CNIL -Les fiches pratiques IA

Ed01

1 -Introduction : QUEL EST LE PÉRIMÈTRE DES FICHES PRATIQUES SUR L'IA ?

La CNIL apporte des réponses concrètes pour la constitution de bases de données utilisées pour l'apprentissage des systèmes d'intelligence artificielle (IA), qui impliquent des données personnelles.

1 – 1 - Quel est le périmètre des fiches pratiques sur l'IA ?

La CNIL souhaite apporter des clarifications et des recommandations concrètes pour le développement des systèmes d'intelligence artificielle (IA) et la constitution de bases de données utilisées pour leur apprentissage, qui impliquent des données personnelles. La CNIL définit le périmètre des différentes fiches pratiques.

Les fiches suivantes concernent uniquement la phase de développement de systèmes d'IA, et non celle de déploiement, lorsque celle-ci implique le traitement de données à caractère personnel (« données personnelles »). Elles se limitent aux traitements de données soumis au règlement général sur la protection des données (RGPD).

Ces fiches doivent permettre d'accompagner les professionnels aux profils aussi bien juridique que technique (délégués à la protection des données, professionnels du droit, personnes disposant de compétences techniques spécifiques ou non à l'IA, etc.).

1 – 2 - La présence de données personnelles

Les fiches pratiques se rapportent aux activités de constitution de base de données et de leur utilisation dans le cadre du développement de systèmes d'IA lorsque tout ou partie de ces données sont des données personnelles. Dans la pratique, trois cas peuvent être rencontrés :

Il est certain qu'aucune donnée personnelle n'est présente dans la base de données : les fiches ne s'appliquent pas (même si certaines recommandations, de l'ordre de la bonne pratique, peuvent être pertinentes).

Il est certain que des données personnelles sont présentes : les fiches s'appliquent. C'est le cas des systèmes d'IA développés à partir de vidéos ou d'images de personnes, d'enregistrements de voix, de données personnelles structurées, etc. - A noter que les textes européens posent la règle selon laquelle les jeux de données comprenant des données personnelles et non personnelles, dits « mixtes », sont régis par le RGPD, si les deux types de données sont inextricablement liés.

Il est possible que des données personnelles soient présentes : c'est un cas fréquent pour lequel la collecte de données personnelles n'est pas expressément souhaitée. Par exemple :
présence résiduelle de personnes ou de plaques d'immatriculation dans des images ;
occurrences de noms, prénoms, adresses, etc. dans des données textuelles de types commentaires ou prompt, etc.
Dans ces cas, sous réserve d'avoir anonymisé les données personnelles originales et pour les opérations de traitement ultérieures à cette suppression, les données perdent leur caractère personnel et les fiches ne s'appliquent plus.
par vérification manuelle, par exemple à l'occasion de l'annotation des données ;
par vérification automatique, par exemple par l'utilisation de techniques de détection de personnes/visages dans les images, par des méthodes de reconnaissance d'entités nommées (named-entity recognition), etc.
Les questions relatives aux risques liés à l'utilisation du système feront l'objet de fiches publiées ultérieurement

1 – 3 - Les systèmes d'IA concernés

Les fiches pratiques de la CNIL concernent le développement de systèmes mettant en œuvre des techniques d'intelligence artificielle impliquant un traitement de données personnelles. Ceux-ci sont qualifiés de **« systèmes d'IA ».**

La définition des systèmes d'IA concernés par ces fiches pratiques est alignée avec celle de la proposition de règlement européen sur l'IA en cours d'adoption (voir encadré plus bas). Le Parlement européen considère ainsi que la définition de systèmes d'IA doit se fonder sur « des caractéristiques essentielles du domaine de l'intelligence artificielle, telles que comme ses capacités d'apprentissage, de raisonnement ou de modélisation, afin de la distinguer de systèmes logiciels ou d'approches de programmation plus simples. Les systèmes d'IA sont conçus pour fonctionner avec différents niveaux d'autonomie, ce qui signifie qu'ils ont au moins un certain degré d'indépendance d'action par rapport aux commandes humaines et des capacités à fonctionner sans intervention humaine. »

A noter : cette définition sera mise à jour avec celle qui figurera de façon définitive dans le règlement IA.

En pratique, les systèmes d'IA concernés incluent les systèmes fondés sur l'apprentissage automatique (supervisé, non supervisé, par renforcement) et ceux fondés sur la logique et les connaissances (bases de connaissance, moteurs d'inférence et de déduction, raisonnement symbolique, systèmes experts, etc.), ainsi que les approches hybrides.

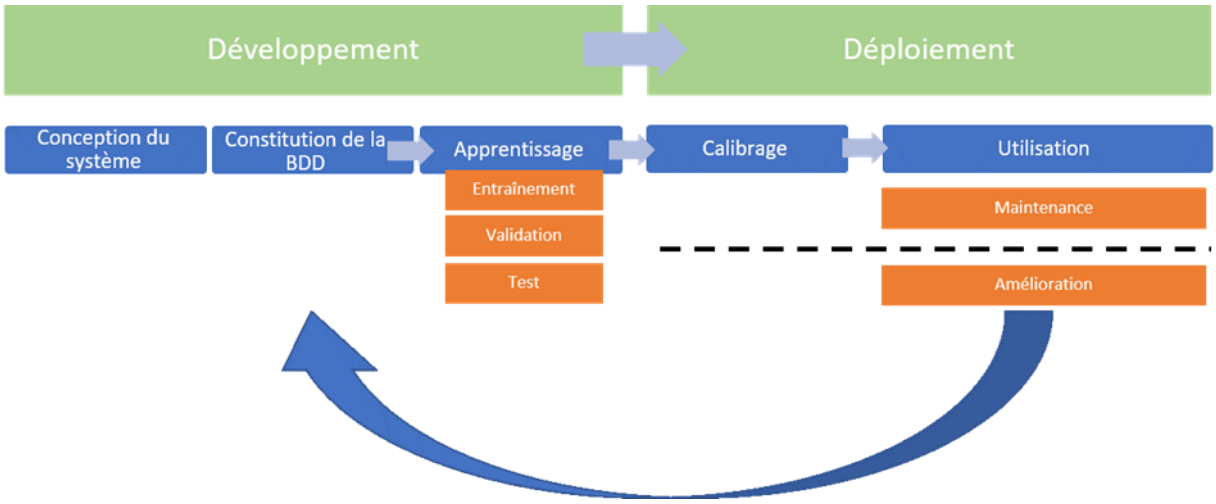
Les fiches pratiques portent sur ces systèmes que l'usage opérationnel en phase de déploiement soit défini dès la phase de développement, ou qu'il s'agisse de systèmes d'IA à usage général(« general purpose AI »), par exemple mettant en œuvre des

modèles « de fondation », du fait de leur capacité à être réutilisés et adaptés pour différentes applications et cas d'usage.

Elles concernent également tous les systèmes d'IA tels que définis ci-dessus, que l'apprentissage soit par exemple réalisé « une fois pour toutes » ou en continu. Dans le cas des systèmes d'apprentissage continu, les données collectées lorsque le système est déployé sont réutilisées pour l'amélioration itérative du système.

Enfin, elles concernent les traitements consistant à entraîner ou à ajuster (fine tuning/transfer learning) des modèles d'IA existants, indépendamment de leur intégration dans un système à proprement parler, dès lors qu'ils impliquent des données personnelles.

1 – 4 - Les phases du traitement concernées



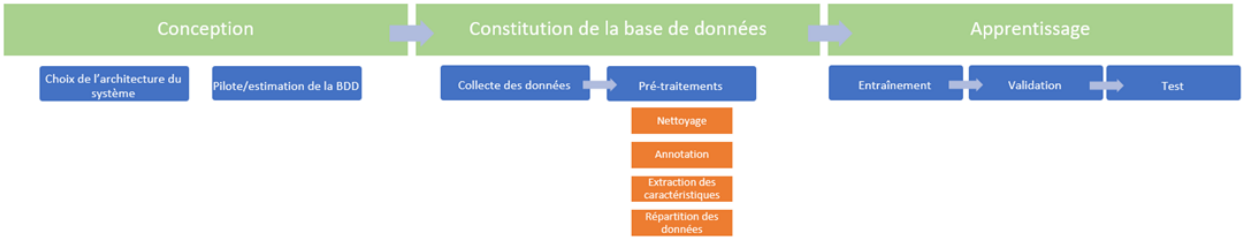
Comme illustré dans le schéma précédent, la mise en place d'un système d'IA reposant sur l'apprentissage automatique nécessite, en principe, la succession de deux phases distinctes :

- **la phase de développement** : elle consiste à concevoir, développer et entraîner un système d'IA.
- **la phase de déploiement** : elle consiste à mettre en usage le système d'IA développé lors de la première phase.

La phase de développement comprend toutes les étapes du développement du système d'IA jusqu'à son déploiement (phase de production), à savoir :

- **la conception du système** : choix de l'architecture, dont la ou les méthodes d'apprentissage et le cas échéant la sélection de modèles pré-entraînés, identification des données nécessaires et premiers tests, pilotes ;
- **la constitution de la base de données** : collecte et prétraitement (nettoyage, annotation, extraction de caractéristiques, répartition des données) ;

- **l'apprentissage** : entraînement du modèle, éventuel ajustement ou fine-tuning), réglage et validation des hyperparamètres, tests de performance ;
- **l'intégration** : lorsque le produit final attendu à l'issue du développement est un système et non un modèle, insertion du modèle entraîné dans le système d'information, connexion aux autres composantes logicielles, développement d'une interface utilisateur, rédaction d'une documentation utilisateur, etc



Dans de nombreux cas, le développement d'un système d'IA reposera sur l'ajustement de **modèles pré-entraînés** (« **fine-tuning** ») ou **l'apprentissage par transfert** (« **transfer learning** »). La CNIL considère que cette phase constitue une deuxième phase de développement, distincte de celle ayant permis la constitution du modèle d'origine.

Dans le cas de l'apprentissage continu, les données sont collectées lors du déploiement et de la mise en production du modèle, pour son amélioration future. L'entraînement en continu est donc inclus dans cette phase de développement via une boucle de rétroaction.

Il convient de noter que s'ajoutent à ces deux phases, une phase d'arrêt du système d'IA ou de suppression des données personnelles qu'il contenait. Les présentes fiches ne précisent pas les opérations à réaliser lors de cette phase, qui relèvent également de la réglementation sur la protection des donnés personnelles.

1 – 5 - Textes applicables :

1 – 5 – 1 -Réglementation relative à la protection des données personnelles

Les fiches pratiques concernent, en particulier, les cas d'usages relatifs à la phase de développement d'un système d'IA (recherche scientifique, recherche et développement, personnalisation d'un produit commercial, amélioration du service public rendu à l'usager, etc.) pour lesquels le RGPD est applicable.

Rappel sur le champ d'application du RGPD

Le RGPD s'applique à toute organisation :
• publique et privée, quel que soit sa taille (entreprise, administration, collectivité, association, etc.) ;
• qui traite des données personnelles pour son compte ou non ;

• établie sur le territoire de l'Union européenne dès lors que le traitement est effectué dans le cadre des activités d'un de ses établissements sur le territoire de l'Union, que le traitement ait lieu ou non dans l'Union ;
• ou qui, non établie sur le territoire de l'Union européenne, cible directement des personnes physiques dans l'Union européenne ou opère un suivi de leur comportement.

Exemples :

- Application du RGPD à la réutilisation de bases de données constituées hors de l'UE : le RGPD est applicable à la réutilisation de bases de données par un responsable de traitement ou un sous-traitant

  Sous\-traitant

  Le sous-traitant est la personne physique ou morale (entreprise ou organisme public) qui traite des données pour le compte d'un autre organisme (« le responsable de traitement »), dans le cadre d'un service ou d'une prestation. Les sous-...> En savoir plus

  établi dans l'Union européenne dès lors que le traitement est effectué dans le cadre des activités d'un de ses établissements sur le territoire de l'Union, même si ces bases de données ont été constituées hors de l'UE et qu'elles contiennent les données personnelles de personnes ne se trouvant pas sur le territoire de l'UE. Dans ce cas, le responsable du traitement est donc tenu de respecter la réglementation applicable en matière de protection des données.

- Application du RGPD à la réutilisation de modèles entraînés hors de l'UE : le RGPD s'applique à l'utilisation des modèles entraînés hors de l'Union européenne par un responsable de traitement ou un sous-traitant établi au sein de l'Union européenne dès lors qu'ils contiennent des données personnelles (voir la fiche outil à venir sur le statut des modèles) et que le traitement est effectué dans le cadre des activités d'un de ses établissements sur le territoire de l'Union.

Les traitements de données en phase de développement du système d'IA soumis au champ de la directive « police-justice » ainsi que ceux qui intéressent la sûreté de l'État et la défense nationale sont donc exclus du périmètre de ces fiches. Toutefois, les recommandations des présentes fiches peuvent servir d'inspiration pour ces traitements

1 – 5 – 2 - Autres réglementations applicables

Si ces fiches visent à clarifier comment le développement de systèmes d'IA peut se conformer aux obligations en matière de protection de données personnelles, d'autres réglementations, que ces fiches n'abordent pas directement, sont susceptibles de s'appliquer. C'est par exemple le cas de la réglementation relative au droit de la propriété intellectuelle ou encore le règlement sur la gouvernance des données qui encadre notamment les services d'intermédiation de la donnée ou l'atruisme des données.

D'autres sont encore en cours d'élaboration. C'est en particulier le cas de la proposition de règlement européen sur l'IA qui a vocation à encadrer le développement et le déploiement de systèmes d'IA au sein de l'Union Européenne.

Enfin, des réglementations sectorielles s'appliquent aux systèmes d'IA développés ou déployés pour certaines applications soumises à une réglementation spécifique (santé, finance, systèmes critiques, etc.). Il appartient à chaque responsable de traitement de déterminer les réglementations applicables et de se tourner vers les régulateurs compétents

1 – 5 – 3 - Articulation des fiches avec la proposition de règlement sur l'IA

La proposition de **règlement européen** distingue notamment plusieurs catégories de systèmes selon leur niveau de risque au regard de la sécurité des produits et des droits fondamentaux : les systèmes interdits, les systèmes à haut risque, les systèmes nécessitant des garanties de transparence et les systèmes à risque minimal. Il prévoit ainsi différents degrés d'obligations reposant principalement sur les fournisseurs de systèmes d'IA.

S'il faut attendre son adoption définitive, les présentes fiches ont été élaborées en vue d'une articulation intelligible avec ces futures obligations (par exemple en matière de qualification des acteurs et d'évaluation des risques).

Il est toutefois à noter que ces fiches s'appliquent, à droit constant, à tout traitement de données soumis au RGPD dans le cadre du développement d'un modèle ou d'un système d'IA, indépendamment de l'entrée en application des règles européennes sur l'intelligence artificielle. La CNIL rappelle, par ailleurs, que la proposition de règlement sur l'IA n'a pas vocation à remplacer les obligations en matière de protection des données mais bien à les compléter.

L'élaboration de règles plus précises sur l'articulation entre ces différentes exigences fait l'objet de travaux européens auxquels la CNIL participe activement et qui donneront lieu à des publications ultérieures

2 - DÉTERMINER LE RÉGIME JURIDIQUE APPLICABLE

La CNIL vous aide à déterminer le régime juridique applicable aux traitements de données personnelles en phase de développement.
Lorsqu'elle contient des données personnelles, la constitution de bases de données pour l'apprentissage d'un système d'IA puis la phase d'apprentissage elle-même doivent respecter la réglementation relative à la protection des données. La CNIL vous aide à déterminer le régime applicable aux traitements de données en phase de développement

2 – 1 – PRINCIPE

Les phases de développement et de déploiement d'un système d'IA constituent des traitements de données personnelles distincts, soumis à la réglementation en matière de protection des données personnelles. Il existe des régimes juridiques différents selon les traitements :

- **le régime résultant du règlement général sur la protection des données (RGPD)** qui a vocation à s'appliquer à l'ensemble des traitements de données

personnelles, à la fois dans le secteur public et le secteur privé, à l'exception des traitements relevant des deux régimes spécifiques suivants ;

- **le régime spécifique aux secteurs « police-justice** » (titre III de la loi « informatique et libertés ») ;

- **le régime intéressant la défense nationale ou la sûreté de l'Etat** régi par les dispositions de la loi « informatique et libertés ».

Cette fiche a pour objectif de définir les cas où le traitement en phase de développement est soumis au même régime juridique que le traitement en phase de déploiement, et les cas où ils sont soumis à des régimes distincts.

Pour rappel : les principes et recommandations formulés dans les fiches suivantes ne concernent que les traitements qui relèvent du RGPD.

En pratique

Pour déterminer le régime applicable aux traitements de données en phase de développement, il faut distinguer deux cas.

- **Cas n°1 : l'usage opérationnel du système d'IA en phase de déploiement est défini dès la phase de développement**

  Dans le cas où l'usage opérationnel du système d'IA en phase de déploiement est identifié dès le développement et si les traitements mis en œuvre en phase de développement poursuivent exclusivement la même finalité que ceux en phase de déploiement, il est possible de considérer qu'ils relèvent, généralement, du même régime juridique (v. Conseil d'État, 22 juillet 2022, n° 451653).

  Ce sera notamment le cas lorsque le choix du développement d'un système d'IA spécifique fait partie des moyens identifiés pour atteindre la finalité fixée pour le système à déployer.

