CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS

PROTÉGER les données personnelles
ACCOMPAGNER l'innovation
PRÉSERVER les libertés individuelles

Intelligence artificielle : les premières recommandations de la CNIL pour développer un système d'IA en conformité avec le RGPD

Webinaire du 23 avril 2024

Alexis LEAUTIER, Ingénieur au service de l'intelligence artificielle
Sarah ARTOLA, Juriste au service des affaires économiques

Plan de la présentation

I. Introduction
II. Le périmètre des fiches pratiques
III. Les recommandations de la CNIL
IV. Les travaux à venir

I. INTRODUCTION

Concilier intelligence artificielle et RGPD

- A travers la publication de 8 fiches pratiques, **la CNIL explique comment concilier IA et RGPD**.
- En effet, si le traitement des données personnelles implique le respect des grands principes établis par le RGPD, il s'agit toutefois de principes **généraux qui ont vocation à s'appliquer à une très grande variété de systèmes et technologies.**
- En pratique, **l'application de ces principes nécessite de les articuler avec les spécificités du dispositif concerné et son contexte de mise en œuvre.**

Apporter de la sécurité juridique

- De nombreux acteurs ont fait part à la CNIL de **questionnements concernant l'application du RGPD à l'IA**, en particulier depuis l'émergence de systèmes d'IA génératives (*generative AI*).

- En mai 2023, la CNIL a publié son « plan IA » et a lancé un **travail de clarification du cadre juridique afin de sécuriser les acteurs**.

- Les fiches pratiques ont ainsi vocation à répondre aux principales interrogations :
  - Comment entraîner un modèle d'IA sur **de grands volumes de données en respectant le principe de minimisation** ?
  - Comment définir **la finalité d'un système d'IA à usage général ou d'un modèle de fondation** ?
  - A quelles conditions **la réutilisation des données est-elle possible ?**
  - Combien de temps est-il possible de **conserver une base de données d'apprentissage** ?

Consultation publique

- Ces recommandations ont été élaborées après une série de rencontres avec les différents acteurs de l'écosystème ainsi qu'une **consultation publique de deux mois**.

- Lors la consultation publique, **43 contributions ont été reçues** par la CNIL, émanant d'acteurs variés de l'écosystème de l'IA.

- Cela a permis **d'enrichir ses recommandations, publiées dans leur version finalisée.** Plusieurs précisions et modifications ont donc été apportées, par exemple sur **le périmètre des recommandations et leur articulation avec le projet de règlement IA, l'utilisation d'outils de moissonnage (*web scraping)*, la publication des AIPD,** etc.

- Une synthèse des contribution est disponible sur le site de la CNIL.

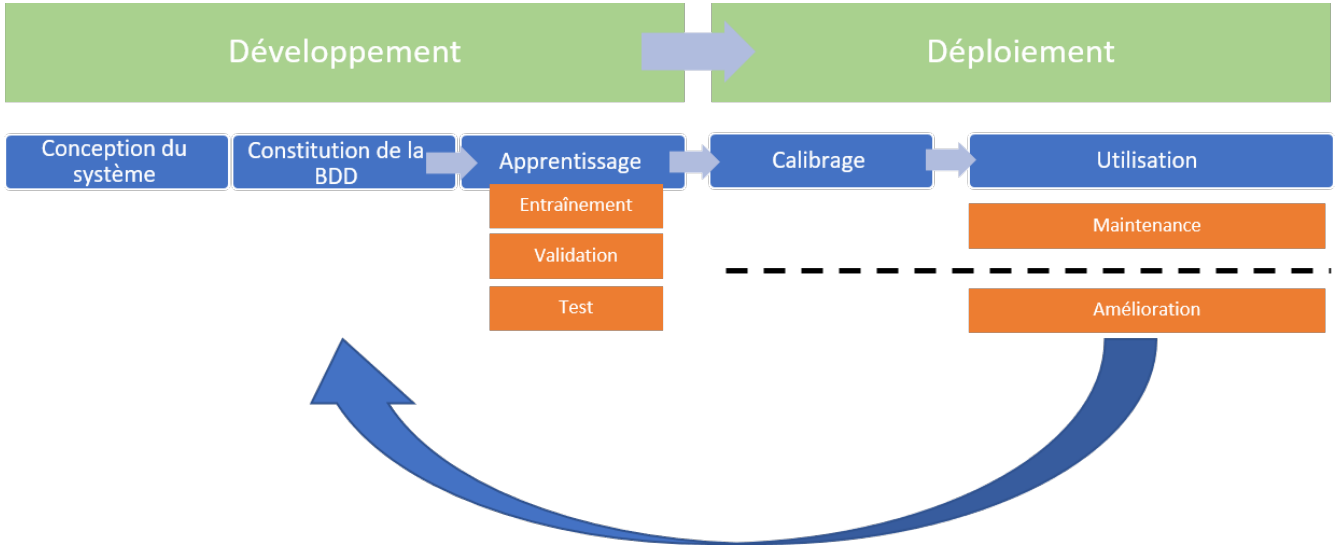
II. PÉRIMÈTRE DES FICHES

Les systèmes d'IA concernés

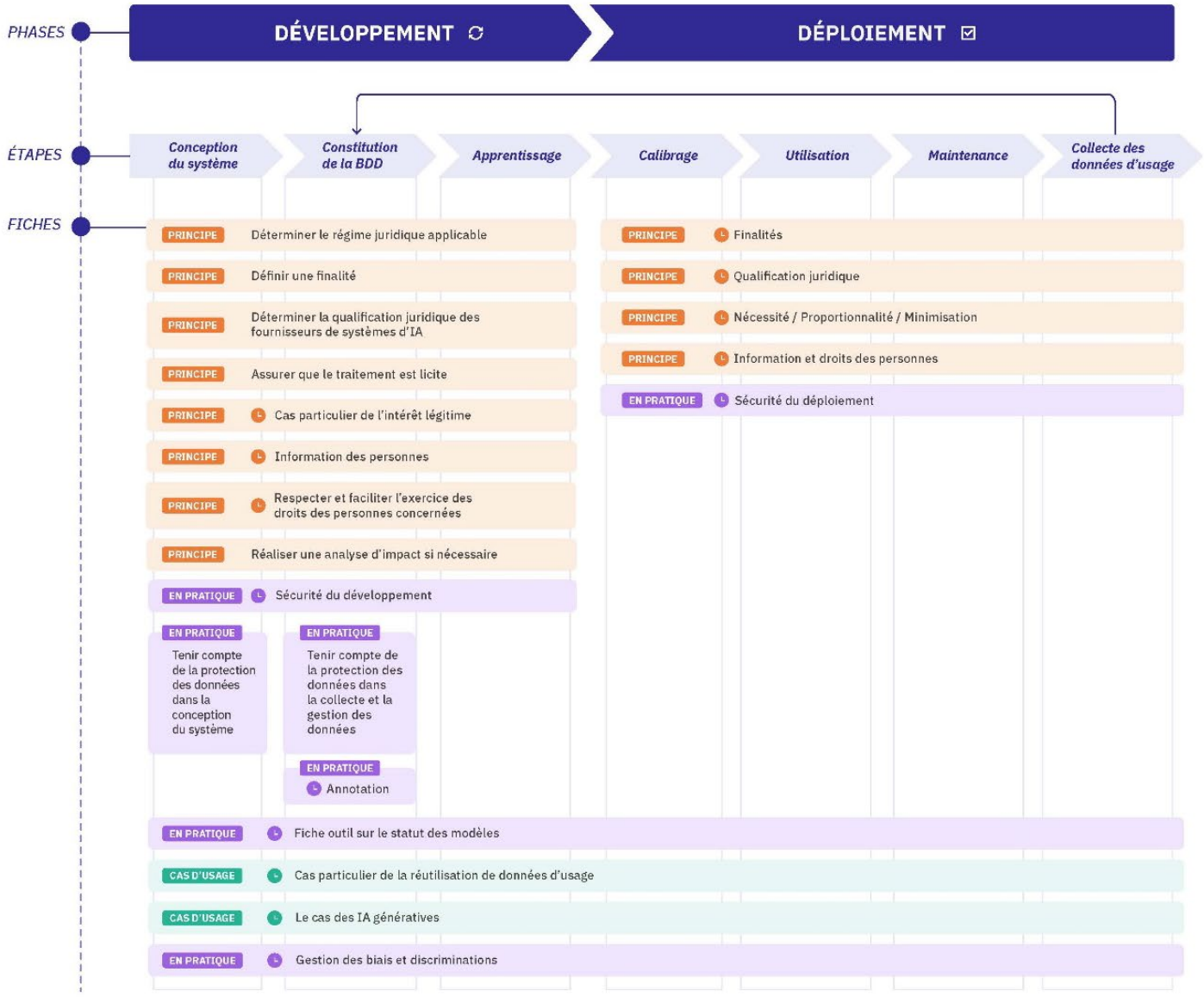
- Ces recommandations concernent le développement de systèmes reposant sur des techniques d'IA et impliquant un **traitement de données personnelles**.

