# IA : Respecter et faciliter l'exercice des droits des personnes concernées

*07 février 2025*

*Les personnes dont les données sont collectées, utilisées ou réutilisées pour développer un système d'IA disposent de droits sur leurs données qui leur permettent d'en conserver la maîtrise. Il appartient aux responsables des traitements de les respecter et d'en faciliter l'exercice.*

*L'exercice des droits est étroitement lié à l'information donnée sur les traitements, qui constitue également une obligation.*

## Rappel général sur les droits applicables

Les personnes concernées disposent des droits suivants sur leurs données personnelles :

- droit d'accès (article 15 du RGPD) ;
- droit de rectification (article 16 du RGPD) ;
- droit à l'effacement, également appelé droit à l'oubli (article 17 du RGPD) ;
- droit à la limitation du traitement (article 18 du RGPD) ;
- droit à la portabilité des données lorsque la base légale du traitement est le consentement ou le contrat (article 20 du RGPD) ;
- droit d'opposition lorsque le traitement est fondé sur l'intérêt légitime ou la mission d'intérêt public (article 21 du RGPD) ;
- droit de retirer son consentement à tout moment lorsque le traitement de données à caractère personnel est fondé sur le consentement (article 7.3 du RGPD).

**Les personnes concernées doivent pouvoir exercer leurs droits à la fois** :

- **sur les bases de données d'apprentissage** et,
- **sur les modèles d'IA si ces derniers ne sont pas considérés comme anonymes** comme précisé dans l'avis du CEPD 28/2024 sur certains aspects de la protection des données liés au traitement des données à caractère personnel dans le contexte des modèles d'IA.

Si l'exercice d'un droit d'accès, de rectification ou d'effacement sur une base de données d'entraînement présente des problématiques assez comparables par rapport à d'autres grandes bases de données, l'exercice des mêmes droits sur le modèle d'IA lui-même présente des difficultés particulières et inédites. Cela doit conduire à **l'adoption de solutions réalistes et proportionnées, afin de garantir les droits des personnes sans empêcher l'innovation en matière d'intelligence artificielle. La CNIL estime ainsi que le RGPD permet d'appréhender les spécificités des modèles d'IA pour l'exercice des droits. La complexité et les coûts des demandes qui s'avèreraient disproportionnées sont ainsi des facteurs qui peuvent être pris en compte.**

## En matière de développement de modèles ou de systèmes d'IA : comment répondre aux demandes d'exercice de droits ?

En pratique, il est assez différent de répondre à des demandes d'exercice de droits selon qu'elles portent sur les données d'apprentissage ou sur le modèle lui-même.

La CNIL recommande que les responsables de traitement informent clairement les personnes exerçant leurs droits sur comment ils interprètent leur demande (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'entrainement 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 décrites ci-dessous).

### Pour l'exercice des droits sur les bases de données d'apprentissage

## Sur les difficultés pour identifier la personne concernée

**Lorsque le responsable du traitement n'a pas ou plus besoin d'identifier les personnes concernées, qu'il peut démontrer qu'il n'est pas en mesure de le faire, il peut l'indiquer en réponse aux demandes d'exercice de droit.**

Ce sera souvent le cas pour la constitution et l'utilisation de bases de données d'apprentissage. En effet, le fournisseur d'un système d'IA n'a en principe pas besoin d'identifier les personnes dont les données sont présentes dans son jeu d'entraînement. Lorsqu'il n'y a plus de données d'identification, certaines métadonnées et informations précises sur les sources de chaque donnée peuvent permettre de retrouver les personnes dans une base d'apprentissage mais, là encore, la conservation de ces informations n'est pas obligatoire (sur la question de l'information sur les sources, cf. supra). Par exemple :

- Un organisme n'a pas à conserver les visages d'une base d'images à la seule fin de permettre l'exercice des droits sur un jeu de photographies si le respect du principe de minimisation lui impose de flouter les visages.

- Un organisme n'a pas à conserver des données d'identification dans le jeu de données à la seule fin de permettre le retrait éventuel du consentement des personnes si elles ont été dûment informées de cet aspect et qu'elles y ont consenti de manière éclairée (par exemple en étant informées lorsque l'effacement des données mémorisées par le modèle d'IA sera impossible).