  A noter, le régime « police-justice » (titre III de la loi « informatique et libertés ») est susceptible de s'appliquer aux traitements en phase de développement si les conditions suivantes sont remplies :

  - l'usage opérationnel du système d'IA est identifié de manière unique dès la phase de développement, de sorte que les traitements mis en œuvre en phase de développement poursuivent exclusivement la même finalité que ceux en phase de déploiement ;

  - l'utilisation du système d'IA développé poursuit exclusivement des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales, y compris la protection contre les menaces pour la sécurité publique et la prévention de telles menaces ;

- le responsable du traitement en phase de développement est une « autorité compétente ».

- **Cas n°2 : l'usage opérationnel du système d'IA en phase de déploiement n'est pas clairement défini dès la phase de développement (système d'IA à usage général)**

  La phase de développement et la phase de déploiement du système d'IA peuvent être décorrélées.

  Il n'est pas toujours possible d'identifier clairement la finalité du traitement en phase de déploiement dès la phase de développement. Certains systèmes d'IA (des systèmes d'IA dits « à usage général » ou « general purpose AI ») sont, en effet, développés sans qu'un usage opérationnel précis ne soit défini.

  Le régime juridique de la phase de développement n'est donc pas systématiquement le même que celui déterminé en phase de déploiement.

  On considère en général dans cette hypothèse, sous réserve d'une analyse au cas par cas, que les traitements en phase de développement sont soumis au RGPD.

  Exemple 1 : un organisme souhaite développer un modèle de reconnaissance vocale capable d'identifier un locuteur et sa langue afin de le commercialiser pour différents usages opérationnels en phase de production (p. ex. : des outils d'identification des personnes par des assistants vocaux ou des applications de traduction vocale sur un terminal mobile, etc.).

  Dans ce cas, la constitution de la base de données pour l'apprentissage du modèle relève du RGPD.

  Cela n'exclut toutefois pas, selon l'usage opérationnel du système d'IA, que les traitements de données en phase de déploiement soient soumis au régime « police-justice », s'ils sont mis en œuvre par une autorité compétente à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales.

  Exemple 2 : une entreprise développe un système de classification d'images permettant de détecter le franchissement d'une zone. Il le commercialise ensuite à plusieurs entités :

  - dans le premier cas, il le commercialise à une entreprise qui l'utilise à des fins statistiques pour mesurer l'affluence de personnes entrant dans un centre commercial. Dans ce cas, les traitements en phase de développement et de déploiement seront soumis au RGPD.

  - dans le deuxième cas, il le commercialise à un service de police nationale qui l'utilise afin de détecter le franchissement par les personnes de zones interdites à des fins de poursuites judiciaires. Dans ce cas, les traitements

en phase de développement seront soumis au RGPD, mais les traitements en phase de déploiement seront soumis au régime « police-justice ».

2 – 2- DÉFINIR UNE FINALITÉ

La constitution d'une base de données contenant des données personnelles pour le développement d'un système d'IA est un traitement de données personnelles qui, en application du RGPD, doit poursuivre une finalité (ou objectif) qui soit déterminée, explicite et légitime. La CNIL vous aide à définir la ou les finalités en tenant compte des spécificités du développement de systèmes d'IA.

Le principe

La finalité du traitement est l'objectif poursuivi par l'utilisation des données personnelles. Cet objectif doit être déterminé, c'est-à-dire établi dès la définition du projet. Il doit également être explicite, c'est-à-dire connu et compréhensible. Il doit enfin être légitime, c'est-à-dire compatible avec les missions de l'organisme.

Les données ne doivent pas être traitées ultérieurement de façon incompatible avec cet objectif initial : le principe de finalité limite la manière dont le responsable du traitement peut utiliser ou réutiliser ces données dans le futur.

L'exigence d'une finalité déterminée, explicite et légitime est particulièrement importante, car elle conditionne l'application d'autres principes du RGPD, dont notamment :

- **le principe de transparence** : l'objectif du traitement doit être porté à la connaissance des personnes concernées afin qu'elles soient en mesure de connaître la raison de la collecte des données les concernant et de comprendre l'utilisation qui en sera faite ;

- **le principe de minimisation** : les données sélectionnées doivent être adéquates, pertinentes et limitées à ce qui est nécessaire au regard des objectifs pour lesquels elles sont traitées ;

- **le principe de limitation des durées de conservation** : les données ne peuvent être conservées que pour une durée limitée, définie selon l'objectif pour lesquelles elles ont été collectées.

En savoir plus : Définir une finalité

Comment définir la finalité du traitement lorsque l'usage opérationnel est identifié dès la phase de développement ?

Ce cas de figure concerne les systèmes d'IA qui sont développés pour servir à un usage opérationnel précis en phase de déploiement. Cela exclut les systèmes d'IA qui sont développés sans qu'un usage opérationnel ne soit défini dès la phase de développement (voir section suivante).

Lorsqu'un système d'IA est développé pour un seul usage opérationnel, on considère que la finalité en phase de développement est directement liée à celle poursuivie par le traitement en phase de déploiement. Il en résulte que si la finalité en phase de déploiement est elle-même déterminée, explicite et légitime, la finalité en phase de développement le sera également.

Dans ce cas, ces deux phases consistent tout de même en des traitements distincts dont la conformité aux obligations du RGPD doit être analysée séparément (en termes notamment d'identification de la base légale, d'information des personnes, de minimisation des données collectées, de définition des durées de conservation, etc.).

Exemple : un organisme souhaite constituer une base de données composée de photos de rames de trains en service – c'est-à-dire avec des personnes présentes – afin d'entraîner un algorithme en vue de mesurer l'affluence et la fréquentation des trains à quai dans les gares. La finalité en phase de développement peut être considérée comme déterminée, explicite et légitime au regard de l'usage opérationnel identifié.

Dans certains cas, un système d'IA peut être développé pour plusieurs usages opérationnels définis dès la phase de développement. Dans ce cas, le développement d'un tel système d'IA peut poursuivre plusieurs finalités correspondant aux usages opérationnels identifiés (un traitement de données peut en effet poursuivre simultanément plusieurs finalités si elles sont toutes déterminées, explicites et légitimes).

Comment définir la finalité du traitement pour le développement de systèmes d'IA à usage général ?

Ce cas concerne les systèmes d'IA dont l'usage opérationnel en phase de déploiement n'est pas clairement identifié dès la phase de développement. Sont visés, ici, les systèmes d'IA à usage général et les modèles de fondation utilisables pour une grande variété d'applications pour lesquels il peut être difficile de définir une finalité suffisamment déterminée et explicite au stade du développement.

Exemples :

- Un organisme peut constituer une base de données pour l'entraînement d'un modèle de classification d'images (personnes, véhicules, aliments, etc.) et le rendre publiquement accessible, sans qu'aucun usage opérationnel spécifique ne soit prévu lors du développement du modèle.

- Ce modèle peut être librement réutilisé conformément à la licence associée (en étant éventuellement adapté, par exemple à l'aide de techniques d'apprentissage par transfert ou transfer learning), et à la réglementation relative au droit à l'image et à la propriété intellectuelle, par des organismes tiers pour le développement de systèmes de vision par ordinateur. Les finalités du système d'IA peuvent être variées : détection de personnes par des systèmes de caméras

augmentées pour la mesure de l'affluence sur des quais de gare ou encore détection de défauts sur des images prises dans le cadre de contrôles de la qualité de produits.

- Un organisme constitue une base de données pour l'entraînement d'un modèle de langage permettant d'identifier le registre de langue d'un texte. Ce modèle peut être utilisé pour diverses tâches : la rédaction et la relecture d'articles, de courriers, de discours, l'apprentissage du français, etc.

La finalité du traitement en phase de développement peut être considérée comme déterminée, explicite et légitime si elle est suffisamment précise, c'est-à-dire lorsqu'elle se réfère cumulativement :

- au « type » de système développé, comme, par exemple, le développement d'un modèle de langage de grande taille, d'un système de « vision par ordinateur » ou encore d'un système d'IA générative d'images, de vidéos ou de sons. Les types de systèmes doivent être présentés de manière suffisamment claire et intelligible pour les personnes concernées, compte tenu de leurs complexités techniques et des évolutions rapides dans ce domaine.

- aux fonctionnalités et capacités techniquement envisageables, ce qui implique pour le responsable du traitement de dresser une liste des capacités qu'il peut raisonnablement prévoir dès la phase de développement.

Ces critères permettent de prendre en compte le fait que le responsable du traitement ne puisse pas définir au stade du développement d'un système d'IA l'ensemble de ses applications futures, tout en garantissant que le principe de finalité soit respecté.

Exemples de finalités considérées comme explicites et déterminées :

- Développement d'un grand modèle de langage (LLM) capable de répondre à des questions, générer du texte en fonction de contexte (courriels, lettres, rapports, y compris du code informatique), effectuer des traductions, résumés et corrections de texte, faire de la classification de texte, de l'analyse de sentiments, etc. ;

- Développement d'un modèle de reconnaissance vocale capable d'identifier un locuteur, sa langue, son âge, son genre, etc. ;

- Développement d'un modèle de vision par ordinateur capable de détecter différents objets comme des véhicules (voitures, camions, scooters, etc.), des piétons, du mobilier urbain (poubelles, bancs publics, abri-vélos, etc.) ou des éléments de signalisation routière (feux tricolores, panneaux routiers, etc.).

À l'inverse, se référer uniquement au type de système d'IA que l'on souhaite concevoir, sans se référer aux fonctionnalités et capacités techniquement envisageables, ne permet pas de considérer la finalité comme suffisamment précise.

Exemples de finalités qui ne sont pas considérées comme explicites et déterminées :

- Développement d'un modèle d'IA générative (les capacités envisageables ne sont pas définies) ;

- Développement et amélioration d'un système d'IA (ni le type de modèle ni les capacités envisageables ne sont définies) ;

- Développement d'un modèle permettant d'identifier l'âge d'une personne (le « type » n'est pas défini).

Point de vigilance : le responsable du développement du système d'IA à usage général devrait rappeler aux utilisateurs du système leur obligation de définir aussi précisément que possible la finalité pour laquelle le déploiement est prévu et d'en assurer la conformité. Cette conformité dépendra notamment de la prise en compte des risques spécifiques liés à cette finalité. Certains de ces risques devraient être anticipés dès la phase de développement : la CNIL recommande de prendre en compte dès la phase de développement les risques liés aux cas de déploiements connus ou raisonnablement envisageables, quand bien même l'utilisateur du système serait un autre responsable de traitement. Le cas échéant, la licence donnée à des utilisateurs tiers devrait permettre aux personnes concernées de connaître l'étendue de ces risques.

Comment définir la finalité du développement d'un système d'IA à des fins de recherche scientifique ?

Le responsable du traitement doit toujours définir l'objectif poursuivi par la recherche et le traitement de données mis en œuvre. Toutefois, en matière de recherche scientifique, il peut être admis que le degré de précision de cet objectif soit moins précis ou que les finalités de recherche ne soient pas spécifiées dans leur intégralité, compte tenu des difficultés que les chercheurs peuvent avoir à la cerner entièrement dès le début de leurs travaux. Il sera alors possible de fournir des informations pour préciser l'objectif à mesure que le projet progresse.

Rappel : qu'est-ce qu'une « recherche scientifique » au sens du RGPD ?

La notion de « recherche scientifique » bénéficie d'une acception large dans le RGPD. En synthèse, la recherche a pour objet de produire des connaissances nouvelles dans tous les domaines dans lesquels la méthode scientifique est applicable.

Afin d'aider les responsables de traitement à déterminer s'ils peuvent bénéficier des dispositions relatives à la recherche scientifique, la CNIL propose un faisceau de critères permettant d'aider le responsable de traitement à déterminer si le traitement qui poursuit une finalité de recherche, relève de la recherche scientifique :

Dans certains cas, il sera possible de présumer que la constitution de bases de données d'apprentissage pour l'IA poursuit une finalité de recherche scientifique en raison de la nature de l'organisme (par exemple, une université ou un centre de recherche public) ou du mode de financement (par exemple, financement par l'Agence nationale de la Recherche, ANR).

À défaut, notamment pour la recherche scientifique privée ne bénéficiant pas de financement public, il convient d'examiner conjointement les critères suivants (fondés sur le Manuel de Frascati de l'OCDE et sur sa définition de la R&D). Ces critères étant cumulatifs, le responsable de traitement devra en principe démontrer qu'ils sont tous remplis pour que le traitement puisse être considéré comme relevant de la recherche scientifique au sens du RGPD. Lorsque ce n'est pas le cas, une analyse au cas par cas est nécessaire pour qualifier le traitement.
La nouveauté : le traitement doit viser à obtenir des résultats nouveaux (une nouveauté pouvant aussi résulter d'un projet qui amène à constater des divergences potentielles avec le résultat censé être reproduit). L'objet de la recherche peut aider à la qualification de la recherche scientifique. À cet égard, la publication d'articles dans une revue à comité de lecture ou l'octroi d'un brevet permet de qualifier le critère de nouveauté.
La créativité : ce critère repose sur des notions et hypothèses originales et non évidentes – l'apport des travaux à la connaissance scientifique ou à l'état de la technique. Le développement d'un savoir collectif qui ne profite pas seulement à l'entité morale porteuse du projet de recherche est un indice fort pour qualifier celle-ci de scientifique.
L'incertitude : le traitement doit revêtir un caractère incertain quant au résultat final.
La systématicité : le traitement doit s'inscrire dans une planification et une budgétisation et mettre en œuvre une méthodologie scientifique. Le respect de normes sectorielles pertinentes de méthodologie et d'éthique est un indice fort pour qualifier la recherche de scientifique. C'est par exemple le cas des exigences méthodologiques particulières pour les traitements mis en œuvre à des fins de recherche, d'étude ou d'évaluation dans le domaine de la santé qui résultent notamment des articles 72 et suivants la loi « informatique et libertés ».
La transférabilité/reproductibilité : le traitement doit déboucher sur des résultats qu'il est possible de reproduire ou de transférer dans un champ plus large que celui de la recherche mise en œuvre. À titre d'exemple, la publication de l'étude réalisée et la présentation de la méthodologie de recherche adoptée est un indice fort permettant de souligner la volonté de partage du ou des porteurs de projet.
Exemple :

Pourrait être considéré comme poursuivant des fins de recherche scientifique le développement d'un système d'IA pour une preuve de concept destinée à démontrer la robustesse d'un apprentissage automatique nécessitant moins de données d'entraînement, qui s'inscrirait dans une démarche scientifique documentée ayant vocation à faire l'objet d'une publication.

3 - Déterminer la qualification juridique des acteurs

Responsable de traitement, responsable conjoint ou sous-traitant : la CNIL aide les fournisseurs de systèmes d'IA à déterminer leur qualification.

Les organismes constituant des bases de données d'apprentissage contenant des données personnelles doivent déterminer leur qualification au sens du RGPD : ils peuvent être responsable de traitement, responsable conjoint ou sous-traitant.

La CNIL vous aide à déterminer vos obligations en fonction de votre responsabilité et des modalités de collecte ou de réutilisation des données.
Plusieurs acteurs peuvent intervenir dans le développement d'un système d'IA, avec divers degrés d'implication sur les traitements de données personnelles. Il y a notamment :

- **le fournisseur de système d'IA** qui développe ou fait développer un système et qui le met sur le marché ou le met en service sous son propre nom ou sa propre marque, à titre onéreux ou gratuit.

- **les importateurs, distributeurs**, et les utilisateurs de ces systèmes (entendus comme les personnes déployant les systèmes d'IA).

La qualification des acteurs impliqués pour chaque traitement, au sens du RGPD, doit faire l'objet d'une analyse au cas par cas.

3 – 1 - Le responsable du traitement

3 – 1 – 1 - Le principe

Le responsable du traitement est la personne physique ou morale qui détermine les objectifs et les moyens du traitement, c'est-à-dire qui décide du « pourquoi » et du « comment » de l'utilisation de données personnelles.

Les moyens essentiels du traitement sont ceux qui sont étroitement liés à l'objectif et à la portée du traitement, tels que le type de données personnelles qui sont collectées et utilisées, les supports matériels et logiciels utilisés pour le traitement ainsi que leur sécurisation, la durée du traitement, les catégories de destinataires et les catégories de personnes concernées.

3 – 1 – 2 - En pratique

Certains indices peuvent aider à mener l'analyse au cas par cas pour déterminer qui est responsable du traitement.

Un fournisseur qui est à l'initiative du développement d'un système d'IA et qui en constitue la base de données d'apprentissage à partir de données qu'il a sélectionnées pour son propre compte, peut être qualifié de responsable de traitement.

Il en ira de même d'un fournisseur qui confie la constitution d'une telle base à un prestataire à travers des instructions documentées suffisamment précises (voir le rôle de sous-traitant ci-dessous).

A noter que dans certains cas, un fournisseur aura recours à un prestataire qui a déjà constitué un jeu de données en tant que responsable du traitement (de sa propre initiative). Il conviendra alors d'identifier les traitements pour lesquels le fournisseur est responsable, comme la réutilisation pour son propre compte d'un jeu de données déjà constitué.

Exemples de responsables de traitement :

- Une plateforme de vidéo à la demande souhaite développer un système d'IA de recommandations. Pour cela, il réutilise une base de données sur ses clients ayant été initialement collectée à des fins de fourniture du service.

  La plateforme de vidéo à la demande qui constitue la base de données pour entraîner son système d'IA de recommandations est responsable de ce nouveau traitement puisqu'elle a décidé de l'objectif (entraîner un système d'IA de recommandations) et des moyens essentiels du traitement (à savoir la base de données collectée pour une autre finalité).

