

Consultation publique

Fiches pratiques sur l'information et l'exercice des droits pour le développement de systèmes d'IA

Synthèse des contributions

Février 2025

La CNIL a lancé, le 10 juin 2024, une consultation publique sur le développement de systèmes d'IA.

Les contributions ont nourri les travaux du deuxième lot de huit fiches pratiques, dont celles sur l'information des personnes et l'exercice des droits pour leur publication définitive sur le site de la CNIL.

La synthèse en chiffres

Le 8 avril 2024, à la suite d'une consultation publique, **la CNIL a publié [un premier lot de fiches pratiques sur le développement des systèmes d'IA](#)**. Elles répondent en particulier aux interrogations portant sur l'application des principes de finalité, de minimisation et de durée de conservation pour le développement des systèmes d'IA et précisent certaines règles applicables à la recherche scientifique, à la réutilisation de bases de données ou à la réalisation d'analyse d'impact sur la protection des données (AIPD).

Dans la continuité de ces travaux, **la CNIL a lancé, le 10 juin 2024, [une nouvelle consultation publique sur un questionnaire concernant l'application du RGPD aux modèles d'IA et les 7 fiches suivantes](#)** :

- Base légale de l'intérêt légitime et développement de systèmes d'IA.
- Intérêt légitime : focus sur la diffusion des modèles en source ouverte (open source).
- Intérêt légitime : focus sur le moissonnage (*web scraping*).
- Informer les personnes concernées.
- Respecter et faciliter l'exercice des droits des personnes concernées.
- Annoter les données.
- Garantir la sécurité du développement d'un système d'IA.

Cette synthèse se concentre sur les retours qui ont concerné les fiches « Informer les personnes concernées » et « Respecter et faciliter l'exercice des droits des personnes concernées ».

À l'issue de la consultation publique, le 1^{er} octobre 2024, la CNIL a reçu 62 contributions représentatives de différents secteurs :

- des organismes à but lucratif :
 - 28 entreprises privées, dont des fournisseurs de systèmes d'IA, des utilisateurs de systèmes d'IA, ou encore d'autres entreprises de différents secteurs (IA, finance, santé, aéronautique, opérateurs de plateformes en ligne, publicité en ligne, audiovisuel, etc.) ;
 - 11 organisations représentantes des professionnels ;
 - 4 cabinets d'avocat ou de conseil.
- des organismes à but non-lucratif :
 - 3 associations représentatives de la société civile ;
 - 5 chercheurs ou groupement de chercheurs académiques et 1 *think tank* ;
- 6 particuliers ;
- 4 administrations publiques.

Ces contributions ont permis à la CNIL :

- de faire évoluer ses projets de fiches en apportant des clarifications supplémentaires et consolider ses analyses ;
- d'apporter des réponses, dans la synthèse ci-dessous, aux préoccupations les plus fréquemment partagées par les contributeurs sur l'information et les droits des personnes.

Fiche sur l'information des personnes

A. Adopter une position équilibrée sur l'information relative aux sources des données, préservant l'objectif de la réglementation

Synthèse des contributions

De nombreuses contributions ont remis en cause le fait de devoir fournir une information « suffisamment précise » sur les sources des données (notamment les noms de domaine et adresses URL des pages moissonnées et les moyens de contact des responsables de traitement des jeux de données), pour les motifs suivants :

- cette exigence ne serait pas explicite dans l'article 14 du RGPD ;
- le niveau de transparence des données d'entraînement des modèles d'IA à usage général devrait être déterminé par le modèle de résumé de l'article 53 du règlement sur l'intelligence artificielle (RIA) qui sera adopté par le bureau de l'IA ;
- ce niveau de précision porterait atteinte à d'autres droits, tels que les droits de propriété intellectuelle et secrets commerciaux ;
- le niveau de détail devrait dépendre des risques pour les droits personnes (au regard du contexte, des finalités et usages et des fonctionnalités du modèle, du volume des données) ;
- cela poserait des difficultés pratiques ou techniques ;
- cela pourrait augmenter le risque d'attaques visant à extraire certaines données du modèle.