Lorsque le responsable n'est pas en mesure d'identifier les personnes concernées, il n'a pas non plus à conserver les données d'apprentissage pour répondre aux demandes d'exercice des droits (par exemple pour envisager de ré-entrainer un modèle d'IA comme cela est développé ci-dessous).

Toutefois, ce n'est pas parce que le responsable du traitement ne peut pas identifier de son propre chef les personnes concernées que cette identification est impossible. Dans ce cas particulier, le RGPD laisse la possibilité aux **personnes de fournir des informations complémentaires afin d'aider les responsable de traitement à les identifier et ainsi d'exercer leurs droits.**

À cet égard, le RGPD ne prévoit pas de forme particulière. Une personne concernée pourrait donc par exemple fournir une image ou un pseudonyme (tel qu'un nom d'utilisateur sous lequel la personne concernée a une activité en ligne). Ce n'est d'ailleurs que si et seulement si, l'organisme a des doutes raisonnables sur le fait que les données appartiennent bien au demandeur, qu'il peut lui demander de joindre tout document permettant de prouver que les données en question sont bien les siennes.

**Au titre du principe de protection des données dès la conception et de l'obligation de faciliter l'exercice des droits, il est recommandé d'anticiper ces difficultés d'identification en indiquant les informations complémentaires susceptibles de permettre l'identification des personnes concernées** (par exemple en cas d'homonymie). Il peut s'agir de leur permettre de fournir un fichier en particulier (comme une image, un enregistrement audio ou vidéo) mais aussi d'autres informations lorsque cela n'est pas possible.

Ces informations supplémentaires, fournies par les personnes, ne peuvent être traitées à d'autres fins que l'exercice de leurs droits

## S'agissant du droit de recevoir communication d'une copie des données d'apprentissage (au titre du droit d'accès)

**Le droit d'accès permet à toute personne d'obtenir gratuitement une copie de l'ensemble des données traitées la concernant.** Cela suppose le droit d'obtenir des extraits de bases de données d'apprentissage lorsque cela est indispensable pour permettre à la personne concernée d'exercer effectivement ses autres droits (CJUE, 4 mai 2023, aff. C-487/21). À cet égard, la CNIL recommande de fournir les données personnelles, y compris les annotations et métadonnées associées dans un format aisément compréhensible.

Cette communication ne doit pas porter atteinte aux droits et libertés d'autrui, lesquels incluent notamment les droits des autres personnes concernées, ou encore les droits de propriété intellectuelle ou du secret des affaires du titulaire de la base de données.

## S'agissant des informations sur le traitement de données (au titre du droit d'accès) :

Permettre aux personnes concernées de connaître les destinataires précis et les sources de leurs données est essentiel pour leur garantir un contrôle effectif sur leurs données. Ces informations leur permettent d'exercer leurs droits auprès des responsables de traitement qui détiennent leurs données. Cela est particulièrement important du fait de la chaîne d'acteurs souvent complexe dans le développement des systèmes d'IA.

Si la personne est présente et identifiable dans la base de données, le responsable de traitement doit pouvoir fournir l'ensemble des informations prévues par l'article 15 du RGPD. Parmi celles-ci, l'information sur les sources appellent toutefois des précisions propres à ce type de base.

**Lorsque les données n'ont pas été collectées directement auprès des personnes concernées** (par exemple, en cas de transmission par un courtier en données), **le droit d'accès permet d'obtenir toute information « disponible» quant à leur source.**

Toutefois, comme vu précédemment, le responsable du traitement n'aura pas toujours à conserver de telles informations lorsqu'il n'a pas ou plus besoin d'identifier les personnes concernées, et lorsque la traçabilité n'est pas nécessaire à la finalité ou pour assurer la proportionnalité du traitement des données d'entrainement.

Il devra alors fournir toute information utile et à sa disposition aux personnes concernées qui en font la demande, c'est-à-dire sur l'ensemble des sources utilisées.

Une analyse au cas par cas est nécessaire pour déterminer les informations qu'il est raisonnable et proportionné de conserver afin de garantir les droits des personnes sur leurs données. Si, notamment, en cas de collecte à partir de sources publiques, le RGPD ne rentre pas dans ce niveau de détails en ne prescrivant pas la conservation d'informations sur chaque URL utilisée, il est cependant nécessaire de constituer une documentation suffisante sur les sources des données d'entraînement afin de conserver leur consistance aux droits garantis par le règlement, notamment par l'article 15 du RGPD, et de permettre au responsable de traitement de documenter sa conformité au RGPD.