- Le fournisseur d'un agent conversationnel qui entraîne son modèle de langage (« Large Language Model » ou LLM en anglais) à partir de données publiquement accessibles sur Internet, est responsable de traitement de la réutilisation des données personnelles publiquement accessibles sur Internet. En effet, il décide à la fois de l'objectif (proposer un agent conversationnel) et des moyens essentiels du traitement (sélectionner les données qu'il va réutiliser).

- Un fournisseur développe un système d'IA sur la base d'un modèle pré-entraîné avec des données personnelles. Il entend le ré-entraîner ou l'ajuster (fine-tuning ou transfer learning) avec un jeu de données qu'il a lui-même constitué, à son initiative. Dans un tel cas, ce fournisseur devra être qualifié de responsable de traitement, dès lors qu'il poursuit une finalité qui lui est propre et pour laquelle il détermine lui-même les moyens essentiels

La réutilisation de données collectées par un autre organisme

Lorsque le fournisseur entraine son système d'IA avec des données collectées par un autre organisme, il est nécessaire de distinguer :

le diffuseur des données : la personne physique ou morale, publique ou privée, qui met à disposition des données personnelles ou une base de données personnelles à des fins de réutilisation ;
le réutilisateur des données : la personne physique ou morale, publique ou privée, traitant ces données ou bases de données en vue d'une exploitation de celles-ci pour son propre compte.

Le diffuseur et le réutilisateur des données sont, en principe, responsables de traitements distincts puisque chacun détermine les objectifs et les moyens essentiels de son propre traitement.

Le diffuseur des données est, en principe, responsable de traitement de la diffusion, alors que le fournisseur du système d'IA qui réutilise les données, est lui responsable de traitement de la réutilisation. Le diffuseur n'est pas, en principe, responsable des traitements de réutilisation de ses données. Il peut toutefois prévoir des conditions d'utilisation des données diffusées pour en limiter les réutilisations ou prévoir certaines dispositions.

Exemple :

Une administration rend publiques et librement réutilisables (open data) des données immobilières. Une entreprise souhaite réutiliser ces données pour constituer une base de données d'apprentissage afin de développer un système d'IA consistant à prédire certaines évolutions immobilières sur un territoire donné. Le diffuseur et le réutilisateur sont alors responsables de traitement distincts, dès lors que ces deux traitements sont indépendants.

3 – 2 - Les responsables conjoints du traitement

3 – 2 – 1 - Le principe

Lorsque deux responsables du traitement ou plus, déterminent conjointement les finalités et les moyens du traitement, ils sont responsables conjoints du traitement.

Cette qualification peut être plus délicate en présence de plusieurs acteurs exerçant une influence sur la détermination des finalités et des moyens du traitement. Les acteurs doivent notamment déterminer s'ils traitent les données pour des objectifs propres et distincts ou pour un objectif commun.

3 – 2 – 2 - En pratique

Lorsqu'une base de données d'apprentissage d'un système d'IA est alimentée par plusieurs responsables de traitement pour un objectif conjointement défini, ces derniers peuvent être qualifiés de responsables conjoints du traitement.

Exemples :

- **Cas n°1** : des centres hospitaliers universitaires développant un système d'IA pour l'analyse de données d'imagerie médicale choisissent de recourir à un même protocole d'apprentissage fédéré. Ce dernier leur permet d'exploiter des données pour lesquelles ils sont initialement des responsables de traitement distincts, afin de ne pas s'en rendre mutuellement destinataires.
  Ils déterminent ensemble l'objectif (entraîner un système d'IA d'imagerie médicale) et les moyens de ce traitement (à travers le choix du protocole et la détermination des données qu'ils exploitent) : ils sont donc responsables conjoints de ce traitement d'apprentissage.

- **Cas n°**2 : un consortium composé d'une commune, d'une société fournissant un logiciel de traitement automatisé d'images et une société fournissant des dispositifs vidéo mettent en œuvre une expérimentation visant à installer des caméras augmentées pour enregistrer et analyser le flux et le comportement des véhicules empruntant une voie de circulation au sein de la commune. Le contrat passé entre la ville et les deux sociétés prévoit l'utilisation du logiciel par la commune en conditions réelles et la possibilité pour les sociétés d'améliorer le logiciel de traitement automatisé d'images par les données collectées en temps réel. Cette amélioration du logiciel de traitement automatisé bénéficie aussi bien à la commune qu'aux sociétés fournissant le logiciel de traitement automatisé d'images et les dispositifs vidéo.

  La commune et les deux sociétés seraient ainsi responsables conjoints du traitement de constitution de la base de données d'apprentissage du logiciel de traitement automatisé d'images dès lors qu'ils décident conjointement de la finalité et des moyens essentiels du traitement et que les sociétés n'agissent pas uniquement pour le compte de la commune. En effet, il est possible de considérer qu'ils décident conjointement des moyens essentiels du traitement (en choisissant d'alimenter la base de données d'apprentissage du système d'IA par les données collectées en temps réel par les caméras augmentées et par les données déjà collectées par la société fournissant le logiciel de traitement automatisé d'images) et de l'objectif du traitement (entraîner de manière expérimentale un système d'IA qui permet de détecter les comportements particuliers de véhicules et améliorer le logiciel de traitement automatisé d'images).

  A l'inverse, si l'une des sociétés entend réutiliser les données pour une finalité propre dont elle serait la seule à bénéficier (par exemple dans un cadre de recherche et développement), il pourrait alors être envisagé de considérer qu'elle soit responsable d'un traitement distinct. En cas de responsabilité conjointe, les responsables de traitement conjoints doivent s'assurer de la licéité du traitement (c'est-à-dire de sa conformité à la loi), notamment en définissant de manière transparente leurs obligations respectives dans le cadre d'un accord. La forme de cet accord n'est pas précisée par le RGPD. L'accord doit refléter les rôles de chacune des parties prenantes, les responsables conjoints devant préciser de manière claire « qui fait quoi » pour assurer la protection des données traitées.

  À noter : indépendamment des termes de l'accord, la personne concernée par le traitement peut exercer ses droits à l'égard de chacun des responsables conjoints du traitement.

3 – 3 - Le recours à un sous-traitant

3 – 3 - 1 - Le principe

Le sous-traitant est la personne physique ou morale qui traite des données pour le compte du responsable de traitement, dans le cadre d'un service ou d'une prestation.

3 – 3- 2 - En pratique

La qualification du fournisseur de système d'IA doit être appréciée au cas par cas.

Un fournisseur de système d'IA peut être sous-traitant lorsqu'il développe un système d'IA pour le compte d'un de ses clients dans le cadre d'une prestation. Le client est pour sa part responsable de traitement dès lors qu'il détermine la finalité et les moyens du traitement.

Dans d'autres configurations, le fournisseur de système d'IA peut être responsable de traitement des systèmes qu'il conçoit pour les commercialiser.

Un fournisseur de système d'IA peut faire appel à un prestataire pour collecter et traiter les données selon ses instructions documentées (par exemple, de collecter des données publiquement accessibles sur Internet, de réutiliser une base de données spécifique mise à disposition en ligne, etc.). Ce dernier est alors qualifié de sous-traitant. Il est essentiel pour le fournisseur du système d'IA, en tant que responsable du traitement, de s'assurer que son sous-traitant respecte le RGPD et limite le traitement des données à ses instructions, notamment en concluant un contrat de sous-traitance.

Par ailleurs, le fait d'avoir recours à la même base de données pour différents clients, dans le cadre de prestations de services distinctes, est généralement un indice décisif permettant de considérer que le prestataire est responsable d'un traitement distinct, à minima pour la constitution de la base de données.

3 – 3 – 3- Exemples :

Un prestataire s'est vu confier la constitution d'une base de données d'apprentissage par un fournisseur de système d'IA qui lui a indiqué précisément comment elle doit être élaborée (en particulier s'agissant des sources et des catégories de données, avec des exigences de qualité et de documentation). Ce prestataire agira vraisemblablement en qualité de sous-traitant.

A l'inverse, un prestataire qui aurait, à son initiative, constitué un jeu de données qu'il exploite en développant des systèmes d'IA adaptés au besoin de chacun de ses clients sera vraisemblablement responsable du traitement de ce jeu de données, indépendamment de son rôle dans les traitements spécifiques réalisés

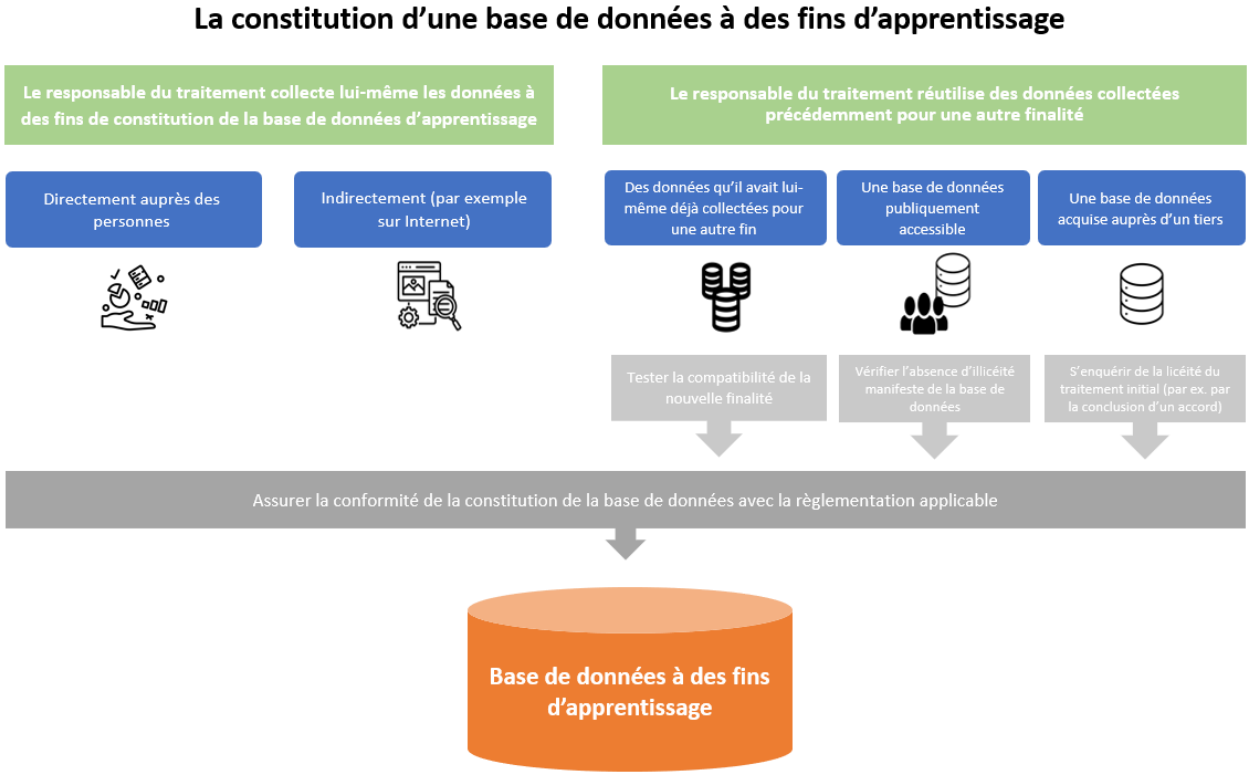
pour ces clients (qu'il pourrait mettre en œuvre en tant que sous-traitant, par exemple sur la base de données fournies par les clients eux-mêmes).

4 – Assurer que le traitement est licite - Définir une base légale

La CNIL vous aide à déterminer vos obligations en fonction de votre responsabilité et des modalités de collecte ou de réutilisation des données.
L'organisme qui souhaite constituer une base de données d'apprentissage contenant des données personnelles puis l'utiliser pour entraîner un algorithme doit s'assurer que le traitement est autorisé par la loi. La CNIL vous aide à déterminer vos obligations en fonction de votre responsabilité et des modalités de collecte ou de réutilisation des données
Le responsable de traitement doit définir, dans tous les cas, une base légale et doit effectuer, en fonction du mode de collecte ou de réutilisation des données, certaines vérifications supplémentaires.

Il existe plusieurs moyens de constituer une base de données à des fins d'apprentissage, qui peuvent être utilisés de manière cumulative :

- les données sont collectées directement auprès des personnes ;
- les données sont collectées à partir de sources ouvertes sur Internet pour cette finalité ;
- les données ont initialement été collectées pour un autre objectif par le responsable du traitement lui-même (par exemple, dans le cadre de la fourniture d'un service à ses utilisateurs) ou par un autre responsable de traitement. Cela implique de prendre des précautions complémentaires.



Définir une base légale

4 – 1 - Le principe

Comme tout traitement, la constitution et l'utilisation d'une base de données pour l'entraînement de modèle ou le développement de systèmes d'IA contenant des données personnelles ne pourra être mise en œuvre que si elle correspond à l'une des « bases légales » prévues par le RGPD.

Concrètement, la base légale est ce qui donne le droit à un organisme de traiter des données personnelles. Le choix de cette base légale est donc une première étape indispensable pour assurer la conformité du traitement. Selon celle qui sera retenue, les obligations de l'organisme et les droits des personnes pourront varier.

Les bases légales les plus pertinentes pour l'entraînement d'un algorithme sont détaillées ci-après.

4 – 2 - En pratique

La détermination de la base légale doit être effectuée de manière adaptée à la situation et au type de traitement. Pour constituer une base de données pour l'entraînement d'un système d'IA, les bases légales suivantes peuvent notamment être envisagées.

4 – 3 - La base légale du consentement

Pour être valide, le consentement des personnes concernées doit remplir quatre critères cumulatifs : il doit être libre, spécifique, éclairé et univoque. Le responsable du traitement doit être en mesure de démontrer la validité du recours à cette base légale en s'assurant que chacune de ces conditions, précisément définies par le RGPD, est remplie.

Exemple : un organisme souhaite filmer ou photographier des volontaires pour constituer une base de données d'images permettant d'entraîner un système à détecter certains gestes spécifiques. Il peut fonder le traitement sur la base de leur consentement. Lorsqu'il constitue une base de données pour l'apprentissage d'un modèle d'IA, un organisme doit s'assurer du consentement recueilli.

Au-delà des obligations liées à la transparence, le consentement doit être accompagné d'un certain nombre d'informations communiquées à la personne avant qu'elle ne consente afin de lui permettre de prendre des décisions en toute connaissance de cause et de savoir comment retirer son consentement.

Le consentement doit porter sur une finalité spécifique (voir la fiche n°2 sur la définition de la finalité).

La liberté du consentement implique, en principe, de garantir la possibilité pour les personnes concernées de donner leur consentement de manière granulaire, lorsque les finalités poursuivies sont distinctes.

Exemple : le consentement de personnes à l'utilisation de leur image, collectée lors d'un événement d'une entreprise à des fins de communication, ne signifie pas qu'ils consentent à une réutilisation des données pour la constitution d'une base de données d'apprentissage ou d'amélioration d'un système d'IA. Dans ce cas, deux consentements distincts doivent être recueillis (par exemple au moyen de deux cases à cocher).
La liberté du consentement doit également faire l'objet d'une certaine vigilance en cas de déséquilibre de rapports de force entre la personne concernée et le responsable du traitement, en particulier si ce dernier est une autorité publique ou un employeur.

Exemple : pour développer un système d'IA, une entreprise souhaite utiliser les données de ses salariés. Leur consentement ne peut alors être valablement recueilli que dans des situations exceptionnelles, lorsqu'ils sont en mesure de refuser de donner leur consentement sans craindre ou encourir de conséquences négatives. En tant que responsable du traitement, l'entreprise devra veiller, en tout état de cause, à ce que les communications destinées à présenter le dispositif aux salariés ne soient ni incitatives, ni contraignantes. Elle devra informer les volontaires de la possibilité de ne plus participer à la collecte de leurs données à tout moment, sans que cela porte à conséquence pour ces derniers.
Il n'apparaît pas possible de recueillir un consentement valide dans certains cas. C'est souvent le cas lorsque le responsable du traitement collecte des données accessibles en

ligne ou réutilise une base de données ouverte, compte tenu notamment de l'absence de contact avec les personnes concernées et de la difficulté à les identifier. Dans ces cas de figure le responsable du traitement doit mobiliser une autre base légale plus adaptée.

Il peut également exister des difficultés liées au droit de retirer son consentement, par exemple du fait d'obstacles techniques à l'identification des personnes concernées. S'il n'est pas possible, pour le responsable du traitement, de garantir la possibilité d'exercer ce droit, il est recommandé de se fonder sur une base légale.

