DÉVELOPPEMENT DES SYSTÈMES D'IA : QUE FAUT-IL VÉRIFIER ?

La CNIL propose une **liste des points à vérifier**, issue de ses **recommandations pour le développement des systèmes d'IA conformes au RGPD.**

Elle s'adresse aux concepteurs et développeurs de systèmes d'intelligence artificielle (chefs de de produit, développeurs, délégué à la protection des données, équipe juridique, responsable de la sécurité des systèmes d'information, etc.) afin de sécuriser l'intégralité des étapes du développement d'un système d'IA, **de la collecte des données, jusqu'à l'intégration**, en passant par l'**apprentissage de modèles** ou l'**annotation**. Cette check-list vise à assurer que, dès le départ, les principes du RGPD sont correctement mis en œuvre : finalité, minimisation, sécurité, information, droits des personnes, transparence et gouvernance.

À noter : les obligations posées par le règlement sur l'intelligence artificielle doivent, le cas échéant, également être prises en compte dans le développement de ces systèmes. Elles ne sont pas couvertes par cette check-list.

FICHES			MESURES	
1	Déterminer le régime juridique applicable et votre responsabilité Définir un cadre pour les utilisateurs	Identifier si le RGPD s'applique	Identifier si la base de données d'entraînement contient des **données personnelles** (y compris issues du *web scraping*).	☐
			Analyser si le RGPD s'applique au modèle appris à partir de bases de données d'entraînement contenant des données personnelles ou si celui-ci peut être présumé anonyme. • Pour cela, établir la nécessité de conduire des attaques en réidentification sur le modèle d'IA, la profondeur de ces attaques, et la vraisemblance d'extraction de données personnelles en détaillant autant que possible par typologie de données.	☐
			Si vous estimez que le système qui intègre un modèle d'IA non anonyme peut permettre de sortir du champ d'application du RGPD, vérifiez que les mesures mises en place sont suffisamment efficaces et robustes, pour rendre insignifiante la vraisemblance de réidentification de personnes. • Cette évaluation implique nécessairement la conduite d'attaques en réidentification sur le système d'IA.	☐
			Mettre en œuvre un processus de réévaluation régulière du caractère anonyme du modèle ou du système.	☐
		Définir les responsabilités des acteurs	Déterminez votre **responsabilité** et celles des autres intervenants dans le traitement des données personnelles (responsable du traitement, responsable conjoint, ou sous-traitant).	☐
			Le cas échéant, assurez-vous de conclure un contrat pour **encadrer la responsabilité conjointe.**	☐
			Le cas échéant, assurez-vous de conclure un contrat pour **documenter les instructions** données à vos sous-traitants.	☐
2	Définir les finalités et choisir la base légale	Définir les finalités	Clarifier **la ou les finalités du projet** dès la phase de conception. • S'il s'agit d'IA à usage général, faites référence au type de système développé (ex. le développement d'un modèle de langage de grande taille, d'un système de vision par ordinateur) ainsi qu'aux fonctionnalités et capacités techniquement envisageables.	☐
		Identifier les bases légales	Identifiez **la ou les bases légales** de chaque traitement (consentement, intérêt légitime, etc.).	☐
			Le cas échéant, documenter les **modalités de recueil du consentement** et en conserver la preuve (article 6.1.a du RGPD).	☐
			Le cas échéant, assurer vous d'avoir un **contrat valide** et que le traitement soit nécessaire pour remplir l'objet du contrat (article 6.1.b du RGPD).	☐
3	Evaluer le cas échéant, la validité de la base juridique de l'intérêt légitime (article 6.1.f du RGPD)	Vérifier l'existence d'un intérêt légitime	Définissez clairement **l'intérêt poursuivi.**	☐
			Vérifier que l'intérêt n'est pas en **contradiction avec d'autres obligations réglementaires** (règlement sur les services numériques, règlement sur l'intelligence artificielle, etc.).	☐
		Évaluer la nécessité du traitement	Vérifiez que le traitement des données personnelles est **nécessaire pour atteindre l'objectif défini.**	☐
			Vérifiez que des **méthodes moins intrusives** (ex. anonymisation, données synthétiques) ne peuvent atteindre les mêmes résultats.	☐
			Vérifier que les techniques algorithmiques utilisées pour le traitement des données (ex. réseau de neurones profonds convolutifs, machines à vecteurs de support SVM, etc.) est la **moins consommatrice de données personnelles** possible pour l'objectif poursuivi. Le cas échéant, documenter la nécessité de recourir à l'apprentissage machine, notamment à l'apprentissage profond.	☐
			Vérifier si des choix de conception peuvent être pris en compte dans une optique de protection des données dès la conception (apprentissage fédéré, calcul multipartite sécurisé, chiffrement homomorphe, etc.).	☐
		Mise en balance entre les intérêts en jeu	S'assurer et documenter le fait que les personnes concernées peuvent **raisonnablement s'attendre** à ce traitement.	☐
			Le cas échéant, mettre en œuvre et documenter **les garanties adaptées et suffisantes pour limiter les impacts du traitement sur les personnes concernées** (ex. prévoir l'anonymisation à bref délai des données collectées ou, à défaut, la pseudonymisation des données collectées, adopter des mesures pour limiter les risques de mémorisation, d'extraction, de régurgitation dans le cadre des IA génératives ou d'attaque des modèles ou systèmes d'IA, prévoir un droit d'opposition discrétionnaire et préalable, etc.).	☐
			Le cas échéant, mettre en œuvre des **garanties adaptées au moissonnage** (*web scraping*) des données (ex. limiter la collecte aux données librement accessibles, établir une liste de sites dont la collecte serait exclue par défaut car contenant des données particulièrement intrusives).	☐