- 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.

La phase de développement



- Ces premières fiches concernent exclusivement les **traitements de données en phase de développement**. Cela inclut :
  - À la fois les cas où l'usage opérationnel est défini dès la phase de développement et **les systèmes d'IA à usage général** (*general purpose AI)*
  - Les systèmes qui impliquent un **apprentissage continu** (où les données collectées lorsque le système est déployé sont réutilisées pour l'amélioration itérative du système)
  - Les traitements consistant à **entraîner ou à ajuster (*fine tuning/transfer learning*) des modèles d'IA existants**, dès lors qu'ils impliquent des données personnelles.


PHASES
DÉVELOPPEMENT
DÉPLOIEMENT
ÉTAPES
Conception du système
Constitution de la BDD
Apprentissage
Calibrage
Utilisation
Maintenance
Collecte des données d'usage
FICHES
PRINCIPE Déterminer le régime juridique applicable
PRINCIPE Définir une finalité
PRINCIPE Déterminer la qualification juridique des fournisseurs de systèmes d'IA
PRINCIPE Assurer que le traitement est licite
PRINCIPE Cas particulier de l'intérêt légitime
PRINCIPE Information des personnes
PRINCIPE Respecter et faciliter l'exercice des droits des personnes concernées
PRINCIPE Réaliser une analyse d'impact si nécessaire
EN PRATIQUE Sécurité du développement
EN PRATIQUE Tenir compte de la protection des données dans la conception du système
EN PRATIQUE Tenir compte de la protection des données dans la collecte et la gestion des données
EN PRATIQUE Annotation
EN PRATIQUE Fiche outil sur le statut des modèles
CAS D'USAGE Cas particulier de la réutilisation de données d'usage
CAS D'USAGE Le cas des IA génératives
EN PRATIQUE Gestion des biais et discriminations
PRINCIPE Finalités
PRINCIPE Qualification juridique
PRINCIPE Nécessité / Proportionnalité / Minimisation
PRINCIPE Information et droits des personnes
EN PRATIQUE Sécurité du déploiement
LÉGENDE
Étape
CLASSIFICATION DES FICHES
PRINCIPE Respect des principes
EN PRATIQUE Mise en oeuvre pratique
CAS D'USAGE Cas d'usage
STATUT DES FICHES
À venir

III. LES RECOMMANDATIONS DE LA CNIL

Fiche 1 - Déterminer le cadre juridique

Cas n°1 : l'usage opérationnel du système d'IA est identifié dès la phase de développement

Si les traitements mis en œuvre en phase de développement poursuivent exclusivement la même finalité que ceux en phase de déploiement, on considère qu'ils relèvent, généralement, du **même régime juridique.**

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

On considère en général dans cette hypothèse que **les traitements en phase de développement sont soumis au RGPD**.

Fiche 2 - Définir une finalité

L'usage opérationnel du système d'IA est précisément identifié dès la phase de développement

Si la finalité en phase de déploiement est déterminée, explicite et légitime, la finalité en phase de développement est également considérée comme telle

L'usage opérationnel du système d'IA n'est pas clairement défini dès la phase de développement (systèmes d'IA à usage général)

La finalité du traitement doit se référer cumulativement :

- au « type » de système développé
- aux fonctionnalités et capacités techniquement envisageables

Il est recommandé que la finalité mentionne également :

- les capacités prévisibles les plus à risque
- les fonctionnalités exclues par conception
- dans la mesure du possible, les conditions d'utilisation du système d'IA

Le système d'IA est développé à des fins de recherche scientifique

Il peut être admis que le degré de précision de la finalité soit plus faible ou que les finalités de la recherche ne soient pas spécifiées dans leur intégralité, compte tenu des difficultés à la cerner entièrement dès le début des travaux.

Fiche 2 - Définir une finalité

L'usage opérationnel du système d'IA est précisément identifié dès la phase de développement

Si la finalité en phase de déploiement est déterminée, explicite et légitime, la finalité en phase de développement est également considérée comme telle

L'usage opérationnel du système d'IA n'est pas clairement défini dès la phase de développement (systèmes d'IA à usage général)

La finalité du traitement doit se référer cumulativement :

- au « type » de système développé
- aux fonctionnalités et capacités techniquement envisageables

Il est recommandé que la finalité mentionne également :

- les capacités prévisibles les plus à risque
- les fonctionnalités exclues par conception
- dans la mesure du possible, les conditions d'utilisation du système d'IA

Le système d'IA est développé à des fins de recherche scientifique

Il peut être admis que le degré de précision de la finalité soit plus faible ou que les finalités de la recherche ne soient pas spécifiées dans leur intégralité, compte tenu des difficultés à la cerner entièrement dès le début des travaux.

Fiche 2 - Définir une finalité

L'usage opérationnel du système d'IA est précisément identifié dès la phase de développement	L'usage opérationnel du système d'IA n'est pas clairement défini dès la phase de développement (systèmes d'IA à usage général)	Le système d'IA est développé à des fins de recherche scientifique
Si la finalité en phase de déploiement est déterminée, explicite et légitime, la finalité en phase de développement est également considérée comme telle	La finalité du traitement doit se référer cumulativement : • au « type » de système développé • aux fonctionnalités et capacités techniquement envisageables	Il peut être admis que le degré de précision de la finalité soit plus faible ou que les finalités de la recherche ne soient pas spécifiées dans leur intégralité, compte tenu des difficultés à la cerner entièrement dès le début des travaux.

Par exemple :

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

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

Fiche 2 - Définir une finalité

L'usage opérationnel du système d'IA est précisément identifié dès la phase de développement

Si la finalité en phase de déploiement est déterminée, explicite et légitime, la finalité en phase de développement est également considérée comme telle

L'usage opérationnel du système d'IA n'est pas clairement défini dès la phase de développement (systèmes d'IA à usage général)

La finalité du traitement doit se référer cumulativement :

- au « type » de système développé
- aux fonctionnalités et capacités techniquement envisageables

Il est recommandé que la finalité mentionne également :

- les capacités prévisibles les plus à risque
- les fonctionnalités exclues par conception
- dans la mesure du possible, les conditions d'utilisation du système d'IA

V

Le système d'IA est développé à des fins de recherche scientifique

Il peut être admis que le degré de précision de la finalité soit plus faible ou que les finalités de la recherche ne soient pas spécifiées dans leur intégralité, compte tenu des difficultés à la cerner entièrement dès le début des travaux.