4 – 4 - La base légale de l'intérêt légitime

L'intérêt légitime du responsable de traitement peut être retenu sous réserve du respect des conditions suivantes :

la légitimité de l'intérêt poursuivi par le responsable de traitement. Par exemple l'intérêt pour un organisme de développer un modèle en vue de la commercialisation d'un système d'IA ou encore en vue de contribuer à l'amélioration de la connaissance scientifique, par exemple par la publication des outils développés (code, modèle, protocole expérimental, etc.) et des résultats de recherche.
la nécessité du traitement de données. Par exemple, le traitement à des fins de constitution d'une base de données d'apprentissage contenant des images de personnes peut être considéré comme nécessaire aux intérêts d'un organisme qui souhaite développer un système de détection de pose, lorsque des données anonymes ou synthétiques ne suffisent pas.
l'absence d'atteinte disproportionnée aux intérêts et droits des personnes concernées, compte tenu de leurs attentes raisonnables à l'égard de ce traitement. La mise en balance des droits et intérêts en cause dépend des caractéristiques concrètes du traitement envisagé et notamment des garanties mises en œuvre pour assurer le meilleur équilibre possible entre ces intérêts et limiter les impacts du traitement sur les personnes concernées.
Le plus souvent, le fait de constituer une base de données pour l'entraînement d'un modèle dont l'usage est lui-même légal peut être regardé comme légitime. Une analyse au cas par cas est cependant nécessaire pour déterminer si l'utilisation de données personnelles à cette fin ne porte pas une atteinte disproportionnée à la vie privée des personnes concernées, et ce même lorsque les données ne sont pas nominatives. Pour assurer que son traitement est proportionné, le responsable de traitement peut notamment recourir à des mesures telles que la pseudonymisation des données, garantir l'absence de données sensibles, définir des critères de sélection permettant de limiter la collecte aux données pertinentes et nécessaires pour le traitement, etc.

Exemples :
Une entreprise souhaite développer un système d'IA capable de prédire le profil psychologique d'une personne à partir de données accessibles en ligne susceptibles de la concerner. Son intérêt commercial à développer un tel système sera vraisemblablement

insuffisant au regard des intérêts, droits et libertés des personnes concernées : une autre base légale devra être recherchée ou le projet abandonné.

Un organisme constitue une base de données d'apprentissage en collectant les commentaires rendus publics et librement accessibles par des utilisateurs en ligne sur des forums, blogs et sites web. La finalité de ce traitement est de concevoir un système d'IA permettant d'évaluer et de prévoir l'appréciation d'œuvres d'art par le grand public. Dans ce cas, son intérêt à développer et éventuellement commercialiser un système d'IA peut être considéré comme légitime. La collecte de commentaires d'appréciation sur les œuvres peut être considérée comme nécessaire pour le développement du modèle, notamment compte tenu de la quantité de données requises pour l'apprentissage. Il convient de noter que la base légale de l'intérêt légitime donne le droit aux personnes concernées de s'opposer au traitement de leurs données (pour des motifs tenant à leur situation particulière).

4 – 5 - La base légale de la mission d'intérêt public

La possibilité de se fonder sur la base légale de la « mission d'intérêt public » suppose :
que la mission dans laquelle s'inscrit le traitement soit prévue par un texte normatif applicable au responsable du traitement ;
que l'utilisation des données permette d'exercer spécifiquement cette mission de manière pertinente et appropriée.
Exemples :
Les chercheurs d'un laboratoire de recherche public sur la langue française souhaitent analyser l'évolution de l'utilisation de la langue en ligne. Ils constituent pour cela une base de données à partir des commentaires publiés librement en ligne sur différents réseaux sociaux (anonymisés à bref délai) afin d'entraîner un modèle qui détecte et analyse automatiquement l'occurrence de certaines expressions ou formes orthographiques.

Dans la mesure où le responsable de traitement est un laboratoire public, les chercheurs peuvent dans ce cas fonder le traitement de données sur la mission d'intérêt public. Cette base légale peut être mobilisée, de manière générale, pour les traitements de données effectués par des laboratoires de recherche publics ou privés investis d'une mission d'intérêt publique, dont les traitements de données sont nécessaires pour leur activité de recherche.

Le pôle d'expertise de la régulation numérique (PEReN) est autorisé à réutiliser, dans certaines conditions, des données publiquement accessibles de certaines plateformes afin de réaliser des expérimentations ayant notamment pour objet de concevoir des outils techniques destinés à la régulation des opérateurs de plateformes en ligne, conformément à l'article 36 de la loi n° 2021-1382 du 25 octobre 2021 et au décret n° 2022-603 du 21 avril 2022.

:

?

4 – 6 - La base légale du contrat

La base légale du contrat pourrait être mobilisée pour la constitution d'une base de données d'apprentissage d'un système d'IA à condition, d'une part, qu'un contrat valide soit conclu entre le responsable et la personne concernée et, d'autre part, que le traitement soit objectivement nécessaire à son exécution.

Les contrats conclus à cette fin doivent respecter les autres règles applicables, en matière de droit du travail ou de propriété intellectuelle par exemple.

Exemples :
Un éditeur de logiciel de traitement de texte propose un service de génération automatisée et personnalisée de courriers, auquel l'utilisateur souscrit contractuellement, et pour lequel il collecte les données des utilisateurs bénéficiaires de ce service.. Le traitement des données pour ce service de personnalisation peut être considéré, sous réserve des caractéristiques spécifiques du traitement de données, nécessaire à l'exécution du contrat.

À l'inverse, l'opérateur d'un réseau social en ligne inscrit dans ses conditions générales d'utilisation qu'il entend réutiliser les données de ses utilisateurs (fournies par ces derniers, observées ou déduites par l'opérateur) pour développer et améliorer de nouveaux produits, services et fonctionnalités utiles pour ses utilisateurs. Il ne peut pas fonder le traitement sur la base légale du contrat dès lors que ce traitement n'est pas objectivement indispensable pour leur offrir son service de réseau social en ligne (CJUE, 4 juillet 2023, Meta Platforms Inc. et a. c/Bundeskartellamt, C-252/21).

Données sensibles : un traitement interdit, sauf exceptions

Les données sensibles sont une catégorie particulière de données personnelles définies à l'article 9 du RGPD. Constituent par exemple des données sensibles des données qui révèlent la prétendue origine raciale ou ethnique des personnes concernées, ou encore des données biométriques aux fins d'identifier une personne physique de manière unique, comme un gabarit facial par exemple.

Le RGPD interdit le traitement de ces données, sauf exception, seulement dans les cas énumérés dans son article 9.2. Ces exceptions incluent notamment :

les traitements pour lesquels la personne concernée a donné son consentement explicite (démarche active, explicite et de préférence écrite, qui doit être libre, spécifique, et informée) ;

les traitements portant sur des données personnelles qui sont manifestement rendues publiques par la personne concernée ;
Dans ses lignes directrices sur le ciblage des utilisateurs des réseaux sociaux, le CEPD fournit une liste d'éléments à prendre en compte pour déterminer si les données sont manifestement rendues publiques : le paramétrage par défaut de la plateforme de réseaux sociaux, la nature de la plateforme, l'accessibilité de la page concernée, la visibilité de l'information sur leur caractère public, le point de savoir si la personne concernée a elle-même publié les données ou si elles l'ont été par un tiers ou déduites.
Il importe de vérifier si la personne concernée souhaitait, de manière explicite et par un acte positif clair, sur la base d'un paramétrage effectué en connaissance de cause, rendre accessibles au grand public ses données personnelles ou, au contraire, à un nombre plus ou moins limité de personnes sélectionnées (CJUE, 4 juillet 2023, Meta Platforms, C-252/21).

les traitements nécessaires pour des motifs d'intérêt intérêt public important, sur la base du droit de l'UE ou d'un État membre ;

les traitements nécessaires à des fins de recherche scientifique sur la base du droit de l'Union européenne ou d'un État membre
Il convient de faire preuve d'une attention particulière à la collecte de données sensibles lors de l'utilisation d'outils de moissonnage (web scraping) qui impliquent le traitement de larges volumes de données. Le responsable du traitement est tenu de mettre en œuvre toutes les mesures permettant d'exclure automatiquement la collecte des données sensibles non pertinentes notamment notamment en appliquant des filtres permettant d'exclure la collecte de certaines catégories de données ou encore d'exclure certains sites comportant des données sensibles par nature.

Si, malgré les mesures prises, l'organismes traite de manière incidente et résiduelle des données sensibles qu'il n'avait pas cherché à collecter, cela n'est pas considéré comme illégal. C'est notamment ce qu'a pu considérer la cour de justice de l'Union européenne en rappelant que cette interdiction s'applique à l'exploitant d'un moteur de recherche « dans le cadre de ses responsabilités, de ses compétences et de ses possibilités » (CJUE, grande chambre, 24 septembre 2019, GC e.a, C-136/17). En revanche, si l'organisme vient à savoir qu'il traite des données sensibles, il est tenu de procéder, autant que possible, à leur suppression immédiate et automatisée de la base de données.

A noter :

Une fiche sur la gestion des biais sera publiée ultérieurement. Elle permettra d'éclaircir la possibilité de traiter des données sensibles à des fins de détection et de correction de biais dans la base de données d'apprentissage.

La CNIL mène actuellement des travaux sur la question de l'IA dans le domaine de la santé, qui feront l'objet d'une publication ultérieure.

4 –7 - La base de l’obligation légale

Si cette base légale peut sembler pertinente dans certains cas pour les traitements de données effectués en phase de déploiement, dans la mesure où l’utilisation d’un système d’IA peut parfois servir au responsable du traitement pour respecter une obligation légale (à condition de démontrer que celle-ci impose un traitement de données personnelles), elle est, en revanche, plus difficile à mobiliser pour fonder son développement.

En effet, pour mobiliser cette base légale, le responsable du traitement doit démontrer en quoi son traitement est nécessaire pour répondre à une obligation légale déterminée à laquelle il est soumis. Le texte sur lequel elle repose doit au moins définir la finalité du traitement et peut l’encadrer de manière plus précise (notamment à travers les types de données à traiter, la limitation des finalités ou d’autres conditions à respecter). Plus l’obligation légale est précise, plus il est facile de justifier en quoi elle impose un traitement de données personnelles.

Toutefois, les obligations n’étant généralement pas suffisamment précises pour prévoir le développement de systèmes d’IA, il conviendra le plus souvent de se fonder sur une autre base légale pour développer ce type de système.

Exemples : Dans le secteur de l’assurance, les études actuarielles reposent sur des modélisations mathématiques, probabilistes et statistiques assimilables à des systèmes d’IA, dont l’objectif est de permettre d’identifier, qualifier et quantifier des risques (ainsi que les montants associés) liés aux contrats d’assurance.

Or, les obligations générales de solvabilité des organismes d’assurance n’étant pas suffisamment précises, il n’est pas possible de considérer que le développement de tels systèmes soit nécessaire à leur respect. L’intérêt légitime apparait alors comme la base légale la plus pertinente.

5 - IA : Assurer que le traitement est licite -En cas de réutilisation des données, effectuer les tests et vérifications nécessaires

En cas de réutilisation de données, le responsable du traitement est tenu d’effectuer certaines vérifications supplémentaires afin de garantir que le traitement de données est autorisé par la loi. La CNIL vous aide à déterminer vos obligations en fonction des modalités de collecte et de la source des données

5 – 1 - Le principe

Dans certains cas, en fonction des modalités de collecte et de la source des données utilisées pour la constitution de la base de données d’apprentissage, le responsable du traitement est tenu d’effectuer certaines vérifications afin de garantir que le traitement de données est autorisé par la loi. Ces vérifications s’ajoutent à l’identification de la base légale du traitement de données.

5 - 2 - En pratique

Le fournisseur réutilise les données qu'il a lui-même collectées initialement pour une autre finalité

Un responsable de traitement peut vouloir réutiliser les données qu'il a collectées pour une finalité initiale (par exemple, dans le cadre de la fourniture d'un service à des particuliers) afin de constituer une base de données à des fins d'apprentissage d'un système d'IA.

Dans ce cas, il doit déterminer si ce traitement ultérieur est compatible avec la finalité pour laquelle les données ont été initialement collectées, lorsque le traitement ne s'appuie pas sur le consentement de la personne concernée ou sur le droit de l'Union ou le droit d'un État membre.

L'obligation d'effectuer ce « test de compatibilité » s'applique aux traitements ultérieurs de données, (au sens de l'article 6.4 du RGPD), c'est-à-dire ceux :

- qui n'ont pas été prévus ni portés à la connaissance des personnes concernées lors de la collecte des données ;
- qui sont effectués par un même responsable de traitement qui décide de réutiliser des données pour une finalité distincte de celle pour laquelle elles ont été collectées, y compris quand il s'agit de les publier sur Internet ou de les partager avec des tiers à des fins de réutilisation pour une autre finalité.

À noter :

Aucun test de compatibilité n'est requis pour les finalités prévues et portées à la connaissance des personnes concernées dès la collecte dans le respect du principe de transparence, y compris lorsque certaines d'entre elles peuvent paraître secondaires ou accessoires. Par exemple, le partage de données par un responsable de traitement avec son sous-traitant pour l'amélioration de la performance de son algorithme ne nécessite pas d'effectuer un test de compatibilité, si cette finalité était prévue et portée à la connaissance de la personne concernée (sous réserve de respecter les conditions de légalité pour cette finalité d'amélioration de l'algorithme).

Pour réaliser ce **« test de compatibilité »** il doit notamment prendre en compte :

- l'existence d'un lien entre la finalité initiale et la finalité du traitement ultérieur envisagé ;
- le contexte dans lequel les données personnelles ont été collectées, en particulier les attentes raisonnables des personnes concernées, en fonction de la relation entre les personnes concernées et le responsable du traitement ;
- le type et la nature des données, en particulier en fonction de leur sensibilité (données biométriques, de géolocalisation, concernant des mineurs, etc.) ;
- les éventuelles conséquences du traitement ultérieur envisagé pour les personnes concernées ;

- l’existence de garanties appropriées (telles que le chiffrement ou la pseudonymisation).

Exemples :
Le fournisseur d’un éditeur de texte grand public lance une fonctionnalité d’IA générative permettant de compléter certaines phrases ou certains paragraphes (auto-saisie). Quelques temps après le déploiement de cette fonctionnalité, il souhaite réutiliser les corrections manuelles apportées par les utilisateurs au contenu des textes ainsi générés, afin de proposer à chaque utilisateur de disposer d’une version personnalisée de son service de recommandation (par exemple pour mieux comprendre et anticiper sa manière d’écrire) sur la base de leurs données respectives.
Une plateforme de streaming vidéo grand public envisage désormais de réutiliser l’historique et les listes de lecture qu’elle a enregistrés dans le cadre de la fourniture du service pour proposer à chaque utilisateur de disposer d’une version personnalisée de son service de recommandation (par exemple pour mieux anticiper et comprendre ses préférences) sur la base de leurs données respectives.

Dans ces deux cas, la nouvelle finalité pourra être considérée comme compatible avec la finalité initiale de la fourniture du service, à condition que les garanties mises en œuvre soient suffisantes (par exemple, grâce à la possibilité de s’opposer à cette réutilisation, sans avoir à fournir de motif) sur la base de leurs données respectives.
Lorsque la réutilisation des données poursuit des fins statistiques ou de recherche scientifique, le traitement est présumé compatible avec la finalité initiale s'il respecte le RGPD et s'il n'est pas utilisé pour prendre des décisions à l'égard des personnes concernées. La réalisation du « test de compatibilité » n’est donc pas nécessaire.

Pour poursuivre une finalité statistique au sens du RGPD, le traitement ne doit tendre qu’à la production de données agrégées pour elles-mêmes : le traitement doit avoir pour unique objet le calcul des données, leur affichage ou publication, leur éventuel partage ou communication (et non à la prise de décisions ultérieures, individuelles ou collectives). Les résultats statistiques ainsi obtenus doivent constituer des données agrégées et anonymes au sens de la réglementation sur la protection des données. Le recours à des techniques statistiques d’apprentissage automatique ne suffit pas à considérer qu’il s’agit de traitements « à des fins statistiques », dans la mesure où la finalité du traitement de données n’est pas de produire des données agrégées pour elles-mêmes. Le recours à cette technique est davantage un moyen mis en œuvre pour l’apprentissage du modèle.

La notion de « recherche scientifique » est entendue largement dans le RGPD. En synthèse, la recherche a pour objet de produire des connaissances nouvelles dans tous les domaines dans lesquels la méthode scientifique est applicable. Tout traitement de données à des fins de recherche scientifique doit être soumis à des garanties appropriées pour les droits et libertés de la personne concernée, telles que l’anonymisation ou la pseudonymisation (mentionnées à l’article 89 du RGPD).

:

À noter :

Même lorsque le traitement ultérieur est compatible, une base légale valable doit toujours être identifiée et les personnes informées, notamment pour pouvoir exercer leurs droits..

Focus : à quelles conditions peut-on réutiliser un jeu de données initialement constitué à des fins de recherche scientifique ?

Le RGPD facilite la réutilisation de données à des fins de recherche scientifique : cette réutilisation est jugée compatible avec la finalité initiale du traitement et certaines dérogations (notamment aux droits des personnes) sont possibles.

En revanche, lorsqu'un responsable du traitement a traité des données à des fins de recherche scientifique et qu'il entend les réutiliser à d'autres fins (pour son propre compte ou pour les transmettre à un tiers), il doit respecter certaines conditions. À cet égard, la réutilisation d'un jeu de données sera possible :

si les données ont été préalablement anonymisées, ou

si la réutilisation est compatible avec la finalité pour laquelle le responsable du traitement a collecté les données (selon le « test de compatibilité » détaillé ci-dessus) et que le nouveau traitement est mis en œuvre dans le respect du RGPD (information des personnes au sujet de cette nouvelle finalité, identification d'une base légale, etc.). Les dérogations permises par le RGPD pour la recherche scientifique ne seront plus mobilisables.

En cas de transmission des données à des tiers, la compatibilité des réutilisations ultérieures avec la finalité de recherche pourra être garantie notamment par une licence de réutilisation.

