

## Fiches pratiques sur la constitution de bases de données d'apprentissage

### Formulaire de la consultation publique

---

La CNIL invite les acteurs à regrouper, si possible, leurs commentaires au sein d'une contribution unique en mutualisant les différents retours internes ou en se rapprochant de leur fédération.

Les contributions transmises à la CNIL dans ce cadre ne sont pas publiques et leur confidentialité sera assurée. Notez toutefois que celles-ci pourront être résumées et restituées dans le cadre de l'éventuelle synthèse de la consultation publique publiée par la CNIL à l'issue de celle-ci.

En effet, toutes les contributions reçues par la CNIL peuvent faire l'objet d'une demande d'accès en tant que documents administratifs (code des relations entre le public et l'administration).

Dans votre contribution, signalez tout élément protégé par des [droits de propriété littéraire ou artistique](#) (précisez, dans ce cas, si vous en permettez ou non la communication), ou [par le secret des affaires](#).

A noter que la CNIL n'est pas tenue de suivre votre évaluation sur ce qui est protégé ou non.

Pour faire parvenir à la CNIL ce formulaire complété, merci de l'envoyer par courriel à l'adresse suivante : [ia@cnil.fr](mailto:ia@cnil.fr)

La CNIL traite les données recueillies à partir de ce formulaire afin d'analyser les observations des participants en vue d'adopter le guide concerné. Les données sont également collectées pour réaliser des statistiques relatives aux contributions et, si nécessaire, pour contacter les contributeurs afin d'approfondir les échanges ou les tenir informés des suites de la consultation. La base légale du traitement est l'exercice de l'autorité publique. Les données sont communiquées aux services de la CNIL en charge de l'analyse des réponses fournies.

**Vous pouvez accéder à vos données, vous opposer à leur traitement, demander leur rectification ou leur effacement. Vous pouvez également exercer votre droit à la limitation du traitement de vos données.**

[En savoir plus sur la gestion de vos données et vos droits.](#)

#### Informations personnelles

Nom : GOUBET

Prénom : Jocelyn

Fonction (**obligatoire**) : Direction droit économique

La CNIL traite les données recueillies à partir de ce formulaire afin d'analyser les observations des participants en vue d'adopter le guide concerné. Les données sont également collectées pour réaliser des statistiques relatives aux contributions et, si nécessaire, pour contacter les contributeurs afin d'approfondir les échanges ou les tenir informés des suites de la consultation. La base légale du traitement est l'exercice de l'autorité publique. Les données sont communiquées aux services de la CNIL en charge de l'analyse des réponses fournies.

Vous pouvez accéder à vos données, vous opposer à leur traitement, demander leur rectification ou leur effacement. Vous pouvez également exercer votre droit à la limitation du traitement de vos données.

[En savoir plus sur la gestion de vos données et vos droits.](#)

Adresse courriel (**obligatoire**) : j.goubet@afep.com

Nom de l'organisme : Association française des entreprises privées

Type d'organisme (**obligatoire**) : Association représentative de la société civile