FICHES			MESURES	
4	En cas de réutilisation des données, effectuer les tests et vérifications complémentaires	Si vous réutilisez vos propres données	Si la finalité d'entrainement de votre modèle n'était pas prévue au moment de la collecte des données, vérifier qu'elle est compatible avec l'objectif initial grâce au **test de compatibilité** (sauf si vous êtes autorisé par les personnes concernées car elles ont consenti ou par un texte ou si vous réutilisez les données dans un objectif de production de statistiques ou de recherche scientifique) : • Existe-t-il un lien entre l'objectif initial et la nouvelle finalité IA ? • Le contexte de la collecte initiale permet-il raisonnablement cette réutilisation ? • Quel est le type et la nature des données (identifiants, sensibles, etc.) ? • Quelles sont les conséquences possibles pour les personnes ? • Quelles garanties techniques et organisationnelles sont en place (pseudonymisation, etc.) ?	☐
		Si vous réutilisez des données publiquement accessibles ou acquises auprès d'un tiers (ex. : data broker)	Vérifier que vous n'êtes pas en train de réutiliser une base de données **dont la constitution était manifestement illicite :** • La source des données est-elle bien identifiée et documentée ? • La base ne résulte pas manifestement d'un crime ou d'un délit (fuite, vol, etc.) 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 une interdiction de l'exploiter ? • Les conditions de collecte des données sont-elles suffisamment documentées ? • La base ne contient pas de données sensibles ou d'infraction, ou des vérifications renforcées sont faites pour s'assurer que le traitement est licite, si c'est le cas ?	☐
5	Limiter les données traitées à celles qui sont pertinentes et nécessaires (minimisation des données)	Sélection des données strictement nécessaires	Identifier les données indispensables à l'atteinte de vos finalités et privilégier des formats moins intrusifs (ex. : tranche d'âge plutôt que date de naissance complète).	☐
			Sur le volume de données, justifier le nombre de personnes concernées, la profondeur historique et la granularité.	☐
			Justifier la nécessité de traiter des **données à caractère hautement personnel.**	☐
			Sur la typologie de données, évaluer l'usage de données réelles et de données synthétiques, pseudonymisées ou anonymisées.	☐
			Recenser **les sources de données.**	☐
		Mettre en œuvre des mesures spécifiques en cas de moissonnage (*web scraping*)	Définir, en amont, des **critères précis** de collecte.	☐
			Exclure la collecte de **certaines catégories de données** lorsqu'elles ne sont pas nécessaires, par l'intermédiaire de filtres lorsque c'est possible ou, à défaut, en excluant certains types de sites qui contiennent structurellement ces catégories de données.	☐
			Exclure de la collecte **les sites qui s'opposent clairement au moissonnage** de leur contenu, par exemple par l'utilisation des fichiers robots.txt ou la mise en place de CAPTCHA.	☐
		Précautions spécifiques pour les données sensibles	Justifier la **nécessité** de traiter des données sensibles.	☐
			Identifier **l'exception au principe d'interdiction de traitement des données sensibles** (article 9.2 du RGPD).	☐
			Prévoir des **mesures renforcées de sécurité** (pseudonymisation, etc.).	☐
			Supprimer immédiatement et, si possible, de manière automatisée les données sensibles collectées de manière incidente et résiduelle lors du moissonnage de données (*web scraping*).	☐
		Organisation de la collecte et préparation des données	Procéder à un **nettoyage** des données (incohérences, doublons, etc.).	☐
			Identifier les **données réellement pertinentes** à la tâche et supprimer les données non pertinentes pour l'apprentissage.	☐
			Appliquer des **techniques de protection dès la conception** (ex. : généralisation, randomisation, pseudonymisation, anonymisation, etc.).	☐
		Justification et validation des choix de conception	Mener une **expérimentation pilote** avec des données fictives, synthétiques ou anonymisées.	☐
			Interroger un **référent ou un comité éthique** sur les enjeux en matière d'éthique et de protection des droits et libertés des personnes.	☐
		Réévaluation continue	Mettre en place un processus de revue régulière de la pertinence des données collectées.	☐
			Mettre en place des mécanismes de suppression des données inutiles ou obsolètes.	☐
6	Définir et encadrer la durée de conservation des données	Définir une politique de conservation claire dès la conception	Définir une durée de conservation spécifique pour chaque phase du cycle de vie du projet IA (développement, maintenance, amélioration, etc.).	☐
		Conservation pendant la phase de développement	Vérifier que les données sont accessibles uniquement aux personnes habilitées pendant le développement.	☐
			Établir un processus pour archiver ou supprimer les données à la fin de la phase de développement, sauf s'il est nécessaire de conserver les données pour la maintenance ou l'améliorer du produit.	☐
		Conservation dans un but de maintenance ou d'amélioration du produit	Documenter la nécessité de conserver les données au-delà du développement notamment pour la maintenance ou l'amélioration du produit.	☐
			Vérifier que les données sont stockées sur un support cloisonné et sécurisé.	☐
			Vérifier que l'accès est strictement limité aux personnes en charge de la maintenance ou de l'amélioration du produit.	☐
			Mettre en place un plan de suppression automatique une fois l'amélioration effectuée.	☐