Par exemple :
En cas de réutilisation d'un jeu de données médicales pour créer une IA d'analyse d'imageries médicales, le responsable du traitement devra conserver les informations sur la source utilisée et lorsque cela est possible le moyen de contact du responsable du traitement source, afin de pouvoir la fournir aux personnes.

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 très difficile 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'entrainement est un moyen d'atteindre cet objectif, toutefois le RGPD n'est pas prescripteur et le responsable de traitement n'est tenu de fournir l'information disponible sur les sources dans le cadre du droit d'accès.

En tout état de cause, des informations sur les sources des données sont utiles pour démontrer la conformité de la base d'apprentissage. Les responsables de traitement devront donc conserver certaines informations sur les sources au titre du principe de responsabilité et de protection des données dès la conception. Dans certains cas, le règlement européen sur l'IA exige également du fournisseur du système ou du modèle d'IA qu'il documente les données d'entrainement, y compris leur provenance (articles 11 et 13 pour les systèmes à haut risque et article 53 pour les modèles d'IA à usage général).

## S’agissant des droits à la rectification, à l'opposition et à l’effacement de ses données contenues dans une base de données d’apprentissage

Les personnes concernées disposent du droit de rectifier les données qui les concernent. En l’occurrence, cela pourra concerner les annotations inexactes que les personnes concernées entendraient voir corriger.

Le droit à l’effacement permet d’exiger la suppression des données dans un certain nombre de cas prévus par l’article 17 du RGPD, par exemple lorsque la personne concernée révèle au responsable du traitement que ce dernier détient **des données sensibles** la concernant (au sens de l’article 9 du RGPD) et pour lesquelles aucune dérogation ne justifiait leur traitement.

Par ailleurs, **lorsque le traitement est fondé sur l’intérêt légitime ou l’exercice d’une mission d’intérêt public, les personnes concernées peuvent, à tout moment, demander à s’y opposer pour des raisons tenant à leur situation particulière**. Le responsable de traitement doit alors y faire droit, sauf motif impérieux et légitime nécessitant de continuer à traiter la donnée.

Par exemple : l’opposition, et l’effacement consécutif sont en principe de droit en cas de conservation de photographies compromettantes dans un jeu de données ou de propos nominatifs tenus sur le web alors que la personne était mineure.

En cas de moissonnage (*web scraping*) de données accessibles en ligne, la CNIL encourage le développement de solutions techniques qui permettraient de faciliter le respect de l’exercice du droit d’opposition en amont de la collecte des données. À titre de comparaison, on peut mentionner les dispositifs d’*opt out* mis en place en matière de propriété intellectuelle. Il existe aussi, 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. De cette manière, le responsable du traitement pourrait respecter l’opposition des personnes en s’abstenant de collecter les données de ces dernières.

## S’agissant de la notification des droits exercés (rectification, limitation et effacement)

L'article 19 du RGPD prévoit qu'un responsable du traitement notifie à chaque destinataire auquel les données personnelles ont été communiquées toute rectification, effacement de données à caractère personnel ou toute limitation du traitement effectué, à moins qu'une telle communication se révèle impossible ou exige des efforts disproportionnés (en fonction des technologies disponibles et des coûts de mise en œuvre). Par exemple, lorsque cela est possible, la CNIL recommande l’usage d’interfaces de programmation applicatives (API) (en particulier dans les cas les plus à risque), ou *a minima* des techniques de gestion des journalisation des téléchargements de données.

En cas de partage d’un jeu de données, une bonne pratique consiste à informer des différentes versions des jeux de données, et de prévoir une obligation contractuelle (par exemple dans la licence de réutilisation du jeu de données) de répercuter les effets de l’exercice des droits d’opposition, de rectification ou d’effacement par leurs réutilisateurs.

## Pour l'exercice des droits sur les modèles dont le traitement est soumis au RGPD

Certains modèles d'IA sont entraînés sur des bases de données contenant des données personnelles mais sont, en eux-mêmes, anonymes.