Fiches pratiques sur la constitution de bases de données d'apprentissage	Commentaires et suggestions
Introduction : Périmètre	<p>Les entreprises de l'Afep accueillent favorablement l'initiative de la CNIL de publier des fiches pratiques sur la constitution de bases de données d'apprentissage. Elles attirent l'attention de la CNIL sur un besoin de « parallélisme des formes » : une fois finalisé ce travail côté développement des outils d'intelligence artificielle, il leur apparaîtrait opportun de développer des fiches pratiques pour les entreprises utilisatrices d'outils d'intelligence artificielle commercialisés par des développeurs.</p> <p>De façon générale, l'Afep relève que l'application du RGPD à l'IA demande une interaction active de la CNIL avec les acteurs publics et privés de l'IA par le biais d'un dialogue ouvert et continu, notamment pour prendre en compte l'évolution technologique rapide dans ce secteur.</p> <p>De plus, les entreprises de l'Afep rappellent la nécessité d'avoir une approche souple et pragmatique de l'IA de manière à favoriser les opportunités d'innovation dans ce domaine. Il est donc important que ces fiches pratiques ne constituent que des recommandations de la CNIL qui ne soient pas exclusives d'autres interprétations sous réserve du respect du RGPD. Elles souhaiteraient, en conséquence, que la typologie de ce document soit clairement mentionnée.</p> <p>L'Afep souhaite, en outre, souligner <b>les risques de décalage entre les décisions européennes et nationales</b>. En effet, cette consultation publique se déroule alors que les débats à Bruxelles sur l'adoption de l'AI Act ne sont pas clos. En conséquence, <b>les fiches de la CNIL devront nécessairement être adaptées après l'adoption définitive de l'AI Act pour s'assurer de leur cohérence avec le texte final du projet de règlement</b>, et notamment avec les systèmes d'IA à haut risque identifiés par ce dernier. Elle préconise que la CNIL s'inscrive dans le cadre européen lorsque celui-ci sera stabilisé afin de proposer un cadre d'application cohérent aux acteurs économiques leur évitant des surcharges administratives potentiellement inutiles si ces recommandations devaient finalement se révéler contraires au cadre européen.</p> <p>A titre d'illustration, concernant le <b>périmètre d'application</b> des fiches pratiques publiées par la CNIL, l'Afep relève que la CNIL distingue deux systèmes d'IA (i) les systèmes d'apprentissage automatique, (ii) les systèmes se fondant sur la logique et les connaissances. <b>Une telle vision ne correspond pas à la définition de l'IA telle que retenue dans le projet d'AI Act, ni à celle communément admise</b>. Dès lors, il semble nécessaire de revoir la définition de l'IA pour écarter l'application de son régime légal à certains systèmes. <b>Les entreprises de l'Afep considèrent que les algorithmes ne constituent pas un système d'IA, et en particulier les arbres de décision</b>. A ce titre, la CNIL pourrait envisager de compléter sa définition de l'IA a contrario, en excluant les systèmes qui n'en relèveraient pas.</p> <p>Enfin, les grandes entreprises déplorent les <b>distorsions juridiques existant déjà au sein des Etats membres</b>. Il en est ainsi de l'interdiction temporaire de l'utilisation de</p>

La CNIL traite les données recueillies à partir de ce formulaire afin d'analyser les observations des participants en vue d'adopter le guide concerné. Les données sont également collectées pour réaliser des statistiques relatives aux contributions et, si nécessaire, pour contacter les contributeurs afin d'approfondir les échanges ou les tenir informés des suites de la consultation. La base légale du traitement est l'exercice de l'autorité publique. Les données sont communiquées aux services de la CNIL en charge de l'analyse des réponses fournies.

**[Vous pouvez accéder à vos données, vous opposer à leur traitement, demander leur rectification ou leur effacement. Vous pouvez également exercer votre droit à la limitation du traitement de vos données.](#)**

[En savoir plus sur la gestion de vos données et vos droits.](#)