À l'inverse, certaines contributions ont souligné le **besoin de transparence en rappelant que les personnes concernées ne sont généralement pas informées du traitement de leurs données à des fins de développement de systèmes d'IA** (inaccessibilité de l'information). Ces contributions ont souligné la pertinence des URLs et noms de domaines.

Réponse de la CNIL

La CNIL a modifié, en partie, l'approche retenue pour cette recommandation.

Elle rappelle que l'objectif de l'obligation d'information est de permettre aux personnes de déterminer si elles sont concernées ou non par le traitement.

En cas de collecte indirecte, les informations supplémentaires à fournir pour atteindre cet objectif diffèrent selon que le responsable du traitement doit fournir une information individuelle ou générale :

- **Lorsque le responsable du traitement doit délivrer une information individuelle**, il est généralement requis de fournir l'ensemble des informations prévues aux articles 13 ou 14 du RGPD. S'agissant des sources des données en cas de collecte indirecte des données, il devra fournir des informations précises sur celles-ci afin de permettre aux personnes concernées de comprendre si leurs données sont traitées et d'assurer son contrôle.
- En revanche, **lorsque l'information individuelle n'est pas possible**, parce que le responsable du traitement ne peut pas identifier les personnes présentes dans la base, ne dispose pas de données de contacts des personnes concernées ou encore que les contacter individuellement requerrait des efforts disproportionnés, des mesures appropriées doivent être prises. Il est nécessaire de publier une notice d'information, par exemple sur un site internet, qui comporte, si possible, les informations qui auraient été fournies en cas d'information individuelle. Toutefois, l'information sur les sources présente des difficultés particulières. Deux cas sont à distinguer :
 - Lorsque le responsable du traitement a utilisé un nombre limité de sources pour constituer sa base de données d'entraînement, il est généralement requis qu'il fournisse des indications précises sur ces sources, sauf exception dont il pourrait justifier.

- Lorsque de nombreuses sources sont utilisées, par exemple un grand nombre de sources accessibles au public en ligne, une information globale, indiquant par exemple leurs catégories, voire les noms de quelques sources principales ou typiques, est généralement suffisante.

La CNIL a également ajouté des précisions relatives à l'articulation avec les exigences du RIA sur la publication d'un résumé suffisamment détaillé des données d'entraînement des modèles d'IA à usage général prévu par l'article 53 du Règlement. Elle précise que **ce résumé pourra également, en principe, servir d'information générale sur les sources de données, pour ces types de modèles d'IA, pour l'application du RGPD sous réserve de la publication définitive de ce résumé par le Bureau de l'IA.**

Enfin, la CNIL a également ajouté des exemples de mentions d'information suffisamment précises et recommande uniquement, à titre de bonne pratique, d'informer sur les noms de domaine et URL des pages web moissonnées.

B. Illustrer les efforts disproportionnés permettant de déroger à une information individuelle

Synthèse des contributions

De nombreuses contributions ont appelé à davantage de précisions et d'illustrations des cas dans lesquels fournir une information individuelle exigerait des efforts disproportionnés.

Certaines ont considéré que **le caractère disproportionné d'une information individuelle devrait être présumé :**

- pour les grands modèles de langage ou les modèles d'IA générative, en raison du volume de données, des efforts à fournir et le risque d'intrusivité pour les personnes pour obtenir leurs données de contact ;
- pour le moissonnage ou lorsque le responsable du traitement ne dispose pas des coordonnées des personnes concernées.

À l'inverse, **certains contributeurs ont considéré que la fourniture d'une information générale ne doit pas devenir le principe :** le caractère disproportionné des efforts doit toujours être démontré par le responsable du traitement, au motif que le volume des données d'entraînement de modèles d'IA générative ne saurait, seul, présumer de ce caractère disproportionné.