Fiche 2 - Définir une finalité

L'usage opérationnel du système d'IA est précisément identifié dès la phase de développement	L'usage opérationnel du système d'IA n'est pas clairement défini dès la phase de développement (systèmes d'IA à usage général)	Le système d'IA est développé à des fins de recherche scientifique
Si la finalité en phase de déploiement est déterminée, explicite et légitime, la finalité en phase de développement est également considérée comme telle	La finalité du traitement doit se référer cumulativement : • au « type » de système développé • aux fonctionnalités et capacités techniquement envisageables	Il peut être admis que le degré de précision de la finalité soit plus faible ou que les finalités de la recherche ne soient pas spécifiées dans leur intégralité, compte tenu des difficultés à la cerner entièrement dès le début des travaux.

- La notion de « recherche scientifique » a une **portée large** dans le RGPD. La CNIL a **précisé les critères de définition de la recherche scientifique**, qui concerne beaucoup d'activités de recherche mais aussi de développement, y compris dans le secteur privé.
- Le statut de recherche scientifique **simplifie certaines obligations**.

Fiche 2 - Définir une finalité

L'usage opérationnel du système d'IA est précisément identifié dès la phase de développement	L'usage opérationnel du système d'IA n'est pas clairement défini dès la phase de développement (systèmes d'IA à usage général)	Le système d'IA est développé à des fins de recherche scientifique
Si la finalité en phase de déploiement est déterminée, explicite et légitime, la finalité en phase de développement est également considérée comme telle	La finalité du traitement doit se référer cumulativement : • au « type » de système développé • aux fonctionnalités et capacités techniquement envisageables	Il peut être admis que le degré de précision de la finalité soit plus faible ou que les finalités de la recherche ne soient pas spécifiées dans leur intégralité, compte tenu des difficultés à la cerner entièrement dès le début des travaux.

Exemple 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'entrainement, dans une démarche scientifique documentée ayant vocation à faire l'objet d'une publication.

Fiche 3 – Déterminer sa responsabilité

Le fournisseur du système d'IA

Responsable du traitement

La personne physique ou morale qui détermine les objectifs et les moyens du traitement

Par exemple :
Si le fournisseur est à l'initiative du développement et qu'il constitue la base de données d'apprentissage pour son propre compte.

Si le fournisseur fait appel à un prestataire pour collecter et traiter les données selon ses instructions, celui-ci sera le sous-traitant du fournisseur.

Fiche 3 – Déterminer sa responsabilité

Le fournisseur du système d'IA

Responsable du traitement conjoint

Par exemple :
Si le fournisseur constitue la base de données d'apprentissage avec d'autres responsables de traitements pour un objectif défini ensemble.

Fiche 3 – Déterminer sa responsabilité

Le fournisseur du système d'IA

Sous-traitant

La personne physique ou morale qui traite des données pour le compte du responsable du traitement

Par exemple :
Si le fournisseur développe un système d'IA pour le compte de l'un de ses clients, qui détermine l'objectif, les moyens et les techniques à utiliser.

En revanche : si le client ne donne qu'un objectif à atteindre mais que le fournisseur conçoit le système d'IA, le fournisseur est responsable du traitement.

Fiche 4 – Définir une base légale (1/2)

Le consentement

L'intérêt légitime

La mission d'intérêt public

Autres bases légales

Fiche 4 – Définir une base légale (1/2)

Le consentement | **L'intérêt légitime** | **La mission d'intérêt public** | **Autres bases légales**

- **Libre**
- **Spécifique**
- **Eclairé**
- **Univoque**

→ Peut-être adapté lorsque les données sont collectées directement auprès des personnes

→ Souvent impossible en pratique (par ex. quand les données sont collectées en ligne)

Par 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.

Fiche 4 – Définir une base légale (1/2)

Le consentement	L'intérêt légitime	La mission d'intérêt public	Autres bases légales
• **Libre** • **Spécifique** • **Eclairé** • **Univoque**	• **Légitimité** de l'intérêt poursuivi • **Nécessité** du traitement de données • **Absence d'atteinte disproportionnée** à la vie privée des personnes		
→ Peut-être adapté lorsque les données sont collectées directement auprès des personnes → Souvent impossible en pratique (par ex. quand les données sont collectées en ligne)	**Par exemple :** Constitution d'une base de données d'apprentissage à partir de commentaires pseudonymisés collectés sur des forums publics pour concevoir un système d'IA permettant de prévoir l'appréciation d'œuvre d'art par le grand public.		

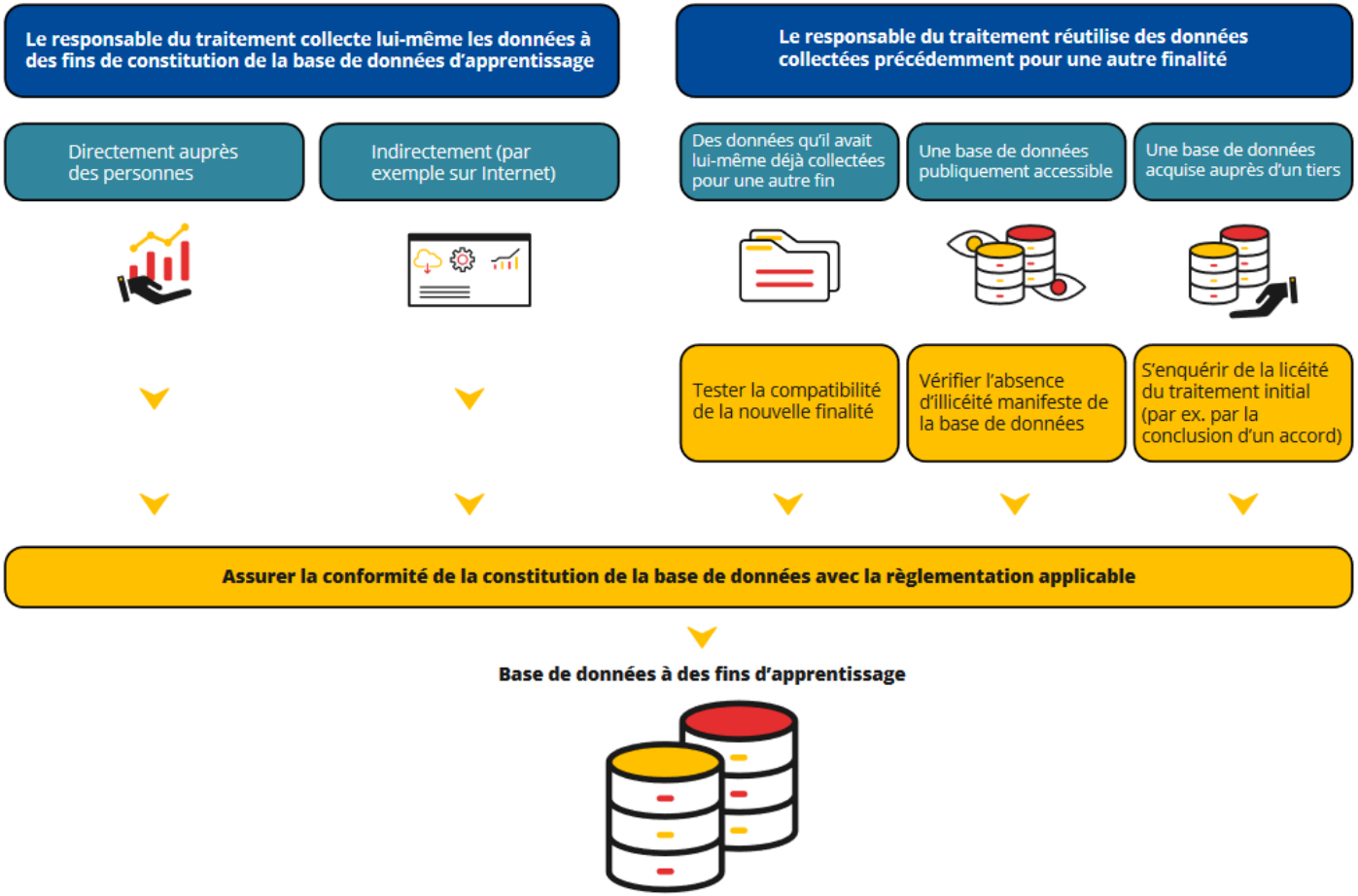
Fiche 4 – Définir une base légale (1/2)