Le fournisseur réutilise une base de données publiquement accessibles

Des bases de données contenant des données personnelles peuvent être librement mises à disposition sur Internet en dehors du cadre légal relatif à l'ouverture des données (« open data »). Le plus souvent, il s'agit de données qui étaient déjà publiquement accessibles et qui constituent une base de données ou un corpus diffusé sur le site web d'une université ou d'une plateforme dédiée au partage de bases de données, pour faciliter leur réutilisation.

Le contrôle du caractère licite de la mise en ligne de la base de données relève en premier lieu du responsable du traitement qui opère cette mise en ligne (le cas échéant en s'assurant qu'il s'agit d'un traitement ultérieur compatible s'il n'a pas initialement collecté les données dans ce but). Cependant, afin de pouvoir se prévaloir d'une base légale au titre du RGPD, le responsable du traitement qui réutilise les données doit s'assurer qu'il n'est pas en train de réutiliser une base de données dont la constitution était manifestement illicite (par exemple, provenant d'une fuite de données).

Le réutilisateur ne peut pas réutiliser une base de données constituée ou mise en ligne dont il ne peut ignorer qu'elle ne respecte pas le RGPD (article 5.1.a du RGPD) ou d'autres règles, telles que celles interdisant les atteintes à la sécurité des systèmes d'information ou les atteintes à des droits de propriété intellectuelle.

En outre, la personne qui télécharge ou réutilise une base de données manifestement illégale risque de se rendre coupable du délit de recel (article 321-1 du code pénal).

Si la possibilité de réutiliser une base de données librement mise à disposition sur Internet n'est pas nécessairement subordonnée à des vérifications approfondies sur le respect de l'ensemble des règles du RGPD ou d'autres règles juridiques applicables (droit d'auteur, données couvertes par le secret des affaires, etc.), vérifications qui relèvent en premier lieu de l'organisme qui met en ligne les données, un organisme ne peut réutiliser une base de données qui serait manifestement illicite.

Cette illicéité manifeste doit s'apprécier au cas par cas. À ce titre, la CNIL recommande aux réutilisateurs de s'assurer :

- **Que la description de la base de données mentionne leur source.**
  Exemple : une base de données dont la description expliquerait qu'elle a été constituée à partir de publications sur un réseau social professionnel nommément désigné.
  À l'inverse, si une base de données contenant des images de vidéosurveillance ne précise pas la source, une telle base ne devrait pas être réutilisée avant d'avoir obtenu davantage de précisions permettant de lever les doutes quant à la conformité de sa constitution et de sa diffusion ;
- **Que la constitution ou la diffusion de la base de données ne résulte pas manifestement** d'un crime ou d'un délit ou a fait l'objet d'une condamnation ou d'une sanction publique de la part d'une autorité compétente qui a impliqué une suppression ou l'interdiction d'exploiter ultérieurement les données ;
  Exemples :
  une entreprise souhaite constituer une base de données pour le développement d'un système d'IA de recommandation qu'il entend utiliser auprès de ses consommateurs. S'il acquiert pour cela une base de données sur le dark web provenant, par exemple, d'une atteinte à un système de traitement automatisé punie par la loi (au sens de l'article 323-1 du code pénal), il ne saurait en ignorer l'origine délictuelle. Dans ce cas, le caractère illicite de la base de données serait alors manifeste.
  Il en irait de même pour une entreprise souhaitant réutiliser une base de données pour laquelle une décision de justice a retenu une atteinte à un droit de propriété intellectuelle comme celui, particulier, des producteurs de bases de données (au sens de l'article L. 342-1 du code de la propriété intellectuelle) ;
  - Qu'il n'y ait pas de doutes flagrants sur le fait que la base de données est licite (notamment que le traitement source ne soit pas manifestement dépourvu de base légale lorsque les données sont tellement intrusives qu'elles ne sauraient être traitées sans le consentement des personnes), en s'assurant en particulier que les conditions de collecte des données soient suffisamment documentées ;

  Exemples :
  Sur une plateforme d'hébergement de bases de données, une entreprise repère un ensemble compilant les trajets domicile-travail de milliers de personnes. Sa description explique qu'il s'agit de données de géolocalisation précises, non anonymes, sans en détailler la source. Dans cette hypothèse, elle ne saurait ignorer qu'il existe un doute sérieux quant à la licéité de la diffusion d'une telle base de données sans le consentement des personnes.

À l'inverse, il serait envisageable de constituer une base de données à partir d'une base de données dont la description ne laisse pas de doute flagrant quant à sa licéité. Par exemple, une base de données pseudonymisées, initialement rendues publiques par les personnes concernées sur un site web identifié et qui ne contiendrait pas de données sensibles.

Il en irait de même pour la réutilisation d'une base de données agrégées que le diffuseur présenterait comme anonymes. Par exemple, un organisme qui souhaite constituer une base de données pour entraîner un système d'IA destiné à prévoir l'impact socio-économique du vieillissement d'une population pourrait réutiliser des bases de données anonymes agrégées contenant notamment des informations démographiques (nombre de personnes actives, âge des personnes, taux de fécondité ou encore taux de dépendance des personnes âgées).

- Que la base de données ne contient pas de données sensibles (données de santé ou révélant des opinions politiques par exemple) ou de données d'infraction (au sens des articles 9 et 10 du RGPD), ou, si elle en contient, il est recommandé de mener des vérifications supplémentaires pour s'assurer que ce traitement était licite (il s'agirait principalement pour les données sensibles de s'assurer du recueil d'un consentement explicite des personnes concernées, ou que les données ont été manifestement rendues publiques par ces dernières comme cela est précisé ci-dessous et pour les données relatives à des infractions qu'une telle utilisation est rendue possible par la loi informatique et libertés).

Exemple : sur un forum en ligne, un chercheur découvre une base de données non anonymes qui contiendrait, selon sa description, les parcours de soin d'une centaine de patients atteints d'une pathologie particulière et qui proviendraient d'hôpitaux français. Dans ce cas, le chercheur devrait sérieusement douter que la diffusion de ce jeu de données soit licite compte tenu de l'encadrement des données de santé prévu par le RGPD et la loi « informatique et libertés ».
Ces vérifications préalables pourraient utilement figurer dans l'analyse d'impact relative à la protection des données (AIPD).

Certains manquements commis par le responsable des traitements de constitution et de diffusion d'une base de données n'impactent pas systématiquement et irrémédiablement la licéité des traitements mis en œuvre par le réutilisateur. Ainsi, un réutilisateur peut utiliser une base de données dont les illicéités sont mineures, à condition que la réutilisation satisfasse les exigences du RGPD.

Exemple : la fourniture de mentions d'informations incomplètes lors de la constitution ou de la diffusion de la base de données, ou un défaut de documentation adaptée de la conformité de ces traitements (qu'il est nécessaire de vérifier avec le diffuseur ou l'éditeur de la base de données).
Le fournisseur réutilise une base de données acquise auprès d'un tiers (courtiers en données, etc.)
Certains fournisseurs souhaitent constituer une base de données d'apprentissage à partir de bases de données détenues par des tiers.

Pour le tiers qui partage des données personnelles, cela implique de s'assurer de la licéité de cette transmission

- **Cas n°1** : les données ont précisément été collectées en vue d'être partagées à des fins de constitution d'une base de données pour l'apprentissage de système d'IA

  Le tiers devra s'assurer de la conformité du traitement de transmission des données au regard du RGPD (définition d'une finalité explicite et légitime, exigence d'une base légale, information des personnes et gestion de l'exercice de leurs droits, etc.) dont il assume la responsabilité.

- **Cas n°2** : le tiers n'a pas initialement collecté les données pour cette finalité

  Lorsque le tiers a initialement collecté les données pour d'autres finalités (par exemple dans le cadre de la fourniture d'un service aux personnes concernées), il lui appartient de s'assurer que la transmission de ces données poursuit une finalité compatible avec celle(s) ayant justifié leur collecte. Il devra donc réaliser un « test de compatibilité ».

À noter que le détenteur initial d'une base de données autorise parfois son utilisation dans le cadre d'un contrat de licence qui en prévoit les termes et les conditions (notamment au titre du droit de la propriété intellectuelle). Ce contrat de licence peut par exemple encadrer cette compatibilité en limitant les réutilisations possibles.

Pour le réutilisateur, cela implique le plus souvent une série de vérifications des traitements du responsable de traitement initial

Le responsable du traitement doit s'assurer qu'il n'est pas en train de réutiliser une base de données dont la constitution ou le partage était manifestement illicite (par exemple, en l'absence d'indication quant à sa source, en cas de doute flagrant sur sa licéité, en particulier dans le cas de traitement de données sensibles, etc.). Cela résulte du principe général de licéité des traitements de l'article 5.1.a du RGPD, outre le risque de se rendre coupable du délit de recel (article 321-1 du code pénal). Cela implique pour le responsable de traitement d'effectuer a minima les mêmes vérifications que celles énoncées dans la partie ci-dessus.

Le réutilisateur d'une base de données transmise de gré à gré par un tiers pourra d'autant moins ignorer qu'elle est constituée ou partagée en méconnaissance du RGPD ou de règles plus générales (telles que celles interdisant les atteintes à la sécurité des systèmes d'information ou des atteintes à des droits de propriété intellectuelle) que sa relation avec ce tiers lui permet de lever les doutes qu'il pourrait avoir.

La conclusion d'un accord entre le détenteur initial des données et le réutilisateur est ainsi recommandée afin de permettre à ce dernier de s'assurer de la licéité de ses propres traitements, quand bien même elle ne serait pas explicitement exigée par le RGPD.

À cet égard, la CNIL recommande de fournir un certain nombre d'indications dans le contrat telles que :

- la source, le contexte de la collecte des données, la base légale du traitement et l'analyse d'impact relative la protection des données (voir notamment la fiche n° 5 sur la réalisation d'une AIPD) si nécessaire, afin d'écarter les risques d'avoir une base de données illicite ;

- les mentions d'information des personnes portées à la connaissance des personnes (en particulier s'agissant de la finalité et des destinataires) ;

- d'éventuelles garanties quant à la licéité de ce partage de données par le détenteur initial des données (par exemple : sur la compatibilité de la finalité, sur la licéité du partage, etc.).

La CNIL fournit un modèle de **fiche descriptive** du jeu de données qui peut utilement être utilisé à cette fin.

À noter : si le réutilisateur souhaite fonder son traitement sur un consentement recueilli par un tiers, il doit être en mesure d'apporter la preuve qu'un consentement valide a bien été recueilli auprès des personnes concernées. L'obligation de rapporter la preuve du consentement ne peut pas être remplie par la seule présence d'une clause contractuelle engageant l'une des parties à recueillir un consentement valable pour le compte de l'autre partie. En effet, une telle clause ne permet pas à l'organisme de garantir, en toutes circonstances, l'existence d'un consentement valide (voir la délibération de la formation restreinte de la CNIL n° SAN-2023-009 du 15 juin 2023). Le contrat pourra, en revanche, être utilisé pour encadrer :

- les mécanismes mis en place pour permettre de démontrer le recueil d'un consentement valide ;

- la mise à disposition des éléments de preuve au profit de l'organisme qui souhaite se prévaloir du consentement ;

- le cas échéant, les conditions dans lesquelles ces éléments de preuve doivent être conservés, notamment afin de conserver leur valeur probante.

Exemple : le fournisseur d'un système d'IA générative d'image se rapproche d'un courtier en données pour constituer une base de données à des fins d'apprentissage comportant notamment des photographies.
Ils concluent pour cela un contrat qui garantit au fournisseur la licéité des données partagées, et encadre la fourniture d'indications cruciales pour la conformité de ses traitements (par exemple : preuves du contexte de la collecte des données pour apprécier son intérêt légitime, garanties s'agissant d'autres réglementations comme celle régissant la cession des droits de propriété intellectuelle, etc.).
Outre ces vérifications préalables, et quel que soit le mode de collecte utilisé, les réutilisateurs doivent s'assurer de la conformité complète de leurs propres traitements.

À noter que cette obligation vaut également lorsqu'ils réutilisent des bases de données dont la constitution et la diffusion ne relèvent pas du droit français ou européen. Pour plus d'informations sur le champ d'application territorial du RGPD, voir la fiche « Quel est le périmètre des fiches pratiques sur l'IA ».

En particulier, le réutilisateur doit veiller au respect des exigences vis-à-vis des personnes dont les données sont présentes dans la base ainsi obtenue : il doit les informer du traitement qu'il souhaite faire des données et leur permettre d'exercer leurs droits.

À noter : une fiche cas d'usage sur la réutilisation des données personnelles sera publiée ultérieurement. Elle permettra de compléter les éléments de mise en conformité introduits dans cette fiche notamment par l'étude de cas pratiques.

6 – IA : Réaliser une analyse d'impact si nécessaire

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.

6 – 1 - La réalisation d'une AIPD pour le développement de systèmes d'IA

6 – 1 – 1 - 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.

6 – 1 – 2 -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.

6 – 1 – 3- L'entraînement d'un système d'IA 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.

6 -1 -4 - 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.

6 – 2 - 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 explor

6 – 3 - 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.

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.

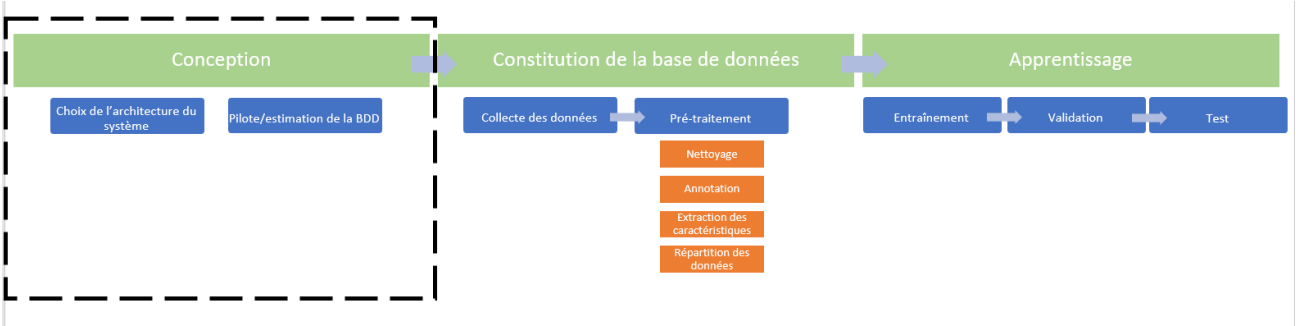
Enfin, la publication de tout ou partie de l'AIPD est recommandée, dans un objectif de transparence : si certaines parties de l'AIPD n'ont pas à être publiées dans la mesure où elles peuvent être couvertes par le secret des affaires ou donner des informations confidentielles sur la sécurité du système, 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.

7 - TENIR COMPTE DE LA PROTECTION DES DONNÉES DANS LA CONCEPTION DU SYSTÈME

Pour assurer le développement d'un système d'IA respectueux de la protection des données, il est nécessaire de mener une réflexion préalable lors de la conception du système. La CNIL en détaille les étapes.

Lors de la réflexion autour des choix de conception d'un système d'IA, les principes de protection des données, et en particulier le principe de minimisation, doivent être respectés. Cette démarche s'opère à cinq niveaux. Un responsable de traitement doit ainsi s'interroger sur :

- l'objectif du système qu'il souhaite développer ;
- la méthode à employer qui aura une incidence sur les caractéristiques de la base de données ;
- les sources de données mobilisées (voir la fiche sur la conformité du traitement à la loi, sur les sources ouvertes, sur les tiers, etc.) et ;
- parmi ces sources, la sélection des données strictement nécessaires, au regard de l'utilité des données et de l'impact potentiel que leur collecte fait peser sur les droits et libertés des personnes concernées ;
- la validité des choix précédemment opérés. Cette validation peut pendre différentes formes (non exclusives) telles que la réalisation d'une étude pilote ou l'avis d'un comité éthique.



7 – 1 - Objectifs du système

L'objectif de cette étape est de concevoir, sur la base de la finalité identifiée (voir Fiche n° 2), un système conforme à un cahier des charges, tout en limitant les conséquences potentielles pour les personnes concernées.

En précisant l'emploi du système d'IA en phase de déploiement (que celui-ci soit mis en œuvre directement par le fournisseur ou par un tiers), le fournisseur du système doit déterminer :

- le type de résultat / sortie attendu ;
- des indicateurs de la performance acceptable de la solution, qu'il s'agisse d'indicateurs quantitatifs (F1-score, erreur quadratique moyenne, temps de calcul) ou qualitatifs (provenant de retours humains par exemple) ;
- le contexte d'utilisation du système permettant d'identifier les informations prioritaires pour son usage opérationnel ;

les contextes d'utilisation exclus et les informations non pertinentes pour le ou les cas d'usage principaux envisagés du système.
Certaines techniques d'IA peuvent permettre de réaliser des tâches complexes qui dépassent les objectifs initiaux des fournisseurs. En définissant précisément la fonctionnalité attendue, il est ainsi possible d'éviter des risques de sur-collecte.

Exemple : pour réaliser l'apprentissage d'un système d'IA visant à permettre le comptage des personnes qui se tiennent debout dans un tramway à partir d'images de caméras de vidéoprotection, les systèmes suivants sont techniquement envisageables :

- un réseau de neurones détectant la présence de personnes dans un wagon, sans analyse de la posture, intégré dans un algorithme réalisant un décompte des personnes debout (le nombre de personnes debout pouvant être déduites grâce au nombre de places assises) ;

- un réseau de neurones réalisant une analyse de la posture des personnes dans un wagon intégré dans un algorithme réalisant un décompte des personnes qui se tiennent debout.

