IA : Réaliser une analyse d'impact si nécessaire

08 avril 2024

La constitution d'une base de données pour l'apprentissage d'un système d'IA peut engendrer un risque élevé pour les droits et libertés des personnes. Dans ce cas, une analyse d'impact sur la protection des données est obligatoire. La CNIL vous explique comment et dans quels cas la réaliser.

L'analyse d'impact sur la protection des données (AIPD), est une démarche qui permet de cartographier et d'évaluer les risques d'un traitement sur la protection des données personnelles et d'établir un plan d'action pour les réduire à un niveau acceptable. Cette démarche, facilitée par les outils mis à disposition par la CNIL, est particulièrement utile pour maîtriser les risques liés à un traitement avant sa mise en œuvre, mais également pour assurer leur suivi dans le temps.

Une AIPD permet notamment de réaliser :

- un recensement et une évaluation des risques pour les personnes dont les données pourraient être collectées, au moyen d'une analyse de leur vraisemblance et gravité ;
- une analyse des mesures permettant aux personnes d'exercer leurs droits ;
- une évaluation de la maîtrise des personnes sur leurs données ;
- une évaluation de la transparence du traitement de données pour les personnes (consentement, information, etc.).

L'AIPD doit être réalisée avant la mise en œuvre du traitement et devra être modifiée de manière itérative au fur et à mesure de l'évolution des caractéristiques du traitement et de l'appréciation des risques.

La réalisation d'une AIPD pour le développement de systèmes d'IA

Identifier quand une AIPD est nécessaire

Le développement de systèmes d'IA nécessite, dans certains cas, la réalisation d'une AIPD. Une AIPD est obligatoire si le traitement envisagé est susceptible d'engendrer un risque élevé pour les droits et libertés des personnes physiques (article 35 du RGPD).

Dans ses lignes directrices concernant l'AIPD, Le Comité européen de la protection des données (CEPD) a identifié neuf critères permettant d'aider les responsables de traitement à déterminer si une AIPD est requise : tout traitement de données personnelles remplissant au moins deux critères de cette liste sera

présumé soumis à l'obligation de réaliser une AIPD. Certains de ces critères sont particulièrement pertinents pour la phase de développement :

- la collecte de données sensibles ou de données à caractère hautement personnel (catégories de données qui peuvent être considérées comme augmentant le risque d'atteinte aux droits et libertés des personnes, telles que des données de localisation ou des données financières, par exemple) ;
- la collecte de données personnelles à large échelle ;
- la collecte de données de personnes vulnérables, comme par exemple les personnes mineures ;
- le croisement ou la combinaison d'ensembles de données ;
- l'utilisation innovante ou l'application de nouvelles solutions technologiques ou organisationnelles.

Dans tous les cas, il convient de s'interroger sur l'existence de risques pour les personnes du fait de la constitution d'une base d'entraînement et de son utilisation : si des risques importants existent, notamment du fait d'un mésusage des données, d'une violation de données, ou lorsque le traitement peut donner lieu à une discrimination, une AIPD doit être réalisée même si deux de ces critères ne sont pas remplis ; à l'inverse, une AIPD n'a pas à être réalisée si deux critères sont remplis mais que le responsable de traitement peut établir de façon suffisamment certaine que le traitement des données personnelles en cause n'expose pas les individus à des risques élevés.

Sur la base de ces critères, la CNIL a publié une liste de traitements de données personnelles pour lesquels la réalisation d'une AIPD est obligatoire (pour plus d'information, voir le site de la CNIL). Parmi ceux-ci, plusieurs peuvent reposer sur des systèmes d'intelligence artificielle, tels que ceux impliquant un profilage ou une prise de décision automatisée : dans ce cas, une AIPD est toujours requise.

L'utilisation d'un système d'intelligence artificielle est-elle un « usage innovant » ?