Le consentement	L'intérêt légitime	La mission d'intérêt public	Autres bases légales
• **Libre** • **Spécifique** • **Eclairé** • **Univoque**	• **Légitimité** de l'intérêt poursuivi • **Nécessité** du traitement de données • **Absence d'atteinte disproportionnée** à la vie privée des personnes	• **Mission d'intérêt public prévue par un texte** • **Nécessité** du traitement pour exercer spécifiquement cette mission, de manière pertinente et appropriée	
→ Peut-être adapté lorsque les données sont collectées directement auprès des personnes → Souvent impossible en pratique (par ex. quand les données sont collectées en ligne)	**Par exemple :** Constitution d'une base de données d'apprentissage à partir de commentaires collectés sur des sites internet pour concevoir un système d'IA permettant de prévoir l'appréciation d'œuvre d'art par le grand public.	→ Souvent adapté pour les acteurs publics	

Fiche 4 – Définir une base légale (1/2)

Le consentement	L'intérêt légitime	La mission d'intérêt public	Autres bases légales
• **Libre** • **Spécifique** • **Eclairé** • **Univoque**	• **Légitimité** de l'intérêt poursuivi • **Nécessité** du traitement de données • **Absence d'atteinte disproportionnée** à la vie privée des personnes	• **Mission d'intérêt public prévue par un texte** • **Nécessité** du traitement pour exercer spécifiquement la mission d'intérêt public	Les bases légales du **contrat** ou de **l'obligation légale** peuvent être plus exceptionnellement mobilisées.
→ Peut-être adapté lorsque les données sont collectées directement auprès des personnes → Souvent impossible en pratique (par ex. quand les données sont collectées en ligne)	**Par exemple :** Constitution d'une base de données d'apprentissage à partir de commentaires collectés sur des sites internet pour concevoir un système d'IA permettant de prévoir l'appréciation d'œuvre d'art par le grand public.	→ Souvent adapté pour les acteurs publics	

Fiche 4 – Effectuer les tests nécessaires pour la réutilisation des données (2/2)



Fiche 4 – Effectuer les tests nécessaires pour la réutilisation des données (2/2)

La réutilisation de jeux de données ouverts

- La CNIL est tout à fait consciente de l'importance de la réutilisation de jeux de données mis en accès ouvert pour l'écosystème IA (pour la transparence, la reproductibilité, l'analyse comparative de solutions (benchmarking), etc.).
- Elle considère que de tels jeux de données peuvent être réutilisés par les acteurs à la condition de **s'assurer qu'il n'y a pas d'illicéité manifeste et évidente**.

Par exemple :

☺ Une base de données, sans données sensibles, constituée à partir de publications sur un réseau social professionnel nommément désigné.

☹ Une base de données contenant des images de vidéosurveillance sans préciser la source 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.

- Le réutilisateur n'a pas à s'assurer de l'ensemble des obligations du RGPD par le responsable du traitement initial. Il doit toutefois garantir lui-même la conformité des usages qu'il fait de ces données (information des personnes, exercice des droits, durée de conservation, etc.).

Fiche 4 – Effectuer les tests nécessaires pour la réutilisation des données (2/2)

La réutilisation un jeu de données initialement constitué à des fins de recherche scientifique

- Lorsque le responsable du traitement a traité des données à des fins de recherche scientifique et qu'il entend les réutiliser à d'autres fins, **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** (« test de compatibilité ») **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**.

Fiche 4 – Effectuer les tests nécessaires pour la réutilisation des données (2/2)

La réutilisation d'une base de données acquise auprès d'un tiers

- Le **réutilisateur** des données doit :
  - S'assurer de **ne pas réutiliser une base de données manifestement illicite**. La conclusion d'un accord entre le détenteur initial des données et le réutilisateur est recommandé.
  - S'assurer de la **conformité** de ses propres traitements de données au RGPD.

Fiche 5 – Réaliser une AIPD si nécessaire (1/2)

Obligatoire
Pour **les traitements à haut risque et les systèmes d'IA à haut risque** selon le Règlement IA

Facultative
Dans les autres cas

Fiche 5 – Réaliser une AIPD si nécessaire (1/2)

Obligatoire
Pour **les traitements à haut risque et les systèmes d'IA à haut risque** selon le Règlement IA

Facultative
Dans les autres cas

Fiche 5 – Réaliser une AIPD si nécessaire (1/2)

Obligatoire
Pour **les traitements à haut risque et les systèmes d'IA à haut risque** selon le Règlement IA

Facultative
Dans les autres cas

Liste de critères

la collecte de **données sensibles** ou de données à caractère **hautement personnel** ;

le traitement de données **à grande échelle** ;

la collecte de données concernant des **personnes vulnérables**, telles que les enfants;

Le **croisement** d'ensembles de données;

Les **traitements innovants** ou l'utilisation de nouvelles solutions technologiques ou organisationnelles;

etc.

Fiche 5 – Réaliser une AIPD si nécessaire (1/2)

Obligatoire
Pour **les traitements à haut risque et les systèmes d'IA à haut risque** selon le Règlement IA

Facultative
Dans les autres cas

Liste de critères

la collecte de **données sensibles** ou de données à caractère **hautement personnel** ;

le traitement de données **à grande échelle** ;

la collecte de données concernant des **personnes vulnérables**, telles que les enfants;

Le **croisement** d'ensembles de données;

Les **traitements innovants** ou l'utilisation de nouvelles solutions technologiques ou organisationnelles;

etc.

Fiche 5 – Réaliser une AIPD si nécessaire (1/2)

Obligatoire
Pour **les traitements à haut risque et les systèmes d'IA à haut risque** selon le Règlement IA

Facultative
Dans les autres cas

Liste de critères

la collecte de **données sensibles** ou de données à caractère **hautement personnel** ;

le traitement de données **à grande échelle** ;

la collecte de données concernant des **personnes vulnérables**, telles que les enfants;

Le **croisement** d'ensembles de données;

Les **traitements innovants** ou l'utilisation de nouvelles solutions technologiques ou organisationnelles;

etc.

La réalisation d'une AIPD est toujours **une bonne pratique**.