Le premier réseau pourrait fournir moins d'informations (notamment le décompte de personnes debout). Toutefois, si l'estimation donnée est suffisante pour le cas d'usage envisagé, en particulier pour le calcul de statistiques d'occupation, il est alors préférable

d’avoir recours à ce modèle. En effet, celui-ci nécessitera une quantité de données plus faible pour son apprentissage tout en permettant de remplir l’objectif poursuivi, alors que le second demande une collecte et une annotation de données spécifiques et de plus grande ampleur. Le principe de minimisation converge alors avec la réduction des coûts de conception du système, sans préjudice pour la précision du système.

7 -2 - La méthode à employer

Bien souvent, une même tâche peut être réalisée par différentes techniques. Toutefois, toutes ne sont pas équivalentes, car elles n’impliquent pas de recourir aux mêmes données et pas nécessairement les mêmes quantités. Elles peuvent ne pas permettre d’atteindre le même niveau de performances, présenter des enjeux plus ou moins importants en termes d’explicabilité, être soumises à différentes contraintes opérationnelles (comme le coût de calcul). Tout en prenant en compte ces enjeux, le fournisseur du système doit sélectionner la technique la plus respectueuse des droits et libertés des personnes afin de respecter le **principe de minimisation des données** en tenant compte de l’objectif recherché. Autrement dit, si une technique remplit la même fonction/permet d’atteindre le même résultat avec moins de données personnelles, elle doit être préférée.

En particulier, les méthodes d’apprentissage machine nécessitent l’utilisation d’un très grand nombre de données. Afin d’assurer le respect du principe de proportionnalité et de minimisation des données, le recours à ces solutions techniques doit donc être justifié. S’il existe une méthode n’utilisant pas l’apprentissage machine et que celle-ci permet de remplir les objectifs poursuivis, celle-ci doit être privilégiée. Le recours à l’apprentissage profond est à justifier et ne doit donc pas être systématique.

Exemples :
Pour assurer la sécurité des employés, un fournisseur veut délimiter une zone dangereuse à ne pas franchir dans son entrepôt. Il souhaite une solution qui permette de détecter la présence d’une personne dans cette zone et de déclencher un signal sonore pour avertir la personne du danger. Le recours à un détecteur de mouvement infrarouge doit être privilégié en lieu et place d’une caméra augmentée, car cette solution ne collecte pas l’image des personnes, conformément au principe de minimisation des données et de protection des données dès la conception.

L’analyse sémantique du contenu d’un texte pourrait être réalisée par un réseau neuronal entraîné sur une base de données textuelles annotées, par une méthode ensembliste tel qu’une forêt d’arbres décisionnels ou encore par un algorithme non supervisé, tel qu’un algorithme de partitionnement des données (« clustering »).
Au stade de l’entraînement du modèle, il y a aussi lieu de tenir compte de l’incertitude éventuelles sur les performances de telle ou telle architecture : le respect du principe de minimisation s’apprécie en fonction des connaissances scientifiques disponibles.

Selon les avancées dans le domaine concerné, cette réflexion doit reposer sur plusieurs facteurs pour chacune des architectures considérées. Cette analyse technique peut se faire grâce à :

- un état de l'art, au moyen, par exemple :
    - d'une étude de la littérature scientifique (recensement et étude des publications académiques ou privées, conférences spécialisées, etc.) ;
    - d'une enquête auprès des professionnels du domaine : la démarche d'ouverture du code informatique (y compris par placement sous licence libre) de certains acteurs du secteur contribue à rendre possible une comparaison des techniques ;
    - de la sollicitation de la communauté spécialisée (compétitions en ligne, forums en ligne, conférences et rencontres dédiées, etc.) ;
- une comparaison des résultats obtenus après l'implémentation de plusieurs architectures sous la forme de « preuves de concept » ;
- une comparaison des résultats obtenus par l'utilisation d'un modèle existant et pré-entraîné (pouvant nécessiter éventuellement un ajustement, ou fine-tuning) et d'un modèle développé par le fournisseur.

Si le choix du modèle d'IA et des algorithmes utilisés peuvent permettre de limiter la collecte de données, d'autres choix de conception sont à prendre en compte, notamment dans une optique de protection des données dès la conception. Le choix du protocole d'apprentissage utilisé notamment, peut permettre de limiter l'accès aux données aux seules personnes habilitées, ou encore de ne donner accès qu'à des données chiffrées. Deux techniques, applicables dans certaines situations, sont particulièrement intéressantes :

- Les protocoles d'apprentissage décentralisés, tels que l'apprentissage fédéré, technique à laquelle un article LINC a été dédié. Cette technique permet d'entraîner un modèle d'IA depuis plusieurs bases, et ainsi à chacun des acteurs de la chaîne de conserver la main sur ses données. Cette technique possède toutefois certains risques, concernant la sécurité des bases décentralisées, ainsi que sur la confiance entre les acteurs parmi lesquels un acteur malicieux pourrait conduire une attaque par empoisonnement des données par exemple.

- Les ressources offertes par la cryptographie. Les progrès scientifiques récents dans le domaine de la cryptographie peuvent permettre d'obtenir des garanties fortes pour la protection des données. En fonction des cas d'usage, il pourra par exemple être pertinent d'explorer les possibilités offertes par le calcul multipartite sécurisé (« secure multi-party computation »), ou encore le chiffrement homomorphe (« homomorphic encryption »). Les techniques utilisées dans ce domaine permettent d'entraîner un modèle d'IA sur des données qui restent chiffrées tout au long de l'apprentissage. Elles demeurent toutefois limitées en ce qu'elles ne peuvent pas être appliquées à tous types de modèles et en raison du temps de calcul supplémentaire qu'elles induisent. Par

ailleurs, certaines d'entre elles, comme le chiffrement homomorphe pour l'entraînement de réseaux de neurones, font encore l'objet d'études. Les évolutions techniques étant fréquentes dans ce domaine, il est conseillé de maintenir une veille active sur ce sujet.

Cette liste de mesures n'est pas exhaustive, des mesures supplémentaires pourraient être citées comme le recours à un environnement d'exécution sécurisé (ou « trusted execution environment »), à la confidentialité différentielle appliquée lors de l'apprentissage ou encore des mesures permettant d'anticiper le désapprentissage machine. Plus généralement, en raison de l'évolution rapide de la technique, il est recommandé de mener une veille technologique sur les pratiques protectrices de la vie privée applicables lors du développement de systèmes d'IA.

7 – 3 - La sélection des données strictement nécessaires

7 – 3 – 1 – Le principe

Le principe de minimisation prévoit que les données personnelles doivent être adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées. Une attention particulière doit être apportée à la nature des données et ce principe doit être appliqué de manière particulièrement rigoureuse lorsque les données traitées sont sensibles (au sens de l'article 9 du RGPD).

7 – 3 – 2 -En pratique

Le principe de minimisation ne signifie pas qu'il est interdit d'entraîner un algorithme avec des volumes très importants de données : il implique d'avoir une réflexion en amont de l'entraînement pour ne pas recourir à des données personnelles qui ne seraient pas utiles au développement du système. Pour identifier les données personnelles nécessaires au développement d'un système d'IA, il convient de prendre en compte quatre dimensions :

- Le volume : nombre de personnes concernées, profondeur historique, précision des données, répartition des données selon les situations et populations et couvrir, etc. Il pourra être justifié, par exemple, par les capacités de calcul limitées des serveurs utilisés pour l'apprentissage, les besoins en termes de représentativité du jeu de données, les pratiques communément admises par la communauté scientifique, une comparaison des résultats obtenus en faisant varier le volume de données, une analyse statistique démontrant qu'un volume minimum de données est nécessaire pour atteindre des résultats significatifs, etc. ;

- Les catégories : âge, sexe, image du visage, activité sur un réseau social, etc. La présence de données sensibles ou de données à caractère hautement personnel doit être examinée et justifiée (voir Fiche n° 3). Cette analyse peut reposer sur la nécessité d'entraîner le modèle sur des données contrefactuelles (susceptibles de donner lieu à des faux positifs en pratique), sur une étude de l'utilité des catégories de données concernées (voir encadré plus bas), etc. Parmi ces catégories de données, il convient de privilégier le format le moins intrusif sans

perte d'information pour l'objectif poursuivi, par exemple l'âge ou tranche d'âge plutôt qu'une date de naissance complète ;

- La typologie : données réelles, de synthèse, augmentées, issues de simulation, des données anonymisées ou pseudonymisées, etc. ;

- Les sources : comme précisé dans la Fiche n° 3, recensement des sources de données auxquelles il est envisagé de recourir, qu'il s'agisse d'une collecte initiale ou d'une réutilisation (données disponibles en source ouverte, collectées précédemment par le fournisseur ou encore auprès de fournisseurs de données).

Bien que la sélection des données soit une phase généralement nécessaire afin de concevoir un système d'IA sur la base de données de qualité, dans certains cas et à titre subsidiaire, il peut être possible de traiter un ensemble de données de manière indiscriminée. La nécessité devra alors en être justifiée.

Outre la prise en compte de ces dimensions d'ordre technique, une attention particulière devra être portée à la nature des données au sens du RGPD, et plus particulièrement s'il s'agit de données sensibles ou hautement personnelles.

À noter :
Les questions relatives à la distribution et à la représentativité des données doivent également être traitées lors de cette étape. Celles-ci sont en effet essentielles pour limiter au maximum les risques de biais de discrimination.

Au sein de cette question se pose notamment celle de l'inclusion de données de type « vrais négatifs » dans la base de données d'apprentissage (notamment pour le test et la validation en vue de vérifier l'absence de certains effets de bord ou d'apprentissage).

7 – 3 – 3 - La validité des choix de conception

À l'issue des trois étapes précédentes, les choix de conception sont théoriquement validés et la collecte des données peut commencer. Afin de valider quantitativement et qualitativement les choix de conceptions, plusieurs mesures sont recommandées à titre de bonne pratique.

Mener une étude pilote

L'objectif du pilote est de s'assurer que les choix d'ordre technique et ceux relatifs aux types de données identifiés sont bien pertinents. Pour ce faire, une expérimentation à petite échelle peut est réalisée. Des données fictives, synthétiques, anonymisées ou à défaut des données personnelles collectées conformément au RGPD peuvent être utilisées.

Exemples :

L'utilisation de données issues de réseaux sociaux sur les pages personnelles de personnes ayant consenti à la collecte de leurs données.
Ce type d'expérimentation n'offre pas toujours une vision représentative de l'activité rencontrée sur les réseaux sociaux, mais elle peut être adaptée à certains cas d'usage comme l'identification de contenus haineux ou l'étude du ciblage publicitaire sur ces réseaux. Cette pratique est bénéfique car elle offre un niveau de transparence largement supérieur à certaines pratiques de moissonnage (« *web scraping* »).

La conception d'un système de recommandation de films
Un organisme peut collecter auprès d'utilisateurs volontaires la liste de films visionnés sur une semaine et ceux visionnés dans les jours qui ont suivi, soit par données déclaratives, soit en collectant leur historique de visionnage sur des sites dédiés. Il peut mener son étude pilote sur les données ainsi collectées en anonymisant les identifiants de chaque utilisateur.

Interroger un comité éthique

L'association d'un comité éthique au développement de systèmes d'IA est une bonne pratique pour garantir que les enjeux en matière d'éthique et de protection des droits et libertés des personnes soient pris en compte en amont.

Le comité éthique peut avoir plusieurs missions :

- **la formulation d'avis** sur tout ou partie des projets, outils, produits, etc. de l'organisme susceptibles de problématiques éthiques, qui lui seraient soumis ;

- **l'animation d'une réflexion** et l'élaboration d'une doctrine interne sur les aspects éthiques du développement de systèmes d'IA par l'organisme (par exemple, quelles conditions pour le recours à la sous-traitance) ;

- **la mise au jour d'attitudes collectives** et individuelles et la recommandation de certains principes, comportements ou pratiques.

La constitution et le rôle de ce comité peuvent varier selon les situations, mais plusieurs bonnes pratiques sont recommandées. Le comité devrait :

- **être pluridisciplinaire** : les profils des membres du comité – employés de l'organisme et/ou personnes externes – doivent être diversifiés. Les personnes amenées à y siéger contribuent aux missions du comité et peuvent mettre à jour des problématiques que les équipes de développement n'avaient pas envisagées. Une bonne pratique est d'attribuer certains sièges du comité aux employés de l'organisme qui l'occuperont chacun leur tour. En outre, la diversité des membres du comité du point de vue du sexe, de l'âge et des origines ethniques et culturelles est vivement encouragée ;

- **être indépendant** : les avis rendus par le comité peuvent avoir des implications importantes, par exemple pour la direction commerciale d'une entreprise et ainsi favoriser ou défavoriser certains de ses projets. Ainsi, les personnes siégeant au comité ne doivent pas être motivées par un éventuel gain (qu'il soit financier ou d'un autre ordre) à tirer par la décision rendue. De même, lorsque des employés siègent au comité, les décisions rendues ne doivent pas avoir de conséquences pour eux ;

- **avoir un rôle clairement défini** : afin de garantir l'intégration systématique du comité, une procédure doit être fixée afin de déterminer les conditions dans lesquelles le comité se réunit et doit être associé. Selon les situations, le comité peut être simplement consultatif ou adopter des avis contraignants : les deux approches présentent des avantages et des inconvénients. Si le comité émet des avis contraignants, son insertion dans la gouvernance d'entreprise doit être particulièrement bien définie au regard des statuts de l'organisme, pour éviter son instrumentalisation. Si le comité est consultatif, son impact doit être garanti, notamment en garantissant sa saisine obligatoire selon des critères précis et la transparence large de ses avis, a minima au sein de l'organisme et éventuellement d'autres mesures comme l'obligation pour le porteur de projet de répondre par écrit aux remarques du comité ;

- **être averti** : le comité est encouragé à s'informer, documenter ses avis et partager ses connaissances. Les risques que comporte l'utilisation de l'IA évoluent avec le développement technique et les nouveaux usages dans ce domaine, et il est nécessaire de s'informer, notamment grâce à la littérature académique et aux publications des entités compétentes dans ce domaine (comme le Défenseur des Droits, ou le Comité national pilote d'éthique du numérique). La diffusion des connaissances acquises permettra d'appuyer les avis rendus et de répandre certaines bonnes pratiques.

Dans le cas du développement d'un système d'IA, l'avis du comité éthique pourrait être sollicité sur plusieurs questions :

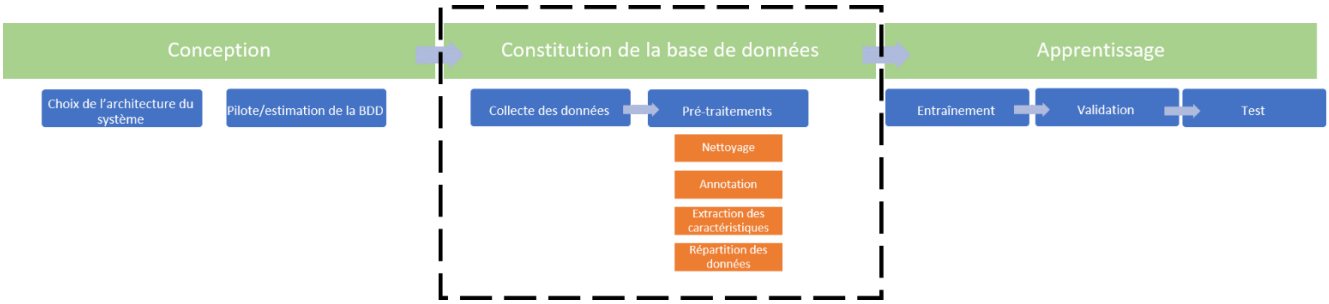
- Les données utilisées pour le développement respectent-elles les critères éthiques de l'organisme ?
- Les usages opérationnels prévus pour le système d'IA pourraient-ils entraîner des conséquences individuelles ou sociétales graves ? Ces conséquences peuvent-elles être évitées ? Ces usages opérationnels peuvent-ils être exclus ?
- Les potentielles utilisations détournées du système d'IA (qu'elles soient volontaires ou accidentelles, en particulier pour les modèles diffusés en source ouverte) pourraient-elles entraîner des conséquences graves pour les personnes ou pour la société ? Quelles mesures permettraient de les éviter ?
- Les choix techniques sont-ils suffisamment maîtrisés par l'organisme (dans le cas du recours à des approches radicalement nouvelles) ?

- Les mesures de transparence, pour l'exercice des droits des personnes ou leur permettant d'exercer un éventuel recours sont-elles suffisantes ?
- Les discriminations que peuvent entraîner l'utilisation du système sont-elles répertoriées et les moyens nécessaires ont-ils été mis en œuvre afin d'éviter leur survenue ?
- L'organisme est-il organisé de manière à prévenir les risques dès la conception (que ce soit en matière de discriminations, de protection des données, de protection du droit d'auteur, de sécurité informatique, etc.) ?

En fonction de la taille des organismes et de la façon dont ils sont structurés, il n'est pas toujours envisageable de constituer un comité éthique. Néanmoins, il est essentiel que de telles réflexions puissent être menées pour accompagner le développement de systèmes d'IA. La nomination d'un « référent éthique » peut être une alternative permettant la prise en compte de ces questionnements.