Réponse de la CNIL

La CNIL considère que le volume des données ne saurait présumer, à lui seul, du caractère disproportionné d'une information individuelle.

Ce sera toutefois le cas pour la collecte licite de données ne présentant pas de risques pour les personnes concernées, à partir d'un grand nombre de sites web à des fins de développement d'un grand modèle de langage, et dont les personnes concernées ne peuvent ignorer qu'elles sont publiquement accessibles, tels que des encyclopédies en sources ouvertes. **La CNIL a inclus un exemple relatif à la collecte de données d'entraînement d'un modèle de fondation qui pourrait se prévaloir de cette dérogation à l'information individuelle.**

C. S'assurer du respect des exigences formelles de l'information

1. Sur les modalités alternatives d'information

Synthèse des contributions

Plusieurs contributions ont souhaité **promouvoir des modalités d'informations alternatives**, telles que les cartes de modèles, les cartes de systèmes, des foires aux questions ou autre documentation, des articles de blog, etc.

Réponse de la CNIL

La CNIL rappelle, dans sa fiche, que s'il est envisageable de fournir ces informations au sein de modèles de documentation existants (comme les cartes de données, de modèles ou de systèmes d'IA), elles doivent être facilement accessibles, claires et compréhensibles pour les personnes concernées. Cela implique que les informations ressortent clairement de ces documentations et qu'elles ne soient pas confondues avec elles.

Ces éléments de transparence supplémentaires sont également à encourager et peuvent constituer des mesures additionnelles pour la mise en balance au titre de l'intérêt légitime, qui seront détaillées dans la fiche correspondante.

2. Sur la distinction entre l'information relative aux traitements pour le développement et les autres traitements

Synthèse des contributions

Certaines contributions ont contesté la pertinence de la recommandation consistant à distinguer clairement l'information portant sur les traitements réalisés à des fins de développement, de celle concernant les autres traitements.

Réponse de la CNIL

La CNIL rappelle le besoin d'intelligibilité de l'information par catégorie de personnes concernées. **Elle conserve donc cette préconisation mais la transforme en bonne pratique.**

D. Permettre la vulgarisation et l'acculturation à l'IA via des recommandations et bonnes pratiques

Synthèse des contributions

Plusieurs contributions ont contesté devoir fournir des éléments d'information et d'acculturation qui ne sont pas explicitement prévus par les articles 13 et 14 du RGPD comme, par exemple, l'information sur les risques liés à la reconstruction des données à partir d'un modèle d'IA. Les contributeurs ont considéré que cela créerait de la confusion, voire de l'inquiétude pour les personnes, ou présenterait trop de complexité.

Plusieurs contributions ont également contesté les autres bonnes pratiques en matière de transparence (campagnes d'information, publication de l'AIPD, développement en source ouverte, etc.) au motif notamment que ces informations seraient confidentielles. A l'inverse, une contribution a encouragé le recours à des outils pédagogiques.

Réponse de la CNIL

Certaines explications résultent de l'importance du principe de transparence et de la nécessité de rendre l'information suffisamment claire et compréhensible par les personnes concernées, particulièrement lorsque les enjeux et risques ne sont pas toujours clairs du fait de l'évolution des technologies.

La CNIL conserve, par ailleurs, des bonnes pratiques préconisées en matière d'information. Certaines d'entre elles, qui pourraient constituer des garanties additionnelles dans le cadre de la mise en balance pour la mobilisation de l'intérêt légitime, apparaîtront dans les contenus dédiés, qui seront publiés ultérieurement.