Fiche 5 – Réaliser une AIPD si nécessaire (2/2)

Lister et évaluer les risques

Prévoir et mettre en œuvre un plan d'action

Fiche 5 – Réaliser une AIPD si nécessaire (2/2)

Lister et évaluer les risques

Prévoir et mettre en œuvre un plan d'action

Fiche 5 – Réaliser une AIPD si nécessaire (2/2)

Lister et évaluer les risques

Prévoir et mettre en œuvre un plan d'action

L'**utilisation abusive ou détournée** des données (violation de données) ;

La discrimination automatisée ;

La production de faux contenus sur une personne réelle ;

La prise de décision automatisée ;

La perte de contrôle sur les données publiées en ligne ;

Les attaques connues (empoisonnement des données, injection de porte dérobée, inversion du modèle) ;

L'extraction de données d'entraînement à partir du modèle.

Fiche 5 – Réaliser une AIPD si nécessaire (2/2)

Lister et évaluer les risques

L'**utilisation abusive ou détournée** des données (violation de données) ;

La discrimination automatisée ;

La production de faux contenus sur une personne réelle ;

La prise de décision automatisée ;

La perte de contrôle sur les données publiées en ligne ;

Les attaques connues (empoisonnement des données, injection de porte dérobée, inversion du modèle) ;

L'extraction de données d'entraînement à partir du modèle.

Prévoir et mettre en œuvre un plan d'action

Fiche 5 – Réaliser une AIPD si nécessaire (2/2)

Lister et évaluer les risques

L'**utilisation abusive ou détournée** des données (violation de données) ;

La discrimination automatisée ;

La production de faux contenus sur une personne réelle ;

La prise de décision automatisée ;

La perte de contrôle sur les données publiées en ligne ;

Les attaques connues (empoisonnement des données, injection de porte dérobée, inversion du modèle) ;

L'extraction de données d'entraînement à partir du modèle.

Prévoir et mettre en œuvre un plan d'action

Prévoir des mesures concernant :

La sécurité, la minimisation, la protection des données dès la conception (anonymisation ou pseudonymisation) ;

Facilitant **l'exercice des droits** des individus ;

L'audit et le test du système ;

Les processus et l'organisation (surveillance et limitation de l'accès aux données en interne, par des tiers et sous-traitants) ;

La gouvernance (comité éthique) ;

La journalisation pour identifier et expliquer les comportements inhabituels ;

La documentation.

Fiche 5 – Réaliser une AIPD si nécessaire (2/2)

Lister et évaluer les risques

L'**utilisation abusive ou détournée** des données (violation de données) ;

La discrimination automatisée ;

La production de faux contenus sur une personne réelle ;

La prise de décision automatisée ;

La perte de contrôle sur les données publiées en ligne ;

Les attaques connues (empoisonnement des données, injection de porte dérobée, inversion du modèle) ;

L'extraction de données d'entraînement à partir du modèle.

Prévoir et mettre en œuvre un plan d'action

Prévoir des mesures concernant :

La sécurité, la minimisation, la protection des données dès la conception (anonymisation ou pseudonymisation) ;

Facilitant **l'exercice des droits** des individus ;

L'audit et le test du système ;

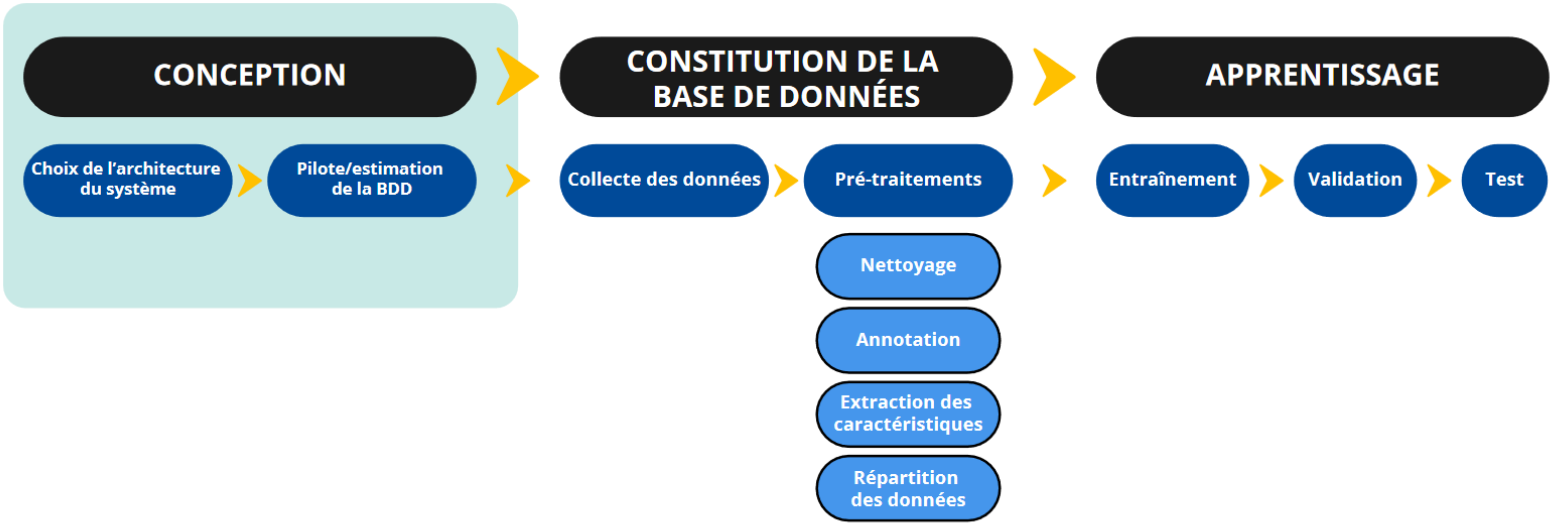
Les processus et l'organisation (surveillance et limitation de l'accès aux données en interne, par des tiers et sous-traitants) ;

La gouvernance (comité éthique) ;

La journalisation pour identifier et expliquer les comportements inhabituels ;

La documentation.

Fiche 6 – Tenir compte de la protection des données dans la conception du système (0/2)



Fiche 6 – Tenir compte de la protection des données dans la conception du système (1/2)

Principe de minimisation (article 5.1):

> *« Les données à caractère personnel sont* ***adéquates, pertinentes et limitées à ce qui est nécessaire*** *»*

Fiche 6 – Tenir compte de la protection des données dans la conception du système (1/2)

Principe de minimisation (article 5.1):

> *« Les données à caractère personnel sont **adéquates, pertinentes et limitées à ce qui est nécessaire** »*

Comment minimiser la collecte de données dans le développement de l'IA?

Fiche 6 – Tenir compte de la protection des données dans la conception du système (1/2)

Principe de minimisation (article 5.1):

> *« Les données à caractère personnel sont **adéquates, pertinentes et limitées à ce qui est nécessaire** »*

Comment minimiser la collecte de données dans le développement de l'IA?

Choisir un modèle adapté

Fiche 6 – Tenir compte de la protection des données dans la conception du système (1/2)

Principe de minimisation (article 5.1):

> *« Les données à caractère personnel sont **adéquates, pertinentes et limitées à ce qui est nécessaire** »*

Comment minimiser la collecte de données dans le développement de l'IA?

Choisir un modèle adapté	**Sélectionner les données pertinentes**

Fiche 6 – Tenir compte de la protection des données dans la conception du système (1/2)

Principe de minimisation (article 5.1):

> *« Les données à caractère personnel sont **adéquates, pertinentes et limitées à ce qui est nécessaire** »*

Comment minimiser la collecte de données dans le développement de l'IA?