8 -TENIR COMPTE DE LA PROTECTION DES DONNÉES DANS LA COLLECTE ET LA GESTION DES DONNÉES

La CNIL donne les bonnes pratiques pour sélectionner les données et limiter leur traitement afin d'entraîner un modèle performant dans le respect des principes de protection des données dès la conception et par défaut.
Une fois les données et leurs sources identifiées, le fournisseur du système d'IA doit mettre en œuvre la collecte et constituer sa base de données. Pour cela il est nécessaire d'intégrer dès leur conception les principes de protection des données personnelles (« privacy by design »



8 - 1 – Collecte

La collecte des données s'accompagne de différentes vérifications et démarches en fonction des modalités et sources de données. Techniquement, il s'agit de s'assurer que les données collectées sont pertinentes compte tenu des objectifs poursuivis, et ainsi d'assurer le respect du principe de minimisation.

Collecte de données par moissonnage (« web scraping »)

Quand le responsable du traitement réutilise des données publiquement accessibles qu'il a lui-même extraites de sites web au moyen d'outils de moissonnage (« web scraping »),

il doit particulièrement s'assurer de minimiser la collecte de données, en tâchant notamment de :

limiter la collecte aux données librement accessibles ;
définir, en amont de la mise en œuvre du traitement, des critères précis de collecte ;
s'assurer de ne collecter que des données pertinentes et de la suppression des données immédiatement après leur collecte ou dès qu'elles sont identifiées comme telles (quand le tri exhaustif n'est pas possible lors de la collecte).
Pour plus d'informations au sujet de la collecte des données publiquement accessibles : voir le projet de guide sur l'ouverture, le partage et la réutilisation des données

8 -2 - Nettoyage, identification et protection de la vie privée dès la conception

8 – 2 – 1 -Nettoyage

Le nettoyage des données permet de constituer une base d'entraînement de qualité. C'est une étape cruciale qui renforce l'intégrité et la pertinence des données en réduisant les incohérences, ainsi que le coût de l'entraînement. Concrètement, il s'agit ainsi de :

- corriger les valeurs vides ;
- détecter les valeurs aberrantes ;
- corriger les erreurs ;
- éliminer les doublons ;
- supprimer les champs inutiles ;
- etc.

8 - 2 -2 - Identification des données pertinentes

La sélection des données et des caractéristiques pertinentes est une procédure classique en IA. Elle vise à optimiser les performances du système tout en évitant les sous- et sur-apprentissage. En pratique, elle permet ainsi de s'assurer que certaines classes inutiles pour la tâche visée ne sont pas représentées, que les proportions entre les différentes classes d'intérêt sont bien équilibrées, etc. Cette procédure vise également à identifier les données non pertinentes pour l'apprentissage. Les données identifiées comme non pertinentes devront alors être supprimées de la base.

En pratique, cette sélection peut trouver à s'appliquer sur trois types d'objets constituant la base de données :

- Les données : il peut s'agir de données « brutes », non-structurées, (extrait audio, image, fichier texte manuscrit, etc.) ou structurées (mesures, observations, etc. au format numérique) ;
- Les métadonnées associées : littéralement « données sur les données », les métadonnées, fournissent des informations de description (quel a été le processus d'acquisition ? par qui a-t-il été réalisé ? à quelle date ? etc.), de structure (comment faut-il les exploiter ?) ou encore de qualité ;

- Les annotations et les caractéristiques extraites de données (« features ») : descriptions attribuées aux données dans le cas des annotations, ou propriétés mesurables extraites à partir des données pour les caractéristiques (informations relatives à la forme ou la texture d'une image, à la hauteur des sons, au timbre ou au tempo d'un fichier audio, etc.).

Plusieurs approches peuvent concourir à mettre en œuvre cette sélection. Citons à titre illustratif :

- L'utilisation de techniques et outils permettant d'identifier les caractéristiques pertinentes (sélection de caractéristiques ou « feature selection »), parfois en amont de l'entraînement. Des analyses de type Analyse en composantes principales (PCA - « Principal Component Analysis »), peuvent également aider à identifier les caractéristiques fortement corrélées d'un jeu de données et ainsi ne conserver que celles qui sont pertinentes. De nombreuses bibliothèques telles que Yellowbrick, Leave One Feature Out (LOFO) ou encore Facets proposent aujourd'hui des implémentations pour la sélection de caractéristiques.

- L'utilisation d'approches d'annotations interactives de données comme l'apprentissage actif (« active learning »), qui permettent une revue des données par l'utilisateur sur la base de la tâche à accomplir et, le cas échéant, la suppression de celles qui sont non-pertinentes. La bibliothèque Scikit-ActiveML en est un exemple.

- L'utilisation de techniques d'ablation des données d'entraînement (« data/dataset pruning ») : cette technique, abordée dans plusieurs publications comme Sorscher et al., 2022 ou Yang et al., 2023, permet de réduire le temps de calcul nécessaire à l'entraînement sans impact significatif sur les performances du modèle obtenu, tout en identifiant les données peu utiles à l'entraînement.

Enfin, dans certains cas spécifiques et pour lesquels la conservation des données pourra s'avérer complexe ou problématique (en raison de la sensibilité des données, de questions liées à la propriété intellectuelle, etc.), le principe de minimisation peut être mis en œuvre par la conservation exclusive des caractéristiques extraites et par la suppression des données sources dont elles sont issues.

Exemple : Pour une étude de la propagation de propos haineux dans les réseaux sociaux, l'analyse des commentaires associés à une publication permet de réaliser une classification des réactions des utilisateurs, mais le contenu des commentaires lui-même pourrait être supprimé après cette analyse.

La constitution d'une base de données d'apprentissage pour l'IA nécessite également souvent l'annotations des données. La production et l'utilisation de celles-ci doivent également faire l'objet de mesures particulières au titre de la protection des données. Celles-ci seront détaillées dans une fiche pratique dédiée.

8 – 3 – 3 -Protection des données dès la conception (« privacy by design »)

Par ailleurs, outre ces étapes nécessaires, le fournisseur du système d'IA doit mettre en œuvre une série de mesures pour intégrer dès leur conception les principes de protection des données personnelles (« privacy by design »).

Elles doivent tenir compte de l'état des connaissances, de leur incidence sur l'efficacité de l'entraînement, des coûts de mise en œuvre et de la nature, de la portée, du contexte et des finalités du traitement ainsi que des risques (dont la vraisemblance et la gravité varient) que présente le traitement pour les droits et libertés des personnes. Ces mesures peuvent inclure :

- **Des mesures de généralisation** : ces mesures visent à généraliser, ou diluer, les attributs des personnes concernées en modifiant leur échelle ou leur ordre de grandeur respectif ;
- **Des mesures de randomisation** : ces mesures visent à ajouter du bruit aux données afin d'en diminuer la précision et d'affaiblir le lien entre les données et l'individu.

Ces mesures sont à mettre en œuvre sur les données ainsi que les métadonnées qui y sont associées.

Dans certains cas, ces mesures peuvent aller jusqu'à l'anonymisation des données, et notamment si l'objectif n'impose pas de traiter des données personnelles : si les traitements de sélection et gestion des données sont des traitements de données personnelles soumis au RGPD et donc aux présentes fiches, les traitements ultérieurs ne seront plus concernés par la réglementation sur la protection des données personnelles.

Exemple : Un organisme souhaite constituer une base de données de code informatique relatif à des machines industrielles (SCADA) issues de dépôts de plusieurs développeurs. Après avoir retiré toute mention relative aux développeurs eux-mêmes, puis avoir vérifié l'absence d'identifiants ou de mentions personnelles dans les commentaires, la base de données ne contient aucune donnée personnelle. Elle n'est plus soumise à la réglementation sur la protection des données.
De plus, certaines mesures permettent de protéger les données lors de l'apprentissage du système d'IA, comme la confidentialité différentielle appliquée durant l'apprentissage du modèle ou l'apprentissage fédéré. Bien que certaines de ces techniques soient encore au stade de la recherche, des outils permettent de les mettre en œuvre afin de tester leur efficacité, comme PyDP ou encore OpenDP.

Les mesures portant sur les données

Les mesures applicables dépendent des catégories de données concernées et doivent être considérées au regard de leur influence sur les performances techniques – théoriques et opérationnelles – du système. L'impact de ces mesures est particulièrement bénéfique en raison :

- d’une part, de leur capacité à réduire les conséquences d’une éventuelle perte de confidentialité des données (par compromission des données contenues dans la base, ou par une attaque portant sur le modèle entraîné tel qu’une attaque par inférence d’attribut) ;
- d’autre part, de la possibilité éventuelle d’utiliser le modèle entraîné en phase opérationnelle sur des données ayant fait l’objet de mesures de protection identiques, offrant ainsi la capacité de mieux les protéger en phase opérationnelle.

Exemple : En généralisant les informations relatives à l’âge de patients dans le cadre du développement d’un système d’IA d’aide au diagnostic, aux champs [mois-année] ou [année] à la place de [jour-mois-année], le fournisseur réduit drastiquement les risque de perte de confidentialité, cela sans préjudice sur la capacité de généralisation de son système.

Les mesures portant sur les métadonnées

Les métadonnées peuvent contenir des informations utiles à un attaquant qui cherche à réidentifier les personnes concernées (comme une date ou un lieu de collecte des données). Le principe de minimisation s’applique également à ces données, et elles devraient ainsi être limitées à ce qui est nécessaire.

Les métadonnées peuvent par exemple être nécessaires au fournisseur pour donner suite à une demande d’exercice des droits, puisqu’elles permettent parfois d’identifier les données se rapportant à une personne. Dans ce cas, une attention particulière devrait être portée à leur sécurité.

Toutefois, si le traitement des métadonnées n’est pas nécessaire et que celles-ci contiennent des données à caractère personnel, leur suppression peut être recommandée dans un objectif de pseudonymisation ou d’anonymisation du jeu de données.

Par exemple : dans le cas où il réutiliserait des images de vidéoprotection pour constituer un jeu de données d’apprentissage, un fournisseur qui généraliserait le lieu de collecte d’une image d’une adresse à une maille IRIS pourrait ne plus être en mesure de donner suite à une demande d’accès aux données.

8 – 3 – 4 - Suivi et mise à jour

Bien que des mesures de minimisation et de protection des données aient été mises en œuvre lors de la collecte des données, ces mesures pourraient devenir obsolètes au cours du temps. En effet, les données collectées pourraient perdre leurs caractères exact, pertinent, adéquat et limité, en raison notamment :

- d’une possible dérive des données en conditions réelles, c'est-à-dire d'un écart entre la distribution des données d'entraînement et la distribution des données en condition d'utilisation. La dérive des données peut avoir de multiples causes :
  - des modifications de processus en amont, comme le remplacement d’un capteur, dont l’étalonnage diffère légèrement de celui qui était précédemment installé ;

- des problèmes de qualité des données, par exemple un capteur cassé qui indiquerait toujours une valeur nulle ;
- l'apparition d'une nouvelle catégorie dans un problème de classification ;
- la dérive naturelle des données, comme la variation de la température moyenne au fil des saisons ;
- la dérive due à de soudains changements, comme la perte de la capacité d'un système à détecter des visages suite au port massif de masques lors de l'épidémie de Covid-19 ;
- la modification de la relation entre caractéristiques ;

un empoisonnement malveillant dans le cadre d'un apprentissage continu, par exemple constaté par des résultats non souhaités.

Des outils existent pour détecter l'apparition d'une dérive des données, tel que Evidently, ou bien la bibliothèque Scipy dont les fonctions de tests statistiques peuvent être utilisés dans cet objectif ;

- d'une mise à jour des données, tel qu'une correction du lieu d'habitation dans le profil public de l'utilisateur d'un réseau social à la suite d'un déménagement ;
- de l'évolution des techniques, qui démontre fréquemment qu'un changement d'approche (utilisation d'un système d'IA différent nécessitant une typologie de données différente, par exemple) peut apporter de meilleures performances au système, ou encore que des performances similaires peuvent être obtenues avec un volume de données moins important (comme l'a montré la technique du « few-shot learning », par exemple).

Ainsi, le fournisseur du système devrait conduire une analyse régulière pour assurer le suivi de la base de données constituée. Cette analyse sera plus poussée et plus fréquente dans les situations où les causes évoquées ci-dessus sont les plus à même d'avoir lieu. Cette analyse devrait reposer sur :

- une comparaison régulière des données ou d'un échantillon de données aux données sources, celle-ci pouvant être automatisée ;
- une revue régulière des données par des agents formés aux questions relatives à la protection des données, ou par un comité d'éthique, en charge de vérifier notamment que les données sont toujours pertinentes et adéquates pour la finalité du traitement ;
- une veille portant sur la littérature scientifique dans le domaine et permettant d'identifier l'apparition de nouvelles techniques plus frugales en données.

8 – 3 – 5 - Conservation des données

Le principe

Les données personnelles ne peuvent être conservées indéfiniment. Le RGPD impose de définir une durée au bout de laquelle les données doivent être supprimées, ou dans certains cas archivées. Cette durée de conservation doit être déterminée par le responsable de traitement en fonction de l'objectif ayant conduit à la collecte de ces données.

En pratique

Le fournisseur doit fixer une durée de conservation des données utilisées pour le développement du système d'IA, conformément au principe de limitation de la conservation des données (article 5.1.d du RGPD).

La fixation d'une durée de conservation impose notamment la mise en œuvre de certaines procédures décrites dans le guide pratique de la CNIL sur les durées de conservation. La CNIL constate que les bases de données publiées en source ouverte évoluent constamment (par amélioration de l'annotation, ajout de nouvelles données, purge des données de mauvaise qualité, etc.) : une durée de conservation de plusieurs années à partir de la date de la collecte devra être justifiée.

Fixer une durée de conservation pour la phase de développement

Tout d'abord, le fournisseur du système d'IA devra fixer une durée de conservation des données pour l'usage fait pour le développement du système. Durant cette phase, le fournisseur utilise les données pour :

- la constitution de la base de données limitée à celles strictement nécessaires, nettoyées, prétraitées et prêtes à être utilisées pour l'apprentissage ;
- l'apprentissage de sa solution, depuis le premier entraînement du modèle d'IA jusqu'à la phase de test permettant de déterminer les caractéristiques et performances du produit fini. Lors de cette phase, les données doivent être conservées de manière sécurisée et être accessibles aux personnes habilitées. Selon les cas, cette phase peut durer de quelques semaines à plusieurs mois, ou au contraire se faire de manière itérative dans le cas de l'apprentissage en continu. Cette durée devrait être définie en amont et justifiée (en tenant compte de l'expérience passée du responsable du traitement, de ses connaissances sur la durée des développements informatiques, des ressources humaines et matérielles qu'il peut mettre à disposition pour les réaliser, etc.).

La conservation des données doit faire l'objet d'une planification en amont et d'un suivi dans le temps. Les durées de conservation définies doivent par ailleurs être appliquées aux données concernées, quel que soit leur support. Le respect des durées de conservation peut parfois être facilité par l'utilisation d'outils de gestion et de gouvernance permettant de définir une durée de conservation de chaque donnée et calculant la durée écoulée depuis la date d'entrée dans la base avant de les supprimer automatiquement. Une attention particulière doit ainsi être portée à la traçabilité des données éventuellement extraites de la base principale et sauvegardées sur des supports tiers, par exemple pour permettre l'analyse d'un échantillon au cas par cas par les ingénieurs. Les mesures recommandées dans la section « Documentation » concernant la traçabilité des données pourront faciliter le suivi des données et de la date prévue pour leur suppression.

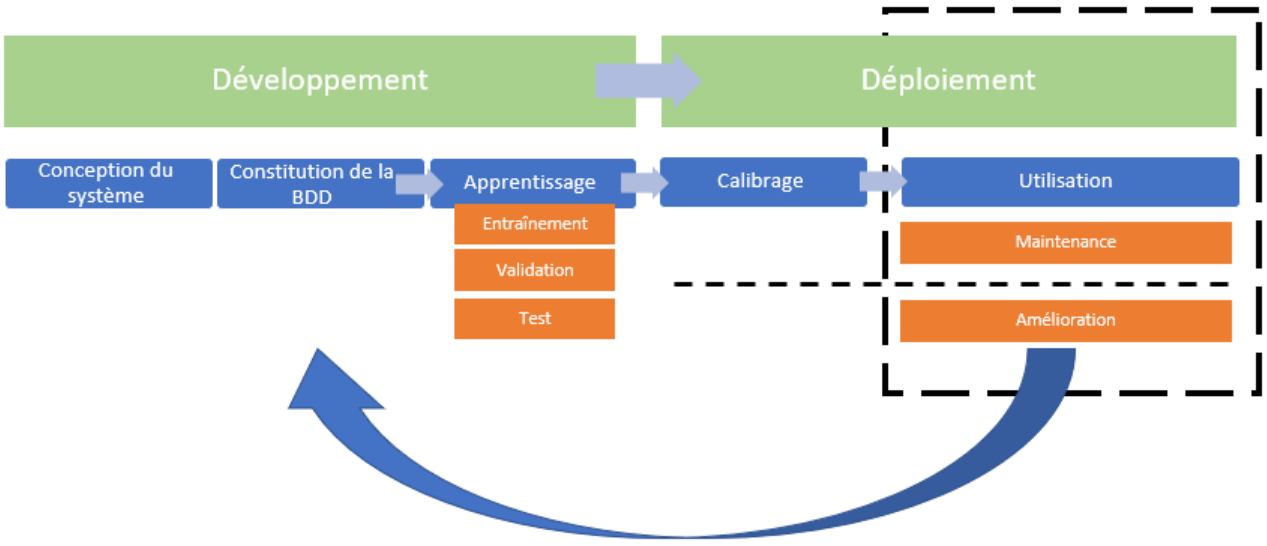
À noter :
Concernant les organismes publics ou les organismes de droit privé chargés d'une mission de service public, les données peuvent également devoir faire l'objet d'un archivage spécifique dans le respect des obligations du code du patrimoine.