E. Proposer un délai indicatif entre l'information et l'entraînement du modèle pour orienter les acteurs sur cette bonne pratique

Synthèse des contributions

La fiche mentionne la bonne pratique qui consiste, lorsque cela est approprié, à observer un délai raisonnable entre le moment où les personnes sont informées que leurs données sont contenues dans une base de données d'apprentissage, et l'entraînement d'un modèle sur cette base. Cette bonne pratique doit permettre aux personnes concernées de s'y opposer, surtout compte tenu des difficultés à exercer leurs droits sur le modèle une fois entraîné.

Plusieurs contributions ont appelé à **une clarification et à une illustration de ce délai**, en le modulant en fonction de la sensibilité des données et des secteurs. D'autres ont appelé à une certaine flexibilité à cet égard, voire à un retrait de cette bonne pratique en raison de ses difficultés pratiques.

Réponse de la CNIL

La CNIL considère que le respect d'un délai raisonnable constitue une bonne pratique qui peut faciliter le respect de l'exercice des droits des personnes. **Il appartient au responsable du traitement de juger du délai approprié selon le cas d'usage.**

Fiche sur l'exercice des droits

A. L'exercice des droits sur les données d'entraînement

1. Sur la collecte ou la conservation de données permettant d'identifier les personnes concernées

Synthèse des contributions

Si le rappel concernant l'article 11 du RGPD a bien été reçu par les contributions, un contributeur a rappelé que l'impossibilité doit être démontrée par les responsables de traitement. Seules deux contributions ont considéré que l'exercice des droits peut justifier la conservation de données d'identification dans certains domaines (en particulier au regard des risques pour les droits de propriété intellectuelle).

Sur la conservation des noms de domaine et adresses URLs pour faciliter l'identification des personnes concernées qui fourniraient des informations complémentaires, les retours sont partagés, compte tenu de la tension entre l'absence d'obligation de conserver les identifiants pour exercer les droits et la conservation de métadonnées pour faciliter l'éventuelle réidentification des personnes.

Réponse de la CNIL

La CNIL rappelle, tout d'abord, que l'article 11 du RGPD ne permet pas d'exiger la conservation des données d'identification uniquement pour répondre à l'exercice des droits. Les personnes concernées conservent, toutefois, la possibilité de fournir des informations complémentaires qui permettent de les identifier à cette fin.

Sur la conservation des noms de domaine et adresses URL, la CNIL a précisé que la conservation des adresses URL pour indexer la base de données d'apprentissage et permettre l'exercice des droits n'est pas obligatoire lorsque le responsable du traitement démontre qu'il n'est pas en mesure d'identifier les personnes. Elle a ajouté également l'exemple suivant dans le corps de la fiche pour illustrer qu'une telle indexation peut toutefois permettre de répondre à l'exercice des droits sans préjudice de la possibilité d'envisager d'autres moyens :

« Le concepteur d'un modèle d'IA ayant licitement collecté des données par moissonnage d'un nombre important de pages web constitue une base de données textuelles dont le volume très important rend

impossible la recherche par mots-clés. Afin de pouvoir vérifier la présence des données indirectement identifiantes des personnes exerçant leur droit d'accès, ce concepteur pourrait demander de fournir la ou les URL des pages web concernées et mettre en œuvre une indexation des URL visées par le moissonnage lui permettant de naviguer facilement dans la base (à la manière de l'indexation réalisée pour Common Crawl). La conservation des adresses URL des pages web moissonnées et l'indexation des données d'entraînement est un moyen d'atteindre cet objectif, toutefois le RGPD n'est pas prescripteur et le responsable de traitement n'est tenu que de fournir l'information disponible sur les sources dans le cadre du droit d'accès. »

De plus, **les responsables restent tenus de documenter la conformité de leur traitement, y compris s'agissant des sources de données utilisées** (quand bien même ils ne seraient pas en mesure d'identifier les sources de données précises de chaque personne).

2. Sur l'identification des personnes dans les données d'entraînement

Synthèse des contributions

Les contributions ont été divisées sur ce point : si une partie d'entre elles ont considéré que l'identification des personnes requiert des efforts disproportionnés, en particulier au regard des risques pour les personnes, d'autres ont souligné que les responsables du traitement doivent prendre en compte l'exercice des droits à chaque étape du développement et ne peuvent pas écarter ces obligations pour des motifs techniques.