Choisir un modèle adapté	**Sélectionner les données pertinentes**	**Valider ses choix de conception**

Fiche 6 – Tenir compte de la protection des données dans la conception du système (2/2)

Choisir un modèle adapté	Sélectionner les données pertinentes	Valider ses choix de conception

Fiche 6 – Tenir compte de la protection des données dans la conception du système (2/2)

Choisir un modèle adapté	Sélectionner les données pertinentes	Valider ses choix de conception

Fiche 6 – Tenir compte de la protection des données dans la conception du système (2/2)

Choisir un modèle adapté | Sélectionner les données pertinentes | Valider ses choix de conception

Quel est l'objectif ?

- le type de **résultat/sortie** ;
- les **performances acceptables**, **quantitatives** (score F1, précision/recall, temps de calcul, etc.) ou **qualitatives** (retour des utilisateurs) ;
- le **contexte d'utilisation prévus** ;
- les **contextes d'utilisation exclus** et les informations qui ne sont pas pertinentes.

Quel modèle choisir ?

Existe-t-il un modèle **plus frugal pour les mêmes performances** ?
un **modèle pré-entraîné** suffirait-il ?

Se fonder sur :

la littérature scientifique, une comparaison des résultats de plusieurs architectures, entre le fine-tuning et l'entraînement « from scratch »…

Fiche 6 – Tenir compte de la protection des données dans la conception du système (2/2)

Choisir un modèle adapté | Sélectionner les données pertinentes | Valider ses choix de conception

Quel est l'objectif ?

- le type de **résultat/sortie** ;
- les **performances acceptables**, **quantitatives** (score F1, précision/recall, temps de calcul, etc.) ou **qualitatives** (retour des utilisateurs) ;
- le **contexte d'utilisation prévus** ;
- les **contextes d'utilisation exclus** et les informations qui ne sont pas pertinentes.

Quel modèle choisir ?

Existe-t-il un modèle **plus frugal pour les mêmes performances** ?
un **modèle pré-entraîné** suffirait-il ?

Se fonder sur :

la littérature scientifique, une comparaison des résultats de plusieurs architectures, entre le fine-tuning et l'entraînement « from scratch »…

Exemple :

Pour un **système de comptage des personnes debout** dans un tramway sur des images de caméras de vidéoprotection

- **Option 1** : un réseau de neurones **détectant les personnes** à l'image (sans analyse de la posture) : le nombre de personnes debout est déduit grâce au nombre de places assises ;

- **Option 2** : un réseau de neurones réalisant **une analyse de la posture** des personnes couplé à un algorithme réalisant un décompte des personnes qui se tiennent debout.

Fiche 6 – Tenir compte de la protection des données dans la conception du système (2/2)

Choisir un modèle adapté	**Sélectionner les données pertinentes**	Valider ses choix de conception

Quel est l'objectif ?

- le type de **résultat/sortie** ;
- les **performances acceptables**, **quantitatives** (score F1, précision/recall, temps de calcul, etc.) ou **qualitatives** (retour des utilisateurs) ;
- le **contexte d'utilisation prévus** ;
- les **contextes d'utilisation exclus** et les informations qui ne sont pas pertinentes.

Quel modèle choisir ?

Existe-t-il un modèle **plus frugal pour les mêmes performances** ?
un **modèle pré-entraîné** suffirait-il ?

Se fonder sur :

la littérature scientifique, une comparaison des résultats de plusieurs architectures, entre le fine-tuning et l'entraînement « from scratch »…

Fiche 6 – Tenir compte de la protection des données dans la conception du système (2/2)

Choisir un modèle adapté	Sélectionner les données pertinentes	Valider ses choix de conception

Quel est l'objectif ?

- le type de **résultat/sortie** ;
- les **performances acceptables**, **quantitatives** (score F1, précision/recall, temps de calcul, etc.) ou **qualitatives** (retour des utilisateurs) ;
- le **contexte d'utilisation prévus** ;
- les **contextes d'utilisation exclus** et les informations qui ne sont pas pertinentes.

Quel modèle choisir ?

Existe-t-il un modèle **plus frugal pour les mêmes performances** ?
un **modèle pré-entraîné** suffirait-il ?

Se fonder sur :

la littérature scientifique, une comparaison des résultats de plusieurs architectures, entre le fine-tuning et l'entraînement « from scratch »…

Quel volume ?

nombre de personnes, profondeur historique, précision, représentativité, etc.

Quelles catégories de données, métadonnées et caractéristiques ?

âge, sexe, photos de visages, activité sur les réseaux sociaux, etc.

De quelle nature ?

données réelles, synthétisées, augmentées, simulées, anonymisées, pseudonymisées, **sensibles ou hautement personnelles**.

Quelles sources ?

collecte ad hoc (scraping, collecte directe) ou **réutilisation** (open source, collecte précédente, data broker).

Fiche 6 – Tenir compte de la protection des données dans la conception du système (2/2)

Choisir un modèle adapté

Quel est l'objectif ?

- le type de **résultat/sortie** ;
- les **performances acceptables**, **quantitatives** (score F1, précision/recall, temps de calcul, etc.) ou **qualitatives** (retour des utilisateurs) ;
- le **contexte d'utilisation prévus** ;
- les **contextes d'utilisation exclus** et les informations qui ne sont pas pertinentes.

Quel modèle choisir ?

Existe-t-il un modèle **plus frugal pour les mêmes performances** ?
un **modèle pré-entraîné** suffirait-il ?

Se fonder sur :

la littérature scientifique, une comparaison des résultats de plusieurs architectures, entre le fine-tuning et l'entraînement « from scratch »…

Sélectionner les données pertinentes

Quel volume ?

nombre de personnes, profondeur historique, précision, représentativité, etc.

Quelles catégories de données, métadonnées et caractéristiques ?

âge, sexe, photos de visages, activité sur les réseaux sociaux, etc.

De quelle nature ?

données réelles, synthétisées, augmentées, simulées, anonymisées, pseudonymisées, **sensibles ou hautement personnelles**.

Quelles sources ?

collecte ad hoc (scraping, collecte directe) ou **réutilisation** (open source, collecte précédente, data broker).

Valider ses choix de conception

Fiche 6 – Tenir compte de la protection des données dans la conception du système (2/2)

Choisir un modèle adapté

Quel est l'objectif ?

- le type de **résultat/sortie** ;
- les **performances acceptables**, **quantitatives** (score F1, précision/recall, temps de calcul, etc.) ou **qualitatives** (retour des utilisateurs) ;
- le **contexte d'utilisation prévus** ;
- les **contextes d'utilisation exclus** et les informations qui ne sont pas pertinentes.

Quel modèle choisir ?

Existe-t-il un modèle **plus frugal pour les mêmes performances** ?
un **modèle pré-entraîné** suffirait-il ?

Se fonder sur :

la littérature scientifique, une comparaison des résultats de plusieurs architectures, entre le fine-tuning et l'entraînement « from scratch »…

Sélectionner les données pertinentes

Quel volume ?

nombre de personnes, profondeur historique, précision, représentativité, etc.

Quelles catégories de données, métadonnées et caractéristiques ?

âge, sexe, photos de visages, activité sur les réseaux sociaux, etc.

De quelle nature ?

données réelles, synthétisées, augmentées, simulées, anonymisées, pseudonymisées, **sensibles ou hautement personnelles**.