L'usage innovant est l'un des 9 critères pouvant entraîner la réalisation d'une AIPD : il est apprécié au regard de l'état des connaissances technologiques et non uniquement du contexte du traitement (un traitement peut être très « innovant » pour un organisme donné, du fait de la nouveauté technologique qu'il y apporte, sans pour autant relever d'un usage innovant en général). L'utilisation de systèmes d'intelligence artificielle **ne relève pas systématiquement de l'usage innovant ou de l'application de nouvelles solutions technologiques ou organisationnelles.** Tout traitement utilisant un système d'IA ne remplira donc pas ce critère. Afin de déterminer si la technique utilisée relève de tels usages, il conviendra de distinguer deux catégories de systèmes :

- Les systèmes qui utilisent des techniques d'IA validées expérimentalement depuis plusieurs années et éprouvées en conditions réelles. Ces systèmes ne relèvent pas de l'usage innovant ou de l'application de nouvelles solutions technologiques ou organisationnelles.

Exemple : certaines techniques de régression ou de partitionnement de données ou clustering, ou encore certaines architectures de modèles comme les forêts aléatoires (ou « *random forests* »), dans les cas où les risques qui sont liés à leur utilisation sont connus ;

- Les systèmes qui utilisent des techniques encore nouvelles, telles que l'apprentissage profond et dont les risques commencent juste à être identifiés aujourd'hui, mais sont encore mal compris ou maitrisés. Ces systèmes relèvent de l'usage innovant.

Exemple : les systèmes d'IA génératives reposant sur un apprentissage portant sur de grandes quantités de données et dont le comportement ne peut être anticipé dans tous les cas d'usage.

À titre d'illustration, un projet de recherche qui vise à développer des outils de traitement automatique du langage pour des applications cliniques dans le domaine médical, à partir de larges volumes de données (données vocales, cas d'études cliniques, résultats médicaux, etc.), peut constituer un usage innovant, notamment compte tenu de l'incertitude qui subsiste quant aux résultats qui seront obtenus.

L'entraînement d'un système d'intelligence artificielle est-il un traitement « à grande échelle » ?

La collecte à grande échelle est l'un des 9 critères pouvant entraîner la réalisation d'une AIPD : si le développement d'un système d'IA repose souvent sur le traitement d'une grande quantité de données, cela ne relève pas nécessairement du traitement à grande échelle qui vise à « *traiter un volume considérable de données à caractère personnel au niveau régional, national ou supranational* [et qui peut] *affecter un nombre important de personnes concernées* » (considérant 91 du RGPD). Pour les systèmes d'IA, il conviendra notamment de déterminer si le développement concerne un très grand nombre de personnes.

Exemples :
Un organisme de recherche souhaite constituer une large base de données de photos de paysages (montagne, océan, désert, villes, etc.) pour permettre d'améliorer les performances de systèmes de vision par ordinateur. Certaines de ces images comportent des images d'individus, parfois reconnaissables.

Quand bien même la base de données compterait des millions d'images couvrant toute la surface de la planète, si le nombre d'images contenant des individus reconnaissables (et donc des données personnelles) est limité (par exemple à quelques milliers), le traitement ne sera pas qualifié de « traitement à grande échelle ». Il n'est cependant pas exclu qu'une AIPD soit exigée selon les autres critères à vérifier.

Lorsqu'un fournisseur d'un agent conversationnel constitue une base de données pour entraîner son modèle de langage (« *Large Language Model* » ou LLM en anglais) à partir d'un volume considérable de données personnelles publiquement accessibles sur Internet collectées par le biais de techniques de moissonnage (« *web scraping* »), le traitement peut être qualifié de « traitement à grande échelle ».