Un fournisseur a estimé, de manière plus générale, qu'il n'existe pas de méthode permettant de lier des informations spécifiques à une personne en particulier avec une certitude raisonnable. Plus précisément, plusieurs fournisseurs ont souligné les difficultés de vérification de l'identité des personnes, en particulier dans les cas d'homonymie lorsque le fournisseur n'est pas en lien direct avec les personnes concernées. Sous un angle différent, une organisation non gouvernementale (ONG) a souligné l'importance de s'assurer que les données sur lesquelles le droit est exercé sont les bonnes.

Réponse de la CNIL

L'article 25 du RGPD prévoit que le droit à la protection des données doit être pris en compte dès la conception et par défaut, ce qui suppose d'anticiper les difficultés techniques. Les limitations sont spécifiquement prévues pour chacun des droits par les dispositions qui les encadrent.

Si la CNIL considère qu'il est possible de déroger au réentraînement des modèles lorsque celui-ci porterait une atteinte disproportionnée à d'autres droits fondamentaux, ces difficultés ne sont pas du tout les mêmes s'agissant de l'exercice des droits sur les bases de données d'entraînement elles-mêmes, en particulier lorsque la personne fournit des informations permettant de l'identifier.

Ce n'est d'ailleurs que si, et seulement si, l'organisme a des doutes raisonnables sur l'identité du demandeur, qu'il peut lui demander de joindre tout document permettant de prouver que les données en question sont bien les siennes. La CNIL a insisté dans la fiche sur cette notion de « doutes raisonnables » sur l'identité de la personne concernée.

3. Sur le droit d'opposition

Synthèse des contributions

De nombreuses contributions ont appelé à clarifier les motifs impérieux que les fournisseurs de systèmes d'IA peuvent valablement mettre en avant pour ne pas répondre favorablement à une demande d'opposition, en citant notamment la liberté d'entreprendre, la liberté d'information, l'amélioration du service et sa sécurisation, les difficultés techniques pour assurer l'exercice des droits, ou encore le risque pour la représentativité des données.

Une ONG a considéré que ces cas devraient rester exceptionnels.

Les contributions n'ont pas été uniformes sur les conditions d'exercice de ce droit pour les personnalités publiques, certaines considérant qu'elles ne devraient pas pouvoir s'opposer aux traitements et d'autres, à l'inverse, ont considéré que les risques sont d'autant plus importants pour elles.

Réponse de la CNIL

La CNIL a **illustré, de manière plus générale, l'appréciation du caractère disproportionné de l'exercice des droits sur un modèle d'IA.**

4. Sur la bonne pratique relative à la liste repoussoir pour le moissonnage

Synthèse des contributions

Plusieurs contributions ont souligné que la liste repoussoir ne serait pas requise par le RGPD et se révélerait techniquement difficile à mettre en œuvre. Certaines ont demandé des modalités plus précises de cette liste repoussoir.

Par ailleurs, deux universitaires ont regretté que cette liste conduise à autoriser un moissonnage en mettant la possibilité de s'opposer à la charge des personnes concernées.

Réponse de la CNIL

De manière générale, en cas de moissonnage de données accessibles en ligne, **la CNIL encourage le développement de solutions techniques qui permettront de faciliter le respect de l'exercice du droit d'opposition en amont de la collecte des données.** Il existe, pour certains traitements, des mécanismes de « liste repoussoir », qui pourraient être transposés lorsque c'est adapté au regard du traitement de données mis en œuvre.

5. Ajouter une bonne pratique pour apporter la preuve du respect des droits

Synthèse des contributions

Une contribution a proposé de **recommander aux fournisseurs de communiquer sur les mises à jour des bases de données** faisant suite à des demandes d'exercice de droit pour que les personnes concernées puissent contrôler que leurs demandes ont été respectées, par exemple dans la documentation du jeu de données ou sur le site web des fournisseurs.

