

Lutter contre les discriminations produites par les algorithmes et l'IA

Dernière mise à jour : 01.02.2024

Les obligations de transparence et de traçabilité, qui dérivent du droit de la non-discrimination, valent aussi pour les outils numériques¹, et notamment les décisions prises via des algorithmes que les organisations publiques comme privées tendent de plus en plus à utiliser. Ces outils constituent souvent de véritables boîtes noires pour les personnes concernées alors même que les systèmes extrêmement complexes dits d'intelligence artificielle (IA) intègrent et rendent invisibles de nombreux biais discriminatoires.

1. La production de discriminations par les dispositifs algorithmiques et l'IA

Comme le soulignait le Défenseur des droits dans sa Déclaration commune avec la CNIL en 2020, « Algorithmes : prévenir l'automatisation des discriminations »², des recherches ont mis à jour, derrière l'apparente neutralité des algorithmes, l'ampleur des biais qui peuvent intervenir lors de leur conception et leur déploiement. Tout comme les bases de données qui les alimentent, ils sont conçus et générés par des humains dont les stéréotypes, en se répétant automatiquement, peuvent engendrer des discriminations.

La mécanique discriminatoire peut reposer sur le caractère biaisé des données sélectionnées et utilisées par l'algorithme. Ce caractère biaisé peut être lié à un manque de représentativité des données par rapport au contexte dans lequel l'algorithme va être déployé. Il peut aussi être lié au fait que les données sont la traduction mathématique de pratiques et comportements passés souvent discriminatoires et des discriminations systémiques opérant au sein de la société. Par ailleurs, la mobilisation de critères neutres en apparence (c'est-à-dire ne relevant pas des motifs prohibés de discrimination) peut avoir des effets discriminatoires. Le plus souvent, c'est la combinaison de plusieurs critères neutres qui peut emporter des effets discriminatoires. Les critères et données en question peuvent même sembler très lointains des motifs prohibés, mais leur corrélation permet des résultats similaires à ceux que l'on aurait obtenus en utilisant directement la caractéristique protégée. Les algorithmes apprenants, et les multiples corrélations de données massives qu'ils permettent, peuvent facilement

¹ [Défenseur des droits, Guide « Recruter avec des outils numériques sans discriminer », 2015.](#)

² [CNIL/Défenseur des droits, Algorithmes : prévenir l'automatisation des discriminations, 2020.](#)

engendrer de tels effets. Ces biais sont néanmoins largement invisibilisés et souvent impossibles à identifier à l'échelle individuelle.

Les obstacles peuvent intervenir sur des algorithmes simples et fermés, tels que Parcoursup qui conditionne l'accès à l'enseignement supérieur : du fait des limites légales posées au principe de transparence pour les algorithmes dits « d'aide à la décision », utilisés par les universités et écoles, qui ont été confirmées par le Conseil d'État et par le Conseil constitutionnel, il peut être difficile de retracer les critères de décision et le poids de chacun de ces critères et d'en contester les biais éventuels. Dans ses décisions de 2019, le Défenseur des droits a pu souligner le manque de transparence des algorithmes locaux utilisés par certains établissements dans le cadre du système Parcoursup pour traiter les afflux de demandes dans les filières les plus demandées³. De plus, la prise en compte du critère, apparemment neutre, de l'établissement d'origine, mobilisé par certains algorithmes locaux, peut être assimilée à une pratique discriminatoire fondée sur le lieu de résidence s'il conduit à favoriser automatiquement les candidats issus des lycées les plus prestigieux au détriment de ceux moins réputés, en particulier ceux situés dans des zones défavorisées⁴.

À l'heure où le déploiement des technologies biométriques s'accélère, le Défenseur des droits a publié un rapport en 2021 pointant les risques considérables de violation du principe de non-discrimination et, plus généralement, des droits fondamentaux que les technologies biométriques, telles que les algorithmes de reconnaissance faciale, représentent pour les personnes qui y sont exposées⁵. Par nature probabiliste, leur utilisation peut entraîner des erreurs aux conséquences multiples selon leur domaine d'usage, mais potentiellement graves (refus d'accès à un lieu, à un emploi, arrestation erronée, etc.). Au-delà du risque inhérent d'atteinte au droit au respect de la vie privée et à la protection des données, l'utilisation même d'outils biométriques d'identification et d'évaluation, dont les études ont démontré les biais avec des taux d'erreur plus importants pour les femmes, les personnes noires ou en situation de handicap, peut générer et amplifier les discriminations.