Quelles sources ?

collecte ad hoc (scraping, collecte directe) ou **réutilisation** (open source, collecte précédente, data broker).

Valider ses choix de conception

Réalisation d'une étude pilote

- dans un **environnement contrôlé, à petite échelle et de courte durée ;**
- de préférence sur des **données fictives, synthétiques ou anonymisées.**

Impliquer un comité éthique

- définir les **valeurs de l'organisation et les appliquer ;**
- **anticiper** les utilisations opérationnelles**,** leurs **conséquences individuelles ou sociétales** et les moyens de **les prévenir ;**
- prévoir et prévenir **les utilisations abusives ou détournées ;**
- garantir **la transparence pour l'exercice des droits ;**
- questionner **les choix techniques et organisationnels.**

Fiche 6 – Tenir compte de la protection des données dans la conception du système (2/2)

Choisir un modèle adapté	Sélectionner les données pertinentes	**Valider ses choix de conception**

Exemple :
Utilisation de données issues **de réseaux sociaux sur les pages de personnes ayant consenti** à la collecte**.**

- Échantillon **non-représentatif** de l'activité sur les réseaux,
- Mais peut être adapté à certains cas d'usage comme **l'identification de contenus haineux ou l'étude du ciblage publicitaire sur ces réseaux**.

Le bénéfice : un niveau de **transparence largement supérieur** au moissonnage.

Réalisation d'une étude pilote

- dans un **environnement contrôlé, à petite échelle et de courte durée ;**
- de préférence sur des **données fictives, synthétiques ou anonymisées.**

Impliquer un comité éthique

- définir les **valeurs de l'organisation et les appliquer ;**
- **anticiper** les utilisations opérationnelles**,** leurs **conséquences individuelles ou sociétales** et les moyens de **les prévenir ;**
- prévoir et prévenir **les utilisations abusives ou détournées ;**
- garantir **la transparence pour l'exercice des droits ;**
- questionner **les choix techniques et organisationnels.**

Fiche 6 – Tenir compte de la protection des données dans la conception du système (2/2)

Choisir un modèle adapté

Quel est l'objectif ?

- le type de **résultat/sortie** ;
- les **performances acceptables**, **quantitatives** (score F1, précision/recall, temps de calcul, etc.) ou **qualitatives** (retour des utilisateurs) ;
- le **contexte d'utilisation prévus** ;
- les **contextes d'utilisation exclus** et les informations qui ne sont pas pertinentes.

Quel modèle choisir ?

Existe-t-il un modèle **plus frugal pour les mêmes performances** ?
un **modèle pré-entraîné** suffirait-il ?

Se fonder sur :

la littérature scientifique, une comparaison des résultats de plusieurs architectures, entre le fine-tuning et l'entraînement « from scratch »…

Sélectionner les données pertinentes

Quel volume ?

nombre de personnes, profondeur historique, précision, représentativité, etc.

Quelles catégories de données, métadonnées et caractéristiques ?

âge, sexe, photos de visages, activité sur les réseaux sociaux, etc.

De quelle nature ?

données réelles, synthétisées, augmentées, simulées, anonymisées, pseudonymisées, **sensibles ou hautement personnelles**.

Quelles sources ?

collecte ad hoc (scraping, collecte directe) ou **réutilisation** (open source, collecte précédente, data broker).

Valider ses choix de conception

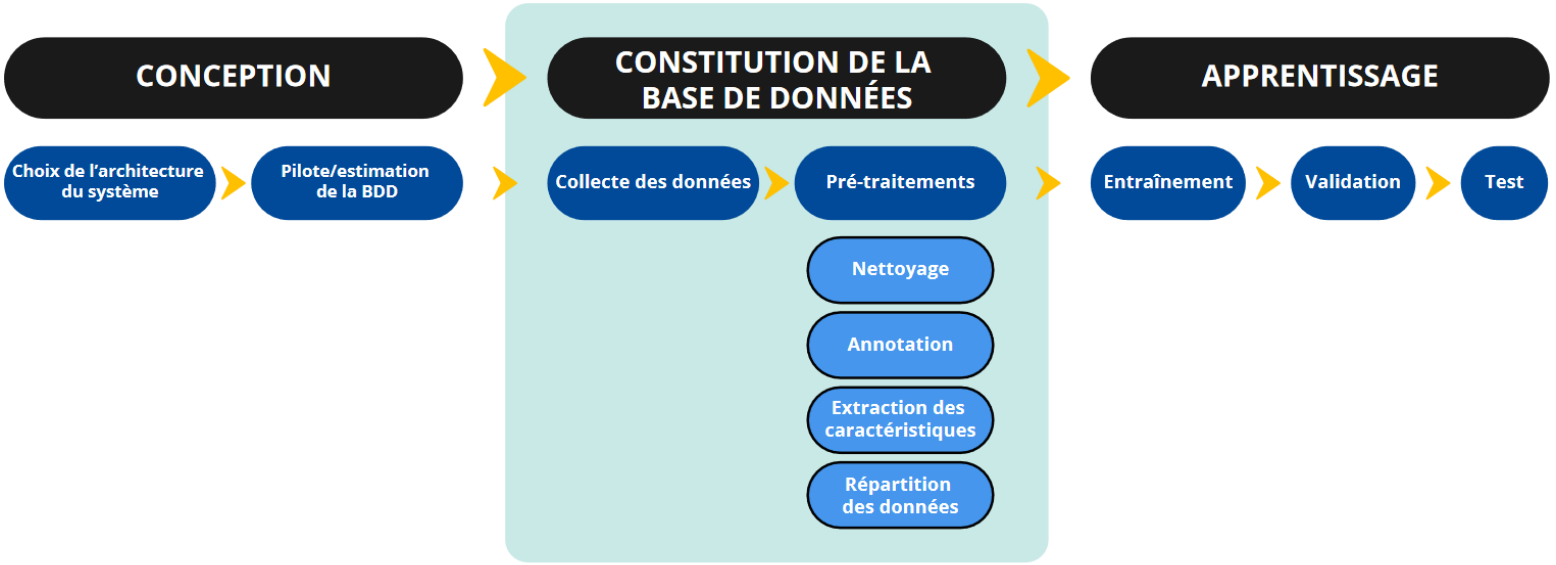
Réalisation d'une étude pilote

- dans un **environnement contrôlé, à petite échelle et de courte durée ;**
- de préférence sur des **données fictives, synthétiques ou anonymisées.**

Impliquer un comité éthique

- définir les **valeurs de l'organisation et les appliquer ;**
- **anticiper** les utilisations opérationnelles**,** leurs **conséquences individuelles ou sociétales** et les moyens de **les prévenir ;**
- prévoir et prévenir **les utilisations abusives ou détournées ;**
- garantir **la transparence pour l'exercice des droits ;**
- questionner **les choix techniques et organisationnels.**

Fiche 7 – Tenir compte de la protection des données dans la collecte et la gestion des données (0/2)



Fiche 7 – Tenir compte de la protection des données dans la collecte et la gestion des données (1/2)

Lors du prétraitement

Fiche 7 – Tenir compte de la protection des données dans la collecte et la gestion des données (1/2)

Lors du prétraitement

Nettoyer les données,

Identifier les caractéristiques pertinentes, par des techniques comme l'analyse en composantes principales (PCA) ;

Sélectionner les données pendant l'apprentissage, par des techniques comme l'*active learning*.
Réduire le volume de données sans impact sur les performances.