Par ailleurs, un contributeur a suggéré que les responsables de traitement **informent les personnes concernées de ce que la modification de leurs données dans les jeux d'entraînement n'est pas nécessairement (ou immédiatement) répercutée dans le modèle ou dans le système d'IA déjà développé sur cette base.** Le cas échéant, cela leur permettrait d'exercer leurs droits directement sur les données contenues dans le modèle d'IA.

Réponse de la CNIL

La CNIL considère que constitue une bonne pratique le fait de communiquer sur les mises à jour des bases de données ou modèles, par exemple dans la documentation du jeu de données ou sur le site web des fournisseurs, pour permettre aux personnes concernées de savoir dans quelle mesure leurs demandes ont été respectées.

Elle considère également que la deuxième mesure est pertinente et recommande que les responsables de traitement informent clairement les personnes qui exercent leurs droits sur l'interprétation qu'ils font de leurs demandes (sur les données d'apprentissage ou sur le modèle d'IA) et sur la manière dont ils y répondent. Cela est particulièrement important lorsque la modification de leurs données dans les jeux d'entraînement n'est pas répercutée (ou du moins pas immédiatement) dans le modèle ou système d'IA déjà développé (et que cela est justifié par l'une des dérogations).

La CNIL a intégré ces éléments dans la fiche.

B. L'exercice des droits sur le modèle d'IA

1. L'application du RGPD aux sorties d'un système d'IA générative

Synthèse des contributions

La fiche entendait exclure de son champ les cas dans lesquels des données de sorties de systèmes ne sont pas soumises au RGPD ou relèvent de la responsabilité du déployeur du système.

Cette partie a suscité des réactions divergentes sur les enjeux de responsabilité et des interrogations sur le statut des données de sortie, sur la question de savoir si elles constituent ou non des données personnelles.

Réponse de la CNIL

La CNIL a **supprimé ces développements qui seront traités dans les contenus publiés ultérieurement sur l'application du RGPD aux modèles d'IA et ses conséquences sur la responsabilité des utilisateurs de modèles ou systèmes d'IA.**

2. Sur le désapprentissage machine

Synthèse des contributions

Plusieurs contributions ont reconnu que le désapprentissage machine est prometteur mais n'est pas assez mature. Une contribution a suggéré de mentionner le RAG (la génération augmentée de récupération), qui permet d'isoler les données personnelles dans un espace vectoriel plus facile à modifier ou reconstruire.

À cet égard, plusieurs contributions ont souligné le besoin de mesurer le niveau d'efficacité de la suppression (c'est-à-dire la capacité d'un modèle à supprimer des données ou des connaissances spécifiques d'une manière efficace et irréversible) qui pourrait satisfaire les demandes d'effacement de données.

Réponse de la CNIL

Dans certains cas, lorsque les paramètres du modèle contiennent explicitement certaines données d'apprentissage ou lorsque le système d'IA est connecté à une base de connaissance (comme notamment le RAG), il sera techniquement possible d'exercer les droits sur les paramètres du modèle. La CNIL a ajouté ces éléments dans la fiche et a précisé certaines considérations sur la distinction entre le modèle et le système où celui-ci est intégré pour distinguer où sont contenues les données personnelles et identifier qui est le responsable du traitement.

3. Sur le réentraînement d'un modèle d'IA

Synthèse des contributions

La fiche s'interrogeait sur les conditions dans lesquelles le réentraînement d'un modèle nécessiterait des efforts disproportionnés. Les contributions ont présenté des positions divergentes de la manière suivante :

- Un premier groupe de contributeurs a considéré, de manière générale, que le réentraînement nécessite des efforts disproportionnés, au moins pour les modèles d'IA à usage général compte tenu de ses besoins financiers, humains, opérationnels et environnementaux.
- Un deuxième groupe a considéré qu'un réentraînement à chaque exercice de droit ou tous les deux mois serait disproportionné, mais qu'un réentraînement périodique (par exemple semestriel) semblerait plus adapté. Une entreprise a souligné à cet égard que la plupart des modèles commerciaux sont régulièrement mis à jour (par opposition à ceux mis en source ouverte).