2. L'action du Défenseur des droits pour prévenir les discriminations algorithmiques, les faire corriger et sanctionner leurs auteurs

Le Règlement général sur la protection des données de l'UE (RGPD) et les textes européens et français apportent un cadre encore lacunaire aux risques majeurs de discrimination. Le développement phénoménal des technologies algorithmiques et des systèmes apprenants impose, pour les institutions et les entreprises, de maintenir une grande vigilance quant aux conséquences de ces évolutions technologiques, mais aussi de les anticiper, pour permettre

³ [Décision 2019-021 du 18 janvier 2019 relative au fonctionnement de la plateforme nationale de préinscription en première année de l'enseignement supérieur \(Parcoursup\)](#).

⁴ J. Grenet, « Les algorithmes d'affectation dans le système éducatif français », in M. Simioni, Ph. Steiner (dir.), *Comment ça matche ? Une sociologie de l'appariement*, Presses de Sciences Po, coll. « Gouvernances », pp. 21-59.

⁵ [Défenseur des droits, Rapport « Technologies biométriques : l'impératif respect des droits fondamentaux », 2021](#).

au débat démocratique de se tenir de manière éclairée tout en pensant un cadre juridique et une régulation protecteurs des droits et libertés.

Dans ce cadre, le Défenseur des droits a recommandé, depuis 2020, de prendre des mesures pour éviter que les discriminations soient reproduites et amplifiées par ces technologies, et notamment de former et sensibiliser les professionnels des métiers techniques et d'ingénierie informatique aux risques discriminatoires des algorithmes. En aval, dans une perspective de responsabilisation, la formation des organisations se procurant des algorithmes s'avère nécessaire puisqu'elles les utilisent aujourd'hui bien souvent sans se préoccuper du fonctionnement du système et des risques discriminatoires induits⁶.

Ces recommandations trouvent un écho dans les résultats de l'enquête « Perception du développement des technologies biométriques en France – Entre manque d'information et demande d'encadrement » publiée par le Défenseur des droits en 2022⁷, d'où ressortaient quatre enseignements majeurs :

- L'existence d'un important déficit d'information du public ;
- Un degré de confiance variable en fonction des entités responsables des déploiements ;
- Une prise de conscience des risques d'atteintes aux droits ;
- Une forte volonté de voir l'encadrement juridique existant renforcé.

Alors que le Conseil de l'Europe travaille à un projet de Convention-cadre sur la conception, le développement et l'utilisation des systèmes d'IA, l'Union européenne a avancé sur son projet de règlement sur l'IA qui devrait être adopté courant 2024 (IA Act). Ce projet incarne la tentative de la Commission européenne de garantir le respect des droits fondamentaux tout en encourageant l'innovation dans le domaine de l'IA. L'objectif affiché du projet est « d'asseoir une vision européenne de l'IA basée sur l'éthique en prévenant les risques inhérents à ces technologies par un règlement commun permettant d'éviter certaines dérives ». Ce texte présente de forts enjeux non seulement parce qu'il sera d'application directe en France (il s'agit d'un règlement) mais aussi parce qu'il vise à encadrer de manière horizontale et transversale toutes les différentes applications de l'IA (sauf exception).

À travers une approche basée sur les risques véhiculés par les différents systèmes d'IA, ce texte organise des contraintes distinctes selon les systèmes en question. Il prévoit notamment l'interdiction explicite de certaines formes d'IA, désignées comme étant contraires aux valeurs de l'Union européenne, et consacre une section entière à la réglementation des systèmes dits « à haut risque ». La proposition prévoit que les fournisseurs de ce type de systèmes seront soumis à diverses exigences en matière de transparence et d'évaluation des risques qu'ils devront remplir avant de mettre ces systèmes en service, ce qui crée une incitation à promouvoir des approches de conformité dès la conception et rejoint certaines des recommandations formulées par le Défenseur des droits.

⁶ Le Défenseur des droits participe à cet enjeu d'acculturation en organisant chaque année depuis 2021, avec le soutien du Conseil de l'Europe, une formation en ligne sur l'intelligence artificielle et les risques de discrimination.