Les critères de risque introduits par le Règlement européen sur l'IA

La proposition de règlement européen sur l'intelligence artificielle, encore en cours d'élaboration, a pour vocation d'encadrer le développement et le déploiement des systèmes d'IA au sein de l'Union Européenne. Ce projet distingue plusieurs catégories de systèmes selon leur niveau de risque : les systèmes interdits, les systèmes à haut risque, les systèmes nécessitant des garanties de transparence et les systèmes à risque minimal. **La CNIL considère que pour le développement de l'ensemble des systèmes à haut risque visés par la proposition de Règlement IA, la réalisation d'une AIPD sera présumée nécessaire lorsque leur développement ou leur déploiement implique un traitement de données personnelles.**

La réalisation de l'AIPD pourra reposer sur la documentation exigée par la proposition de règlement sur l'IA sous réserve de comporter les éléments prévus par le RGPD (article 35 du RGPD). L'élaboration de règles plus précises sur l'articulation entre ces exigences fait l'objet de travaux européens auxquels la CNIL participe activement et qui feront l'objet de publications ultérieures. Ces travaux viseront notamment à éviter toute redondance dans les obligations pesant sur les acteurs en privilégiant la réutilisation d'un cadre à l'autre des éléments constitués.

De plus, **la CNIL considère que le développement d'un modèle de fondation ou d'un système d'IA à usage général, en ce que leurs usages ne peuvent être identifiés de manière exhaustive, nécessite dans la majorité des cas la réalisation d'une AIPD lorsqu'il implique le traitement de données personnelles.** En effet, bien que ces modèles et systèmes ne soient pas considérés comme à haut risque par défaut par la proposition de règlement IA, leur diffusion ainsi que leurs utilisations à venir pourraient comporter des risques pour les personnes dont les données ont été traitées lors du développement, ou pour les personnes concernées par leur utilisation.

La réalisation d'une AIPD pour les modèles de fondation et systèmes à usage général facilitera la mise en conformité des traitements mis en œuvre par leurs utilisateurs. A cet égard, le partage ou la publication des AIPD réalisées pourra faciliter la mise en conformité de tous les acteurs impliqués, notamment dans le cas de la diffusion des modèles en source ouverte, ou de la mise à disposition pour tous des systèmes.

Définir le périmètre de l'AIPD

Le périmètre de l'AIPD peut différer en fonction de la connaissance que le fournisseur a de l'usage qui sera fait, par lui-même ou par un tiers, du système d'IA qu'il développe.

Cas où l'usage opérationnel du système d'IA en phase de déploiement est identifié dès la phase de développement

Lorsque le fournisseur du système est également responsable du traitement pour la phase de déploiement et que l'usage opérationnel du système d'IA en phase de déploiement est identifié dès la phase de développement, il est recommandé de réaliser une AIPD générale pour l'ensemble du traitement. Le fournisseur pourra alors compléter cette AIPD par les risques liés aux deux phases.

Si le fournisseur n'est pas responsable du traitement pour la phase de déploiement mais qu'il identifie les finalités d'usage en phase de déploiement, il peut proposer au responsable du traitement un modèle d'AIPD. Cela peut lui permettre notamment de tenir compte de certains risques qu'il est plus facile d'identifier lors de la phase de développement. Toutefois, l'utilisateur du système d'IA, en tant que responsable de traitement, reste tenu de réaliser une AIPD, par exemple sur la base du modèle du fournisseur, s'il le souhaite.

Il est à noter que, dans certains cas, il n'est pas possible de déterminer, avec précision et de manière préalable, l'encadrement de la phase de déploiement (sur les données, etc.) : par exemple, certains risques peuvent être réévalués à l'issue d'une phase de calibrage du système d'IA dans ses conditions de déploiement. **L'AIPD devra alors être modifiée de manière itérative** au fur et à mesure de la définition des caractéristiques du traitement au stade du déploiement.

Cas où l'usage opérationnel du système d'IA en phase de déploiement n'est pas clairement identifié dès la phase de développement

Dans cette hypothèse, le fournisseur ne pourra réaliser son analyse d'impact que sur la phase de développement. Il appartiendra ensuite au responsable du traitement de la phase de déploiement d'analyser, au regard des caractéristiques du traitement, si une AIPD est nécessaire pour cette phase. Le cas échéant, si les finalités de la phase de déploiement sont multiples, le responsable de traitement pourra décliner une même AIPD générale pour chacun des cas d'usages spécifiques.