Les données peuvent ainsi être versées en archivage définitif dans un service public d'archives selon l'intérêt particulier qu'elles présentent. Lorsque les archives publiques comportent des données personnelles, une sélection est réalisée pour déterminer les données destinées à être conservées et celles, dépourvues d'utilité administrative ou d'intérêt scientifique, statistique ou historique, destinées à être éliminées.

En tout état de cause, les données conservées dans le cadre de l'archivage définitif relèvent d'un traitement à finalité archivistique au sens du RGPD et, dès lors, n'entrent pas dans le cadre des présentes fiches. Par ailleurs, la durée de conservation des données doit être précisée dans les mentions d'information qui seront portées à la connaissance des personnes concernées.

Fixer une durée pour la maintenance ou l'amélioration du produit

Lorsque les données n'ont plus à être accessibles pour les tâches quotidiennes des personnes en charge du développement du système d'IA, elles doivent en principe être supprimées. Elles peuvent toutefois être conservées pour la maintenance du produit (c'est-à-dire pour une phase ultérieure de vérification des performances) ou encore à des fins d'amélioration du système).



Les opérations de maintenance

Le principe de minimisation des données impose de ne conserver que les données strictement nécessaires aux opérations de maintenance (en sélectionnant les données pertinentes, en réalisant une pseudonymisation des données lorsque c'est possible, comme en floutant des images par exemple, etc.).

Ces opérations permettent de garantir la sécurité des personnes concernées par l'utilisation du modèle en phase de déploiement, comme lorsque le système produit un effet sur les personnes, lorsqu'une baisse de performance pourrait entraîner des conséquences graves pour les personnes, ou encore lorsqu'il concerne la sécurité d'un produit. Alors, **la conservation des données d'apprentissage peut permettre d'effectuer des audits, et faciliter la mesure de certains biais**. Dans ces cas, et lorsqu'un résultat similaire ne pourrait être atteint par la conservation d'informations générales sur les données (telles que la documentation réalisée sur le modèle proposé dans la section Documentation, ou encore des informations sur la distribution statistique des données), une conservation des données prolongée peut être justifiée. Cette conservation doit toutefois être limitée aux données nécessaires, et s'accompagner de mesures de sécurité renforcées.

Une fois les données triées, elles peuvent être stockées sur un support cloisonné, c'est-à-dire séparé physiquement ou logiquement des données constitutives de la base. Ce cloisonnement permet de renforcer la sécurité des données et de restreindre leur accès aux seules personnes habilitées. La durée de la phase de maintenance peut varier de quelques mois à plusieurs années lorsque la conservation de ces données emporte peu de risques pour les personnes et que les mesures adaptées ont été prises. Dans le cas de données provenant de sources ouvertes, la durée de conservation prévue par la source des données doit être prise en compte dans la détermination de la durée de la phase de maintenance. Cette durée doit toutefois être limitée, et justifiée par un besoin réel.

L'amélioration du système d'IA

Les données constitutives de la base constituée précédemment peuvent être également nécessaires pour améliorer le produit issu du système d'IA ainsi développé. Cette finalité, pour laquelle une base légale devra être identifiée, devra être portée à la connaissance des personnes concernées, conformément au principe de transparence. Concrètement, seules les données nécessaires à l'amélioration du système d'IA peuvent être extraites de leur espace de stockage cloisonné.

> À noter :
> La possibilité de prolonger le cycle par une nouvelle phase de développement ou de maintenance ne pourra, en aucun cas, permettre une prolongation indéfinie de la durée de conservation, une analyse de la durée nécessaire aux opérations de traitement devra être conduite systématiquement.

8 – 3 - 6 – Sécurité

Le principe

Le responsable du traitement et ses sous-traitants (s'il en a) doivent mettent en œuvre les mesures techniques et organisationnelles appropriées afin de garantir un niveau de sécurité adapté au risque (article 32 du RGPD).

Le choix des mesures à mettre en œuvre doit tenir compte de l'état des connaissances, des coûts de mise en œuvre et de la nature, de la portée, du contexte et des finalités du traitement ainsi que des risques, dont les degrés de vraisemblance et de gravité varient, pour les droits et libertés des personnes concernées.

En pratique

Ainsi, le fournisseur d'un système d'IA doit en particulier prévoir les mesures adaptées afin de sécuriser :

- les techniques de collecte des données employées, au moyen par exemple de méthodes de chiffrement des flux et de méthodes d'authentification robustes permettant de restreindre l'accès au système d'information. Il est recommandé d'utiliser les moyens prévus par le diffuseur pour collecter les données, notamment lorsque ceux-ci reposent sur des API. La recommandation de la CNIL sur l'utilisation d'API devra alors être appliquée ;
- les données collectées, au moyen de méthodes de chiffrement des sauvegardes, de vérification de leur intégrité, ou encore de journalisation des opérations réalisées sur la base de données conformes à la recommandation de la CNIL relative aux mesures de journalisation. Un risque fréquent dans le développement de système d'IA concerne la duplication des données, celles-ci ayant fréquemment à être analysées pour vérifier leur qualité. Les duplications de données devraient être limitées dans la mesure du possible et tracées lorsqu'elles sont inévitables. Des outils dédiés, comme NB Defense, Octopii, ou encore PiiCatcher, ou des techniques telles que la recherche par expressions régulières, ou la reconnaissance d'entités nommées pour les données textuelles, permettent de vérifier la présence de données personnelles dans certains ontextes ;
- le système d'information utilisé pour le développement du système d'IA, au moyen, par exemple, de méthodes d'authentification et de la formation des agents ayant à y accéder, et de la mise en œuvre des bonnes pratiques d'hygiène informatique ;
- le matériel informatique, notamment au moyen de méthodes de restriction d'accès aux locaux et par l'analyse des garanties apportées par l'hébergeur de données lorsque cela est sous-traité à un prestataire.

Les mesures de sécurité spécifiques aux phases de développement et déploiement de systèmes d'IA seront l'objet d'une Fiche ultérieure. Toutefois, les recommandations et bonnes pratiques classiquement mises en œuvre en informatique, telles que celles présentes sur le site de la CNIL, ainsi que les guides RGPD de l'équipe de développement et de la sécurité des données personnelles, constituent un socle de référence utile auquel le fournisseur du système d'IA pourra se référer.

8 – 3 – 7 – Documentation

La documentation des données utilisées pour le développement d'un système d'IA permet de garantir la traçabilité des jeux de données utilisés dont la grande taille rend généralement cette tâche difficile. Elle doit permettre de :

- faciliter l'utilisation de la base de données ;
- démontrer que les données ont été collectées de manière licite ;
- faciliter le suivi des données dans le temps jusqu'à leur suppression ou leur anonymisation ;
- réduire les risques d'une utilisation imprévue des données ;
- permettre l'exercice des droits pour les personnes concernées ;
- identifier les améliorations prévues ou envisageables.

Afin de répondre à ces objectifs, un modèle de documentation pourra être adopté, notamment dans le cas où le fournisseur a recours à de multiples sources de données ou constitue plusieurs bases de données. En s'appuyant sur les modèles existants (tel que ceux proposés par Gebru et al., 2021, Arnold et al., 2019, Bender et al., 2018, le Dataset Nutrition Label, ou encore la documentation technique prévue en annexe IV du projet de règlement européen sur l'intelligence artificielle), **la CNIL fournit ci-après un modèle** qui pourra être utilisé à cet effet, notamment dans le cas où la base de données constituée a vocation à être diffusée. Cette documentation devrait être réalisée par jeu de données lorsque ceux-ci sont constitués, mis à disposition, ou qu'ils proviennent d'un jeu de données existant auquel une modification substantielle a été apportée. Des modèles de documentation plus spécifiques à chacun des cas d'usage, comme le modèle CrowdWorkSheets, particulièrement pertinent pour documenter la phase d'annotation, pourront compléter le modèle proposé.

Les objectifs de cette documentation sont de nourrir la réflexion interne du responsable de traitement sur ses pratiques, d'informer les utilisateurs du jeu de données sur les conditions de sa constitution et des recommandations concernant son traitement, et enfin, d'informer les personnes dans un but de transparence. Ainsi, il est recommandé de fournir cette documentation aux utilisateurs du jeu de données ou des modèles qu'il a servi à concevoir.

Annexe

:***Fiche descriptive du jeu de données***

Référence du jeu :

1. ***Synthèse***

Description du jeu de données :
Durée de conservation prévue :
Restrictions d'accès :
Restrictions d'utilisation :
Présence de données personnelles, hautement personnelles, sensibles ou protégées par d'autres dispositions :
Version de la fiche descriptive et dernière mise à jour :

2. ***Contexte et motivation***

Identité de l'organisme ayant constitué la base de données

- *Nom de l'organisme :*
- *Statut :*
- *Adresse de contact :*
- *Liens avec d'autres organismes :*

Motivation pour la constitution de la base de données

- *Finalité :*
- *Problématique motivant la collecte :*
- *Résultat attendu à la suite du traitement de la base (tâche ou fonctionnalité du système) :*
- *Plus-value au regard des bases existantes :*

Contexte de la collecte

Source des données	*Méthode de collecte employée*	*Personnes concernées*	*Profondeur historique ou période concernée*
Source 1			
Source 2			
…			

Destination du jeu

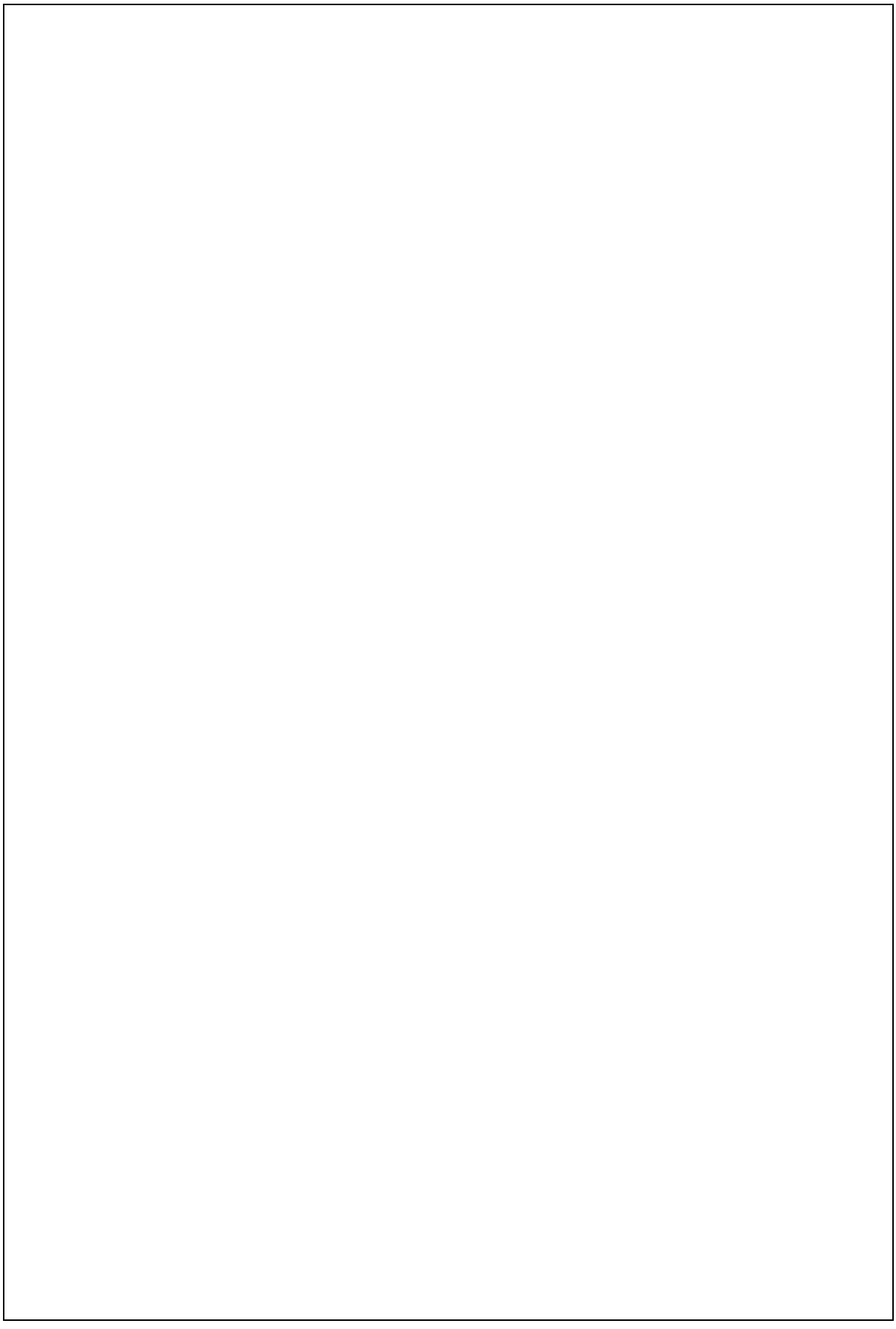
- *Domaine scientifique concerné :*
- *Entité fonctionnelle interne concernée (R&D, marketing, RH, production, etc.) :*
- *Catégories et nombre estimé de réutilisateurs internes et externes (entreprises, chercheurs, particuliers, etc.) :*

3. ***Composition de la base de données***

Description de la typologie des données (données tabulaires, images, séries temporelles, enregistrements vidéo ou sonores, etc.) :
Description des catégories de données (variables, classes, etc.) :
Description des métadonnées :
Description des liens, interconnexions et recoupements entre les données (tel qu'un identifiant liant les données relatives à une personne) :

Volume de données

- *Quantité de données ○ Totale : ○ Par classe :*
  - *Par instance :*
- *Pas de temps (séries temporelles) :*
- *Fréquence d'actualisation :*



Représentativité du jeu de données

- Hypothèses relatives à la conception du jeu (tel que les variables de substitution ou *proxy* utilisés) :
- Catégories de données absentes ou exclues du jeu :
- Catégories (d'objets, de situations, de personnes, etc.) pour lesquels la représentativité du jeu a été testée :
- Indicateurs concernant la représentativité (tels que la distribution statistique des catégories de données) :
- Contextes auxquels la distribution statistique du jeu sera extrapolée :
- Biais connus ou envisagés :
- Techniques de mesure et de compensation des biais employées ou recommandées ○ Sur le jeu de données : ○ Sur l'entraînement des modèles :
    - ○ Sur les sorties des modèles :

Qualité des données

- Erreurs connues ou envisagées dans les données :
- Sources de bruit et d'inexactitudes (indiquer leur impact lorsqu'il est connu) :
- Causes pouvant conduire à une perte d'exactitude des données (mise à jour des données réelles, obsolescence, dérives, etc.) :

Division recommandée aux réutilisateurs en jeux d'entraînement, de validation et de test :

4. Conditions d'utilisation

Limitations d'utilisation connues du jeu de données :
Conditions d'utilisation exclues pour le jeu de données :
Licences :
Description de la procédure permettant d'accéder aux données :
Mesures de journalisation concernant l'accès aux données :
Utilisations et projets portant sur les données notables :
Cadre juridique applicable (protection des données personnelles, de la propriété intellectuelle, etc.) :

5. Protection des données personnelles

Base légale de la collecte :

Données se rapportant à des personnes

- Liste des données directement ou indirectement identifiantes :
- Liste des données anonymisées :
- Liste des données sensibles au sens de l'article 9 du RGPD :

Exception mobilisée pour le traitement de données sensibles :
Description de la procédure d'information des personnes et des supports utilisés :

Mesures de protection des données

- Méthodes d'anonymisation ou de pseudonymisation :
- Revue des données selon des critères éthiques ○ Description de la procédure employée :
    - ○ Qualification et compétences de l'entité ou des entités en charge de cette revue :
- Analyse de risque sur les droits et libertés des personnes :
- Mesures de sécurité :
- Cadres de référence, standards, labels applicables :

Mesures d'exercice des droits

- Possibilité d'exercer les droits ○ D'accès :

- De portabilité : ○ De rectification : ○
  D'opposition : ○ A l'effacement :

- Démarche à suivre pour les personnes concernées :
- Démarche prévue suite à la demande d'exercice d'un droit :

Recommandations pour les utilisateurs (durée et conditions de conservation, d'information sur les utilisations, pour l'exercice des droits, etc.) :

6. **Annotation et prétraitement Procédure d'annotation des données**

- Description de la procédure :
- En interne ou par un prestataire :
- Automatisée (pas du tout, partiellement, entièrement) :
- Représentativité démographique de l'équipe d'annotation :
- Garanties apportées sur la responsabilité sociale du prestataire pour l'annotation :

Description de la procédure de vérification des annotations :
Description des méthodes de prétraitement employées :

7. **Maintenance et support**

Description des procédures de maintenance et de support prévues (actuellement et dans l'éventualité où une nouvelle version serait publiée) :
Date prévue pour la fin de la maintenance et du support :
Description de la procédure de mise à jour du jeu de données :
Canaux d'informations sur les mises à jour et évolution du jeu de données :
Description de la procédure permettant de contribuer à l'amélioration ou à la maintenance du jeu de données :

8. **Autres commentaires**