Cependant, dans d'autres cas, le modèle contient en lui-même des données personnelles qui peuvent en être extraites et ne peut, dès lors, être considéré comme anonyme (voir l'avis du CEPD 28/2024 sur certains aspects de la protection des données liés au traitement des données à caractère personnel dans le contexte des modèles d'IA). Il en va notamment ainsi, généralement, pour les grands modèles de langage qui, nourris de sources publiques, peuvent fournir des informations sur des personnes physiques, notamment des personnalités publiques.

Les droits ouverts aux personnes par le RGPD trouvent alors à s'appliquer au modèle mais leur exercice pose des difficultés particulières. Si certaines garanties devront être envisagées dans ce cas, la CNIL estime que les exigences du RGPD sont suffisamment équilibrées pour appréhender les spécificités des modèles d'IA. Il convient notamment de prendre en compte les contraintes des fournisseurs de modèles d'IA dans l'exercice des droits sur les données qu'ils contiennent tels que la difficulté (voire de l'impossibilité) d'identifier les données d'une personne en particulier au sein du modèle ou la complexité et les coûts de réentrainement d'un modèle susceptible d'être disproportionné.

## S'agissant de l'identification de la personne concernée au sein du modèle

**Lorsque le RGPD s'applique au modèle**, deux situations peuvent être distinguées :

- Soit la présence de données de la personne dans le modèle est évidente, (par exemple certains modèles d'IA sont spécifiquement conçus pour fournir en sortie des données personnelles utilisées pour leur entrainement) ;

- Soit ce n'est pas le cas, et se pose la question de l'appréciation de la vraisemblance que des données concernant cette personne soient dans le modèle. **Le responsable du traitement pourra alors démontrer qu'il n'est pas en mesure d'identifier les personnes au sein de son modèle au sens de l'article 11 du RGPD.**

Une fois la présence d'une personne identifiée au sein du modèle, se pose également la question d'identifier les données personnelles qui y sont contenues.

**L'influence des données d'une personne sur les paramètres du modèle**

Lors de l'entraînement des modèles d'apprentissage automatique les plus complexes comme les réseaux de neurones, les données d'entraînement sont utilisées pour optimiser les paramètres du modèle. Toutefois, en raison du nombre important d'itérations nécessaires à l'entraînement du modèle et de la nature des techniques utilisées (comme la descente du gradient), la contribution de chaque donnée au modèle est diffuse et les méthodes permettant d'en retrouver la trace font encore l'objet de recherche.

Le champ du désapprentissage machine, dont le principe, les avantages et les limitations sont décrits dans l'article Linc « **Comprendre le désapprentissage machine** » ainsi que dans le document du **programme Support Pool of Experts du CEPD *Effective implementation of data subjects' rights***, fait fréquemment l'objet d'avancées prometteuses. Par exemple, l'utilisation des fonctions d'influence, qui permettent de conserver une trace de la contribution des données sur les paramètres du

modèle lors de l'entraînement, est l'une des techniques qui pourrait permettre de résoudre la question de la contribution de chaque donnée. Ces techniques permettraient par exemple d'identifier les poids influencés par les données aberrantes contenues dans un jeu de données et ainsi de corriger leur valeur.

**Dans l'attente des avancées de la recherche sur ces sujets, il ne semble aujourd'hui pas possible, dans le cas général, pour un responsable de traitement d'identifier pour une personne en particulier, l'ensemble des données la concernant ayant été mémorisées par le modèle d'IA sans que cette dernière ne fournisse d'informations supplémentaires**. En revanche, il peut parfois lui être possible de déterminer si le modèle a appris des informations sur une personne en conduisant des tests et des attaques fictives, telles que les attaques par inférence d'appartenance ou par reconstruction d'attributs. Ces attaques sont décrites dans l'article Linc « Petite taxonomie des attaques des systèmes d'IA ».

**La CNIL alerte les responsables de traitement sur le fait que ce constat n'est valable qu'au regard de l'état de l'art actuel, et qu'elle encourage activement la recherche et le développement de pratiques permettant de respecter l'exercice des droits et le principe de la protection des données dès la conception.**

À noter qu'il existe toutefois des cas particuliers comme lorsque les paramètres du modèle contiennent explicitement certaines données d'apprentissage (ce qui peut être le cas de certains modèles tels que les machines à vecteurs de support ou SVM, de certains algorithmes de partitionnement des données, ou *clustering*, lorsqu'un système d'IA est connecté à une base de connaissance (comme la génération augmentée de récupération ou RAG) : il sera alors techniquement possible d'exercer les droits sur les paramètres du modèle.