⁷ [Défenseur des droits, Enquête « Perception du développement des technologies biométriques en France – Entre manque d'information et demande d'encadrement », 2022.](#)

Le Défenseur des droits a suivi les négociations sur ce texte et défend une approche basée sur le respect des droits fondamentaux permettant de lutter efficacement contre les discriminations, comme en témoignent les prises de position suivantes :

- Un avis intitulé « Pour une IA européenne protectrice et garante du principe de non-discrimination » publié en 2022, réalisé conjointement avec le réseau Equinet (réseau européen des organismes de promotion de l'égalité), dont il est membre⁸. Les recommandations émises dans cet avis s'inscrivent dans la continuité des travaux antérieurs de l'institution en soulignant la priorité de lutter contre les discriminations algorithmiques et insistent sur le rôle que pourraient jouer les organismes de promotion de l'égalité européens dans ce cadre ;

- Un deuxième avis intitulé « Intelligence artificielle : la Défenseure des droits appelle à garantir le droit de la non-discrimination dans les discussions européennes », publié en septembre 2023 et qui promeut trois principes clés⁹ :

- o Faciliter le recours pour les personnes victimes de discrimination liée à un système d'IA via les organismes nationaux de promotion de l'égalité tels que le Défenseur des droits ;

- o Assurer une application harmonisée et cohérente du principe de non-discrimination par un mécanisme de pleine coopération entre les futures autorités nationales de supervision et les organismes nationaux de promotion de l'égalité, notamment pour le traitement des plaintes concernant des systèmes d'IA présentant des risques pour les droits fondamentaux, mais également pour la formation du personnel des autorités de supervision sur la non-discrimination, et en prévoyant des études d'impact sur les droits fondamentaux pour les systèmes d'IA dits « à haut risque » ;

- o Favoriser la capacité des personnes affectées par un système d'IA à demander une protection contre la discrimination, notamment en s'assurant qu'elles disposent d'une information claire et complète et en prévoyant que les organismes nationaux de promotion de l'égalité pourront saisir les futures autorités de supervision ;

- Une prise de position conjointe avec la CNCDH appelant, en novembre 2023, en lien avec leurs homologues européens, à la nécessaire prise en compte des droits fondamentaux et du principe de non-discrimination, dans le cadre des dernières négociations sur le texte, avec quatre points essentiels¹⁰ :

- o Ajouter à la liste des « systèmes interdits » les systèmes biométriques permettant l'identification des personnes dans l'espace public, catégorisant les personnes et « reconnaissant » leurs émotions, ainsi que les systèmes de police prédictive. Ces systèmes présentent des risques majeurs d'atteinte aux droits fondamentaux des individus ;

- o Garantir les droits de recours individuels et collectifs auprès des futures autorités nationales compétentes et du bureau européen de l'IA. Ces garanties procédurales sont essentielles pour assurer la cohérence avec les recours accordés aux individus ou aux groupes en vertu des droits fondamentaux de l'UE et du droit de la non-discrimination ;

⁸ [Défenseur des droits, avis « Pour une IA européenne protectrice et garante du principe de non-discrimination », 2022.](#)

⁹ [Défenseur des droits, avis « Intelligence artificielle : la Défenseure des droits appelle à garantir le droit de la non-discrimination dans les discussions européennes », 2023.](#)

¹⁰ [Défenseur des droits/CNCDH, Communiqué conjoint, novembre 2023.](#)

- o Retenir les propositions du Parlement européen relatives aux analyses d'impact sur les droits fondamentaux pour les entités qui déploient les systèmes d'IA. Ces propositions apparaissent les seules à même de prévenir et protéger contre les discriminations et les atteintes aux droits fondamentaux d'une manière globale, transparente et cohérente ;
- o Prévoir un contrôle solide et indépendant des « modèles de fondation ». Un modèle de fondation est un modèle d'IA de grande taille, entraîné sur une quantité massive de données brutes. Les dangers inhérents à ce type d'IA sont largement documentés. Un contrôle solide et indépendant de ces systèmes apparaît comme la condition préalable à la prévention des discriminations et à la protection des droits fondamentaux. Pour autant, lors des dernières étapes de négociations du projet d'IA Act, le cadre envisagé ne reposerait que sur des codes de conduite et une autorégulation par les acteurs, sans sanction immédiate en cas de violation.

Le Défenseur des droits compte prendre toute sa part dans la préparation de l'application du texte (qui sera progressive jusqu'au 24^e mois après son adoption) et une fois celui-ci pleinement applicable.

Pour en savoir plus :

- [Avis « Intelligence artificielle : la Défenseure des droits appelle à garantir le droit de la non-discrimination dans les discussions européennes », 2023](#)
- [Enquête « Perception du développement des technologies biométriques en France – Entre manque d'information et demande d'encadrement », 2022](#)
- [Avis « Pour une IA européenne protectrice et garante du principe de non-discrimination », 2022](#)
- [Rapport « Technologies biométriques : l'impératif respect des droits fondamentaux », 2021](#)
- [CNIL/Défenseur des droits, *Algorithmes : prévenir l'automatisation des discriminations*, 2020](#)
- [Guide « Recruter avec des outils numériques sans discriminer », 2015](#)