- Une contribution a contesté la possibilité même de déroger au réentraînement en cas d'atteinte disproportionnée à d'autres droits, en l'absence de fondement explicite dans le RGPD.
- Enfin, deux contributions ont appelé à une clarification des critères d'appréciation de la proportionnalité d'un réentraînement au cas par cas, dont l'une a proposé les critères suivants : le coût financier (en termes de ressources computationnelles), la fréquence des demandes d'opposition, de modification ou d'effacement, l'impact potentiel sur la performance du modèle, et la nature des données faisant l'objet d'une demande de droit et les conséquences pour la personne concernée.

Réponse de la CNIL

Les techniques permettant de garantir les droits des personnes sur leurs données dans les modèles d'IA sont en pleine évolution. Sans préjudice de méthodes efficaces qui pourront être développées par les acteurs, la CNIL identifie, à ce jour, deux techniques principales : le réentraînement et l'application de filtres (voir la partie B.4 ci-dessous).

La CNIL considère que lorsque le responsable du traitement dispose toujours des données d'entraînement, un réentraînement du modèle (qui peut avoir lieu de manière périodique) permet de répondre à l'exercice des droits. Subsidiairement, lorsque le réentraînement du modèle s'avère disproportionné (temporairement ou définitivement), la CNIL recommande la mise en œuvre d'autres types de mesures permettant de protéger les données et la vie privée des personnes. Si l'état de l'art évolue rapidement, la CNIL recommande en particulier l'utilisation de mesures consistant à filtrer les sorties d'un système pour permettre de répondre à l'exercice des droits à la rectification, à l'opposition ou à l'effacement si le responsable du traitement démontre qu'elles sont suffisamment efficaces et robustes (c'est-à-dire qu'elles ne peuvent pas être contournées).

La CNIL précise, en outre, les critères qui doivent être pris en compte pour apprécier le caractère proportionné ou non de la réponse à l'exercice d'un droit.

4. Sur le filtrage des sorties

Synthèse des contributions

Une majorité de contributions était favorable à la possibilité de répondre à des demandes d'exercice de droit par des mesures de filtrage. Deux associations professionnelles ont toutefois pointé les éventuels biais ou impacts de ces mesures sur la performance des systèmes et ont plaidé pour une concertation approfondie de la CNIL sur ces mesures techniques avant de formuler des recommandations. A l'inverse, une autre association professionnelle alerte sur le fait que l'efficacité de telles mesures est pour l'heure sujette à caution.

Réponse de la CNIL

La CNIL identifie, à ce jour, deux techniques principales : le réentraînement, qui permet d'effacer ou de rectifier des données au cœur du modèle et des filtres, qui permettent de remédier à certains effets du modèle sans supprimer les données du modèle lui-même. Ces recommandations sont sans préjudice de méthodes efficaces qui pourront être développées par les acteurs pour répondre à l'exercice des droits des personnes.

La CNIL recommande, à titre subsidiaire, lorsque le réentraînement s'avèrerait disproportionné, de mettre en œuvre d'autres types de mesures permettant de protéger les données et la vie privée des personnes. Si l'état de l'art évolue rapidement, la CNIL recommande en particulier l'utilisation de mesures consistant à filtrer les sorties d'un système pour permettre de répondre à l'exercice des droits à la rectification, à l'opposition ou à l'effacement si le responsable du traitement démontre qu'elles sont suffisamment efficaces et robustes.

La CNIL précise cette position dans la fiche.