Comme pour l'exercice des droits sur les données d'entrainement, **le responsable du traitement doit, si possible informer les personnes concernées qu'il n'est pas en mesure de les identifier par lui-même**, et leur permettre, si c'est envisageable, de fournir des informations complémentaires permettant de les identifier pour qu'ils puissent exercer leurs droits.

**La CNIL recommande au fournisseur du modèle d'indiquer aux personnes concernées les informations complémentaires à fournir pour les identifier.**

- **Lorsque le responsable du traitement dispose encore des données d'entrainement**, il peut être pertinent d'identifier la personne au sein de ces dernières, avant de vérifier si elles ont été mémorisées par le modèle et sont susceptibles d'en être extraites.

- **Lorsque le responsable du traitement ne dispose plus des données d'entrainement**, ce dernier peut s'appuyer sur leur typologie pour anticiper les catégories de données ayant été susceptibles d'être mémorisées, afin de faciliter les tentatives d'identification de la personne (par exemple par des tests et requêtes sur le modèle d'IA en question).

**Dans le cas de l'IA générative**, la CNIL recommande au concepteur d'établir une procédure interne consistant à interroger le modèle (par exemple à partir d'une liste de requêtes judicieusement choisies) pour vérifier les données qu'il aurait pu mémoriser sur la personne grâce aux informations fournies.

Exemple :
Le concepteur d'un grand modèle de langage (« *Large Language Model* » ou LLM en anglais) ayant entraîné son modèle sur des données collectées par moissonnage de divers sites sur le Web pourra indiquer à une personne souhaitant exercer ses droits

qu'il lui sera nécessaire de lui fournir les requêtes (prompts) qui révèleraient selon elle une mémorisation de ses données personnelles par le modèle d'IA.

Pour cela, la personne concernée peut, par exemple et si elle en dispose, se servir de la copie des données d'entrainement la concernant afin de vérifier que telle ou telle information peut être extraite du modèle (par exemple en demandant au modèle de compléter une partie de sa biographie accessible sur une encyclopédie en source ouverte (s'il s'agit d'une personnalité publique), ou un article issu d'un site internet).

En raison du fonctionnement probabiliste de ces modèles, il est possible que le résultat fourni ne résulte pas d'une mémorisation de données d'entrainement mais d'une génération fondée sur des corrélations apprises sur d'autres données. Ces résultats sont parfois qualifiés d'hallucinations lorsqu'ils sont factuellement infondés mais présentés comme véridiques.

Dans certains cas, il sera important de bien distinguer le modèle du système au sein duquel il est intégré. En effet, le système – qui peut être considéré comme une sorte d'interface entre l'utilisateur et le modèle d'IA qu'il intègre – peut ajouter différentes données à la requête du modèle ou à son résultat (par exemple à travers une recherche sur le web ou depuis une base de connaissance dans le cas du RAG). Il est donc possible qu'une personne croit à tort qu'un modèle d'IA contient des données la concernant, alors que les résultats en question sont occasionnés par d'autres éléments constitutifs du système. Il est donc essentiel de garder cette distinction à l'esprit pour identifier le bon responsable du traitement. Par exemple si une base de connaissance (dans le cas du RAG) est intégrée dans le système par son fournisseur, ce dernier devra aussi prendre en compte les demandes d'exercice de droit la concernant. A l'inverse, si cette base de connaissance a été ajoutée par un tiers, la personne concernée devra se tourner vers ce dernier. Une fiche pratique dédiée viendra prochainement clarifier les enjeux de responsabilité en la matière.

## S'agissant des informations sur le traitement du modèle (au titre du droit d'accès) et du droit d'obtenir une copie des données :

Lorsqu'il a pu identifier la personne concernée et vérifier qu'une mémorisation des données avait eu lieu, le responsable doit le lui confirmer.

Lorsque la personne a exercé son droit d'obtenir une copie de ses données, ce dernier devrait lui fournir le résultat de ses investigations la concernant, dans un format compréhensible, en particulier les exemples de sorties contenant ses données personnelles dans le cas des systèmes d'IA génératifs.