Les risques liés à l'IA à prendre en compte dans une AIPD

Les traitements de données personnelles reposant sur des systèmes d'intelligence artificielle présentent des risques spécifiques qu'il convient de prendre en compte :

- les risques pour les personnes concernées liés à des mésusages des données contenues dans la base d'apprentissage, notamment en cas de violation de données ;
- le risque d'une discrimination automatisée causée par un biais du système d'IA introduit lors du développement, par exemple lié à une performance moindre du système pour certaines catégories de personnes ;
- le risque de produire du contenu fictif erroné sur une personne réelle, particulièrement important dans le cas des systèmes d'IA génératives, et pouvant avoir des conséquences sur sa réputation ;
- le risque de prise de décision automatisée causée par un biais d'automatisation ou de confirmation dans le cas où les mesures d'explicabilité nécessaires ne sont pas prises lors du développement de la solution (comme la remontée d'un score de confiance, ou d'informations intermédiaires tel qu'une carte de saillance ou « *saliency map* ») ou si un agent utilisant le système d'IA ne peut pas prendre une décision contraire sans que cela ne lui porte préjudice ;
- les risques liés aux attaques connues spécifiques aux systèmes d'IA tel que les attaques par empoisonnement des données, par insertion d'une porte dérobée, ou encore par inversion du modèle ;
- les risques liés à la confidentialité des données susceptibles d'être extraites depuis le système d'IA ;
- les risques éthiques systémiques et graves liés au déploiement du système, tels que les impacts sur le fonctionnement démocratique de la société, ou encore sur le respect des droits fondamentaux (par exemple en cas de discrimination), et pouvant être pris en compte lors de la phase de développement.
- Enfin, le risque d'une perte de contrôle des utilisateurs sur leurs données accessibles en ligne, une collecte à large échelle étant souvent nécessaire à l'apprentissage d'un système d'IA, notamment lorsque celles-ci sont collectées par moissonnage ou *web scraping*.

Lorsque plusieurs sources de données sont utilisées pour le développement du système d'IA, les risques cités sont à prendre en considération pour chacune des sources, mais également pour l'ensemble ainsi constitué. De plus, lorsque le système est développé sur la base d'un modèle pré-entraîné fourni par un tiers, le modèle doit tout de même être soumis à l'analyse de risque décrite ci-dessus, par exemple sur la base des informations fournies par l'organisme fournissant le modèle.

Enfin, des analyses provenant de référentiels publiés par la CNIL ou par des tiers pourront être intégrées ou associées à l'AIPD. Parmi ces référentiels, la CNIL recommande d'utiliser :

- le guide d'auto-évaluation publié par la CNIL ;
- les référentiels et cadres recensés par la CNIL sur la page « Autres guides, outils et bonnes pratiques » ;
- la proposition de règlement européen sur l'intelligence artificielle, et notamment son annexe IV détaillant la documentation technique qui doit accompagner la mise sur le marché des systèmes d'IA à haut risque.

Articulation entre les exigences de documentation de la proposition de règlement IA et la réalisation d'une AIPD

Si elles s'inscrivent toutes deux dans une logique d'anticipation des risques et peuvent se recouper, il existe des différences notables entre l'AIPD et la documentation de la conformité à la proposition de règlement sur l'IA.

D'une part, elles diffèrent dans leur champ d'application. Dès lors que certains systèmes d'IA n'étant pas classifiés comme à haut risque reposeront sur des traitements présentant des risques pour la protection des données personnelles, ceux-ci nécessiteront la réalisation d'une AIPD.

D'autre part, il appartiendra au responsable du traitement en cause, que ce dernier concerne le développement ou le déploiement du système, de réaliser une AIPD, alors que les exigences de documentation du projet de règlement sur l'IA pèseront essentiellement sur le fournisseur du système d'IA.