Fiche 7 – Tenir compte de la protection des données dans la collecte et la gestion des données (1/2)

Lors du prétraitement

Nettoyer les données,

Identifier les caractéristiques pertinentes, par des techniques comme l'analyse en composantes principales (PCA) ;

Sélectionner les données pendant l'apprentissage, par des techniques comme l'*active learning*.
Réduire le volume de données sans impact sur les performances.

Suivre les données au cours du temps

Fiche 7 – Tenir compte de la protection des données dans la collecte et la gestion des données (1/2)

Lors du prétraitement

Nettoyer les données,

Identifier les caractéristiques pertinentes, par des techniques comme l'analyse en composantes principales (PCA) ;

Sélectionner les données pendant l'apprentissage, par des techniques comme l'*active learning*.
Réduire le volume de données sans impact sur les performances.

Suivre les données au cours du temps

Identifier une dérive de données : changements de processus (remplacement d'un capteur, changement par rapport à la configuration d'étalonnage), **perte de qualité des données** (capteur détérioré), **dérive naturelle** (variations saisonnières), **changements soudains** (apparition des masques lors du COVID-19), **évolution des corrélations entre les caractéristiques, empoisonnement des données**.

Identifier les évolutions des données : mise à jour dans un profil de réseau social.
Nouvelles méthodes nécessitant moins de données.

Fiche 7 – Tenir compte de la protection des données dans la collecte et la gestion des données (1/2)

Lors du prétraitement

Nettoyer les données,

Identifier les caractéristiques pertinentes, par des techniques comme l'analyse en composantes principales (PCA) ;

Sélectionner les données pendant l'apprentissage, par des techniques comme l'*active learning*.
Réduire le volume de données sans impact sur les performances.

Suivre les données au cours du temps

Identifier une dérive de données : changements de processus (remplacement d'un capteur, changement par rapport à la configuration d'étalonnage), **perte de qualité des données** (capteur détérioré), **dérive naturelle** (variations saisonnières), **changements soudains** (apparition des masques lors du COVID-19), **évolution des corrélations entre les caractéristiques, empoisonnement des données**.

Identifier les évolutions des données : mise à jour dans un profil de réseau social.
Nouvelles méthodes nécessitant moins de données.

Fiche 7 – Tenir compte de la protection des données dans la collecte et la gestion des données (2/2)

Documenter les données

Fiche 7 – Tenir compte de la protection des données dans la collecte et la gestion des données (2/2)

Documenter les données

A destination des **équipes techniques et des utilisateurs,**
Pour **faciliter l'utilisation, démontrer que la collecte est légale,** faciliter **le suivi** des données au fil du temps**,** réduire le **risque d'utilisation imprévue, permettre l'exercice des droits,** identifier **les améliorations prévues ou possibles.**

Gebru et al., 2021 («Datasheets for datasets»), Arnold et al., 2019, Bender et al., 2018, Dataset Nutrition Label, CrowdWorkSheets.

Fiche 7 – Tenir compte de la protection des données dans la collecte et la gestion des données (2/2)

Documenter les données

Pour **faciliter l'** … s données au fil du temps, réd… **s,** identifier **les**

Gebru et al., 2021 … rowdWorkSheets.

Description of the dataset
Dataset reference:

1. **Synthesis**

Description of the dataset:
Time limits for erasure:
Access restrictions:
Use restrictions:
Presence of personal, highly personal, sensitive or otherwise protected data:
Version of the description sheet and last update:

2. **Context and motivation**

Identity of the dataset provider

- Name of the organisation:
- Status:
- Contact address:
- Relation to other bodies:

Motivation for the establishment of the dataset

- Purpose:
- Issue motivating the collection:
- Expected result of processing of the dataset (task or functionality of the system):
- Added value compared to existing datasets:

Fiche 7 – Tenir compte de la protection des données dans la collecte et la gestion des données (2/2)

Documenter les données

A destination des **équipes techniques et des utilisateurs,**
Pour **faciliter l'utilisation, démontrer que la collecte est légale,** faciliter **le suivi** des données au fil du temps**,** réduire le **risque d'utilisation imprévue, permettre l'exercice des droits,** identifier **les améliorations prévues ou possibles.**

Gebru et al., 2021 («Datasheets for datasets»), Arnold et al., 2019, Bender et al., 2018, Dataset Nutrition Label, CrowdWorkSheets.

Limiter la conservation dans le temps

Fiche 7 – Tenir compte de la protection des données dans la collecte et la gestion des données (2/2)

Documenter les données

A destination des **équipes techniques et des utilisateurs,**
Pour **faciliter l'utilisation, démontrer que la collecte est légale,** faciliter **le suivi** des données au fil du temps**,** réduire le **risque d'utilisation imprévue, permettre l'exercice des droits,** identifier **les améliorations prévues ou possibles.**

Gebru et al., 2021 («Datasheets for datasets»), Arnold et al., 2019, Bender et al., 2018, Dataset Nutrition Label, CrowdWorkSheets.

Limiter la conservation dans le temps

Définir une durée **pour la conception,**
et **une durée pour la maintenance** (comme pour la surveillance des biais)**.**

Adapter **les conditions de conservation et de sécurité à chaque phase,**
Effectuer **une nouvelle sélection** des données **entre les deux phases.**

Supprimer ou anonymiser les données après cela.

Fiche 7 – Tenir compte de la protection des données dans la collecte et la gestion des données (2/2)

Documenter les données

A destination des **équipes techniques et des utilisateurs,**
Pour **faciliter l'utilisation, démontrer que la collecte est légale,** faciliter **le suivi** des données au fil du temps**,** réduire le **risque d'utilisation imprévue, permettre l'exercice des droits,** identifier **les améliorations prévues ou possibles.**

Gebru et al., 2021 («Datasheets for datasets»), Arnold et al., 2019, Bender et al., 2018, Dataset Nutrition Label, CrowdWorkSheets.

Limiter la conservation dans le temps

Définir une durée **pour la conception,**
et **une durée pour la maintenance** (comme pour la surveillance des biais)**.**

Adapter **les conditions de conservation et de sécurité à chaque phase,**
Effectuer **une nouvelle sélection** des données **entre les deux phases.**

Supprimer ou anonymiser les données après cela.

IV. LES TRAVAUX À VENIR

Publication d'une deuxième série de fiches pratiques IA

- Ces premières recommandations seront complétées par **une deuxième série de fiches pratiques** :
  - Mobiliser la base légale de l'intérêt légitime pour le développement des systèmes d'IA, avec un focus sur l'*open source* et le moissonnage des données (*web scraping*) ;
  - Informer les personnes ;
  - Garantir et faciliter l'exercice des droits des personnes ;
  - L'annotation des données ;
  - La sécurité des traitements en phase de développement

- **Ces fiches seront publiées prochainement pour consultation publique.**

- La CNIL **poursuit également ses travaux doctrinaux** pour les traitements en phase de déploiement, pour certains cas d'usage plus précis (réutilisation des données, l'IA générative, etc.) et l'articulation entre les exigences du RGPD et du futur règlement européen sur l'IA.

Poursuite des travaux doctrinaux

- La CNIL **poursuit également ses travaux doctrinaux** pour les traitements en phase de déploiement et pour certains cas d'usage plus précis (réutilisation des données, l'IA générative, la gestion des biais, etc.). Ces travaux feront l'objet de publications ultérieures.

- La CNIL participe aussi activement aux **travaux du CEPD sur l'articulation entre le RGPD et le futur règlement européen sur l'IA**.

CNIL.

Contact : ia@cnil.fr

CNIL.

MERCI DE VOTRE ATTENTION !