Lorsqu'il n'a pas pu vérifier la présence d'une mémorisation, mais que le responsable du traitement n'a pas pu exclure que celle-ci soit possible (notamment en raison des limitations techniques des méthodes actuelles), la CNIL recommande d'indiquer aux personnes qu'il n'est pas impossible que des données d'entraînement le concernant aient été mémorisées par le modèle.

Si le responsable du traitement dispose encore des données d'entraînement, la CNIL recommande qu'il réponde à la personne en lui confirmant que ses données ont été traitées pour l'apprentissage et qu'elles sont susceptibles d'avoir été mémorisées.

Le cas échéant des informations portant spécifiquement sur le modèle devront lui être communiquées, comme l'information sur les destinataires du modèle (15.1.c), sa durée de conservation ou les critères utilisés pour la déterminer (15.1.d), les droits qui peuvent être exercés sur le modèle (15.1.e et f), ainsi que sa provenance lorsqu'il n'a pas été conçu par le responsable de traitement (15.1.g).

## S’agissant de l’exercice des droits à la rectification, d’opposition, ou à l’effacement sur le modèle

Tous les droits des personnes sur leurs données ne sont pas absolus. C’est notamment le cas du droit à l’effacement, qui résulte essentiellement du droit d’opposition (dont les conditions sont développées plus spécifiquement ci-dessous) : il convient de mettre en balance le droit de la personne et l’éventuel « motif légitime et impérieux », selon l’expression du RGPD, du responsable de traitement.

Bien que l’identification de données à partir des paramètres d’un modèle présente des difficultés techniques importantes à ce jour, certaines solutions pratiques permettent de répondre à l’exercice des droits (notamment le réentraînement ou l’application de filtres) mais elles sont, à ce jour, soit particulièrement coûteuses et délicates, soit imparfaites.

Ainsi, le caractère proportionné ou non de la réponse à l’exercice d’un droit dépendra :

- de la sensibilité des données et des risques que leur régurgitation ou divulgation ferait peser sur les personnes.

  Par exemple une personne peut souhaiter s’opposer au traitement de ses données en raison d’hallucinations ou d’inexactitudes à son égard susceptibles de lui causer un préjudice (par exemple en attribuant à tort des comportements répréhensibles ou criminels ou en présentant la personne comme étant décédée).

- de l’atteinte à la liberté d’entreprendre du responsable de traitement, qui dépend de la faisabilité pratique et du coût des mesures à prendre pour accéder à la demande de la personne, notamment s’agissant d’un éventuel réentrainement du modèle (en termes de ressources computationnelles, environnementales, humaines et financières).

Par exemple : un grand modèle de langage est nourri d’un grand nombre de sources publiques et renseigne correctement les utilisateurs sur les fonctions de certaines personnalités publiques. Une personnalité publique demande à être effacée du modèle. Le réentraînement du modèles présente un coût très important. La demande d’effacement peut en principe être rejetée.

Par exemple : un modèle d’IA destiné à aider les compagnies d’assurances a été entraîné sur un ensemble de dossiers réels. Le modèle n’est pas anonyme car son anonymat aurait nuit à sa performance. Une personne, qui n’avait pas fait opposition à sa présence dans la base d’entraînement, se rend compte que des informations sur le règlement financier de son sinistre sortent dans les résultats du modèle lorsque l’on pose certaines questions. Le responsable de traitement doit alors en principe proposer des solutions pour répondre à la demande d’effacement des données du modèle.

Les techniques permettant de garantir les droits des personnes sur leurs données dans les modèles d’IA sont en pleine évolution : les responsables de traitement doivent rester attentif au fait que certaines demandes, qui pourraient être refusées aujourd’hui, devront être traitées demain si l’état de l’art,

notamment sur les techniques de réidentification et de désapprentissage,  progresse. Sans préjudice de méthodes efficaces qui pourront être développées par les acteurs, la CNIL identifie, à ce jeux, deux techniques principales :

- la plus efficace est le réentraînement, qui permet d'effacer ou de rectifier des données au cœur du modèle ;

- subsidiairement, lorsque le réentraînement n'est pas possible, des filtres permettent de remédier à certains effets du modèle sans supprimer les données du modèle lui-même.

Elles seront examinées successivement.

**Lorsque le responsable du traitement dispose toujours des données d'entraînement, un réentraînement du modèle permet de répondre à l'exercice des droits** suite à leur prise en compte sur le jeu de données.