Toutefois, il est prévu que dans les cas où un fournisseur de système d'IA soumis aux obligations de documentation du règlement IA doit également réaliser une AIPD, il soit encouragé à reprendre des éléments issus du premier document dans le second. L'élaboration de règles plus précises sur l'articulation entre ces exigences fait l'objet de travaux européens auxquels la CNIL participe activement et qui feront l'objet de publications ultérieures. La possibilité de ne travailler dans ses cas que sur un unique document intégrant les exigences de l'AIPD et de la documentation du règlement IA sera ainsi explorée.

En savoir plus : Les guides AIPD de la CNIL

Les mesures à prendre en fonction des résultats de l'AIPD

L'AIPD est un exercice qui permet d'abord de déterminer le niveau de risque lié à un traitement de données à caractère personnel. Une fois ce niveau déterminé, il convient de concevoir dans l'AIPD un ensemble de mesures visant à le réduire et à le maintenir à un niveau acceptable. Ces mesures doivent intégrer les recommandations de la CNIL venant à s'appliquer, qu'elles portent sur les techniques d'IA utilisées ou non.

En savoir plus : AIPD – Les bases de connaissances

Par ailleurs, certaines mesures spécifiques au domaine de l'IA – en particulier d'ordre technique – pourront être mise en œuvre, parmi lesquelles :

- des mesures de sécurité, telles que le chiffrement homomorphe ou l'utilisation d'un environnement d'exécution sécurisé ;
- des mesures de minimisation, telles que le recours à des données synthétiques ;
- des mesures d'anonymisation ou de pseudonymisation, telles que la confidentialité différentielle ;
- des mesures de protection des données dès le développement, telles que l'apprentissage fédéré ;
- des mesures facilitant l'exercice des droits ou les recours pour les personnes, telles que les techniques de désapprentissage machine, ou les mesures d'explicabilité et de traçabilité des sorties du systèmes d'IA ;
- des mesures d'audit et de validation, reposant par exemple sur des attaques fictives de type « *red teaming* », notamment pour identifier et corriger les biais ou les erreurs en défaveur de certaines personnes ou catégories de personnes.

D'autres mesures, plus génériques, pourront également être appliquées :

- des mesures organisationnelles, telles que l'encadrement et la limitation de l'accès aux bases de données d'apprentissage et pouvant permettre une modification du système d'IA, la limitation de l'accès aux données par les tiers et les sous-traitants ;
- des mesures de gouvernance, telles que la mise en place d'un comité éthique ;
- des mesures de traçabilité des actions effectuées afin d'identifier et d'expliquer les comportements anormaux ;
- des mesures prévoyant une documentation interne, comme la rédaction d'une charte informatique.

Ces mesures devront être sélectionnées au cas par cas afin de réduire les risques spécifiques au traitement de données considéré. Elles devront être intégrées dans un plan d'action et faire l'objet d'un suivi. De plus, étant destinées à protéger les données lors du développement du système d'IA et notamment lors de la constitution de la base données, elles pourront être complétées d'autres mesures spécifiques à l'IA, à appliquer lors de la phase de déploiement. En particulier, une description des mesures spécifiques au déploiement d'une IA générative sera fournie dans une fiche ultérieure.

Enfin, la publication de tout ou partie de l'AIPD est recommandée, dans un objectif de transparence : si certaines parties de l'AIPD n'ont pas à être publiées dans la mesure où elles peuvent être couvertes par le secret des affaires ou donner des informations confidentielles sur la sécurité du système, d'autres présentent les risques et les mesures prises pour les limiter et leur publication présente un intérêt pour les utilisateurs du système et le public.

< Fiche précédente : Assurer que le traitement est licite - En cas de réutilisation des données

Sommaire

Fiche suivante : Tenir compte de la protection des données dans la conception du système >