York, NY, USA, 1–18.*
- *Wilson, C. AI worse than humans in every way at summarising information, government trial finds. Crikey.*

Sommaire