Ce réentraînement peut avoir lieu de manière périodique afin de prendre en compte plusieurs demandes d'exercice de droit à la fois. En principe, le responsable du traitement doit répondre à une demande le plus rapidement possible, et au plus tard dans le délai d'un mois. Le RGPD prévoit toutefois que ce délai peut être prolongé de deux mois compte tenu de la complexité et du nombre de demandes (par exemple en fonction de l'ampleur du réentraînement à mener), à condition que la personne en soit informée. Lorsqu'un réentraînement n'est pas prévu ni possible dans ce délai, il peut être envisagé de recourir à des filtres (cf. ci-dessous) en attendant le prochain entraînement du modèle.

Il conviendra alors au responsable du traitement de fournir une version actualisée du modèle d'IA à ses utilisateurs, le cas échéant en leur imposant par voie contractuelle de n'utiliser qu'une version régulièrement mise à jour. Une bonne pratique consiste à communiquer plus largement 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. Cela implique également d'inciter les destinataires des versions antérieures à les supprimer ou à les remplacer par la dernière version.

**Subsidiairement, lorsque le réentrainement 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).

En pratique, les mesures les plus robustes en la matière ne sont pas appliquées au modèle lui-même mais au système d'IA qui met en œuvre le modèle, afin de limiter ses sorties. Le fournisseur du modèle doit donc s'assurer que ces mesures sont effectivement mises en place, par ses soins et/ou par les utilisateurs, en fonction du mode de déploiement. Une fiche dédiée à la répartition des responsabilités entre le fournisseur et le déployeur d'un modèle d'IA non anonyme sera prochainement publiée.

- Il est recommandé d'utiliser des règles générales prévenant la génération de données personnelles, plutôt qu'une « liste noire » des personnes ayant exercé leurs droits.

- Ces règles générales devraient chercher à détecter et pseudonymiser les données personnelles concernées dans les sorties lorsque la production de données personnelles n'est pas la finalité visée,

par exemple par des techniques de reconnaissance d'entités nommées.

- À défaut, une liste noire devra être envisagée. Il conviendrait d'en assurer la sécurité et d'évaluer l'impact du filtre sur les sorties, notamment en évaluant l'augmentation du risque d'attaque par inférence d'appartenance qu'elle peut induire en modifiant la distribution statistique des sorties, permettant ainsi d'identifier les personnes s'étant opposées.

Le fournisseur du modèle ou du système d'IA devrait alors fournir à son déployeur les moyens de mettre en œuvre ces mesures pour lui permettre de répondre à ses propres obligations.

**Bonne pratique**

Les solutions techniques aujourd'hui disponibles n'étant pas satisfaisantes dans tous les cas où un modèle est soumis au RGPD, **la CNIL invite en priorité les fournisseurs** à anonymiser les données d'entrainement ou, à défaut de pouvoir le faire, à s'assurer que le modèle d'IA est anonyme à l'issu de son entrainement.

---

## Dérogations à l'exercice des droits sur les bases de données ou sur le modèle d'IA

**Point de vigilance** : lorsqu'il peut se prévaloir de certaines dérogations, le responsable du traitement doit informer préalablement (c'est-à-dire au moment de la fourniture de l'information) les personnes que leurs droits font l'objet de restrictions et expliquer les motifs du refus de l'exercice d'un droit aux personnes qui en ont fait la demande.

En dehors des situations dans lesquelles **le responsable du traitement n'est pas en mesure d'identifier les personnes concernées** (voir les développements dédiés ci-dessus), le responsable du traitement peut déroger à l'exercice des droits dans les cas suivants :

- **La demande est manifestement infondée ou excessive** (article 12 du RGPD),
- **L'organisme qui reçoit la demande n'est pas le responsable du traitement en cause,**
- **L'exercice d'un ou plusieurs droits est exclu par le droit français ou européen** (au sens de l'article 23 du RGPD),
- **Pour les traitements à des fins de recherche scientifique ou historique, ou à des fins statistiques : lorsque l'exercice des droits risquerait de rendre impossible ou d'entraver sérieusement la réalisation de ces finalités spécifiques** (au sens de l'article 116 du décret d'application de la loi « informatique et libertés »).

En savoir plus : Respecter les droits des personnes

---

< Fiche précédente : Informer les personnes concernées

Sommaire

Fiche suivante : Annoter les données >