FICHES			MESURES	
7	Assurer la transparence des traitements		**Informer** les personnes concernées de manière claire et aisément accessible de l'ensemble des informations prévues aux articles 13 et 14 du RGPD.	☐
			Lorsque les données ne sont pas collectées directement auprès des personnes, documenter, le cas échéant, le fait qu'une information individuelle **exigerait des efforts disproportionnés** (article 14.5.b du RGPD) et rendre les informations publiquement disponibles (site web, etc.).	☐
			Si le moissonnage (*web scraping*) concerne un nombre limité de sites, **fournir une information précise** sur les sources utilisées. Si les sources sont très nombreuses, **fournir les catégories de sites sources**, au moins celles présentant le plus de risques pour les personnes.	☐
			Lorsque le modèle est en lui-même soumis au RGPD, **informer les personnes sur les données mémorisées** et fournir l'ensemble des informations imposées par les articles 13 et 14 du RGPD.	☐
8	Respecter les droits des personnes	Mise en place des procédures pour la gestion des droits	Informer les personnes du **risque de régurgitation de données** dans le cas de l'IA générative, des **mesures prises afin de limiter ces risques** et les **mécanismes de recours existants** (ex. la possibilité de signaler à l'organisme une occurrence de régurgitation)	☐
			Mettre en place une **procédure pour notifier les destinataires, notamment les utilisateurs, de toute demande** de rectification, effacement ou limitation des données sauf si cette communication se révèle impossible ou exige des efforts disproportionnés.	☐
		Gérer l'identification des personnes	Dans le cas de l'IA générative, **établir une procédure interne consistant à interroger le modèle** (par exemple à partir d'une liste de requêtes choisies) pour vérifier les données qu'il aurait pu mémoriser sur la personne grâce aux informations fournies.	☐
			S'il n'est pas possible d'identifier une personne au sein du modèle mais qu'elle est identifiée dans une base de données d'entraînement, l'informer du risque de mémorisation du modèle.	☐
			S'il n'est **pas possible d'identifier une personne** dans la base d'apprentissage ou au sein du modèle, l'en informer.	☐
			Informer la personne **des éléments complémentaires** qu'il est possible de fournir pour aider à leur identification (ex : pseudonyme, échantillon de leurs données).	☐
			Mettre en place un processus pour **supprimer ces données complémentaires** après le traitement de la demande.	☐
		Choisir une solution technique pour assurer le respect des droits sur les modèles	En principe, prévoir **un processus de réentraînement** du modèle. • Le réentraînement peut être périodique pour limiter les coûts et satisfaire plusieurs demandes d'exercices de droits en même temps.	☐
			Fournir une **version actualisée du modèle** à ses utilisateurs, éventuellement en leur imposant par voie contractuelle de n'utiliser qu'une version régulièrement mise à jour.	☐
			Le cas échéant, documenter le fait que le **réentrainement du modèle s'avère disproportionné** (temporairement ou définitivement).	☐
			Si le réentrainement est disproportionné, **mettre en place des filtres ou autres mesures robustes** sur les sorties du système IA.	☐
			Le cas échéant, **préférer des règles générales** prévenant la génération de données personnelles à une simple « liste noire » de personnes ayant exercé leurs droits.	☐
9	Assurer la conformité de l'annotation des données		Vérifier que les annotations ne contiennent que les informations nécessaires pour atteindre la finalité et qu'elles sont objectives.	☐
			Mettre en œuvre des **revues régulières** pour s'assurer de la pertinence continue des étiquettes.	☐
			Mettre en place une **procédure continue de vérification** afin de contrôler la qualité de l'annotation : définir un protocole d'annotation, en application des principes d'exactitude et de minimisation et impliquer un référent ou un comité éthique (bonne pratique).	☐
			Informer les personnes concernées de la phase d'annotation des données.	☐
			S'assurer que les **procédures internes de gestion des droits et les modalités d'exercice des droits** incluent l'annotation (droit d'accès, de rectification, à l'effacement, à la limitation, à la portabilité, d'opposition).	☐
			Vérifiez, le cas échéant, que les données sensibles sont traitées conformément à une exception au principe d'interdiction des données (article 9.2 du RGPD).	☐
			Mettre en œuvre des **mesures particulières** au regard du risque accru pour les personnes : annoter selon des critères objectifs et factuels, limiter l'annotation au contexte des données, renforcer l'étape de vérification des annotations, augmenter la sécurité des données annotées (ex. en réalisant le traitement d'annotation en interne, en traitant les données localement et en garantissant leur sécurité par le chiffrement, la journalisation, et par des restrictions d'accès plus fortes) et s'interroger sur le risque de régurgitation et d'inférence des données sensibles sur les modèles entraînés à partir de celles-ci.	☐
			Former les personnes en charge de l'annotation aux principes relatifs à la protection des données.	☐

FICHES			MESURES	
10	Assurer la sécurité des données		Vérifiez que les mesures de sécurités portant sur les données d'entraînement soient mises suffisantes et adaptées (voir le guide de la sécurité des données personnelles).	☐
			Vérifiez que les mesures de sécurités portant sur le développement du système soient suffisantes et adaptées. En particulier, utiliser des outils de développement, des librairies, et, le cas échéant, des modèles pré-entraînés vérifiés.	☐
			Vérifiez que les mesures visant à encadrer le fonctionnement du système soient suffisantes et adaptées. Par exemple : favoriser des formats d'importation et de sauvegarde vérifiés comme *safetensors*, contrôler les sorties du système d'IA en utilisant des filtres, mettre en œuvre des techniques de tatouage numérique (*watermarking*).	☐
			De façon générale, gérer les habilitations à accéder aux données, tracer ces accès et analyser les traces.	☐
			Mettre en œuvre et suivre un plan d'action pour assurer que les exigences de sécurité sont bien satisfaites.	☐
11	Analyser les risques et réaliser une analyse d'impact sur la protection des données (AIPD)		Mener une AIPD si le traitement d'entrainement du modèle présente des **risques élevés** en fonction des critères dégagés par le comité européen de la protection des données (usage innovant, large échelle, données sensibles, personnes vulnérables, etc.).	☐
			Inclure les **risques spécifiques IA** (discrimination automatisée causée par un biais du système introduit lors du développement, risque de produire du contenu fictif sur une personne réelle, risques liés aux attaques connues spécifiques aux systèmes d'IA, etc.) et prendre les mesures adéquates.	☐