	<p>ChatGPT en Italie, alors que l'IA générative ne figure pas à ce stade sur la liste des IA à haut risque de l'AI Act. Cette interdiction (désormais levée) constitue une véritable incertitude juridique dans la coordination entre l'AI Act et le RGPD, qui doit faire l'objet d'échanges et d'éclaircissements au niveau européen afin d'assurer la cohérence entre l'AI Act et le RGPD.</p>
Fiche 1 : Déterminer le régime juridique applicable	<p>Les entreprises de l'Afep souhaitent attirer l'attention de la CNIL sur la complexité de la chaîne de valeur de l'IA qui peut impliquer de nombreux acteurs, et qui induit un dialogue et une répartition des responsabilités entre le développeur du modèle d'IA et l'utilisateur de ce modèle ou client du développeur.</p>
Fiche 2 : Définir une finalité	<p>Les entreprises de l'Afep relèvent que, pour les outils d'IA reposant sur un arbre décisionnel, l'usage opérationnel en phase de développement et en phase de déploiement poursuit une seule et même finalité. Ce ne serait pas le cas pour les outils d'IA reposant sur le big data. <b>Les illustrations développées par la CNIL pour illustrer son propos pourraient être utilement complétées avec des modèles d'IA reposant sur le big data.</b></p> <p>S'agissant plus particulièrement de la définition de la finalité du modèle d'IA, les entreprises relèvent que bien souvent cette dernière n'est pas définie par le développeur du modèle, surtout s'il s'agit d'un modèle d'IA générale, mais par l'utilisateur – le client du développeur.</p>
Fiche 3 : Déterminer la qualification juridique des fournisseurs de systèmes d'IA	<p>Comme évoqué ci-dessus, la qualification des fournisseurs ou développeurs de solutions d'IA variera en fonction de son implication dans la détermination des objectifs et des moyens du traitement.</p> <p>Les entreprises de l'Afep relèvent que la qualification de responsable de traitement, ou de coresponsable, pour les développeurs du système d'IA va changer le paradigme des négociations commerciales entre ces fournisseurs de solutions d'IA et les entreprises utilisatrices.</p> <p>Elles s'interrogent néanmoins sur la nécessité de rendre obligatoire l'analyse d'impact pour le fournisseur qu'il soit responsable de traitement ou sous-traitant. En effet, la bonne utilisation du système d'IA repose sur la connaissance du fonctionnement du système d'IA et de sa construction.</p>
Fiche 4 : Assurer que le traitement est licite	<p>Constituer une base de données sur la base légale du consentement peut conduire à la constitution d'un modèle biaisé, dans la mesure où certaines catégories socio-professionnelles ont plutôt tendance à consentir ou à refuser de consentir à la réutilisation de leurs données. Un tel fondement conduirait à la constitution de biais discriminatoires. Dès lors, <b>la base légale du consentement doit être maniée avec précaution.</b></p> <p>Sur ce point, les entreprises comprennent, du dernier paragraphe relatif à la base légale du consentement, que la CNIL ferait référence aux données librement accessibles (open data). Si c'est bien le cas, elles invitent la CNIL à y faire expressément référence dans un souci de clarté. A ce titre, elles attirent l'attention de la CNIL sur le fait que le développement de l'IA repose en grande partie sur les publications librement accessibles. Quel que soit le mode d'accès aux données (open source ou web scraping), la représentativité des données collectées doit être assurée.</p> <p>Au regard de l'ensemble de ces éléments, et compte-tenu des risques limités pour les personnes en phase d'entraînement du modèle d'IA, <b>les entreprises de l'Afep</b></p>

### 3

La CNIL traite les données recueillies à partir de ce formulaire afin d'analyser les observations des participants en vue d'adopter le guide concerné. Les données sont également collectées pour réaliser des statistiques relatives aux contributions et, si nécessaire, pour contacter les contributeurs afin d'approfondir les échanges ou les tenir informés des suites de la consultation. La base légale du traitement est l'exercice de l'autorité publique. Les données sont communiquées aux services de la CNIL en charge de l'analyse des réponses fournies.

Vous pouvez accéder à vos données, vous opposer à leur traitement, demander leur rectification ou leur effacement. Vous pouvez également exercer votre droit à la limitation du traitement de vos données.

[En savoir plus sur la gestion de vos données et vos droits.](#)

	<p><b>considèrent que la base légale de l'intérêt légitime est la plus adaptée pour le développement d'outils d'IA, et devrait être encouragée par la CNIL afin de soutenir l'innovation.</b> Les entreprises attendent donc avec intérêt la fiche spécifique qui sera publiée ultérieurement par la CNIL sur les conditions à respecter pour mobiliser la base légale de l'intérêt légitime.</p>
Fiche 5 : Réaliser une analyse d'impact si nécessaire	<p>Les entreprises de l'Afep accueillent favorablement l'appréciation menée par la CNIL du caractère innovant d'un système d'IA. <b>Elles partagent la position selon laquelle l'utilisation d'un système d'IA ne relève pas systématiquement d'un usage innovant ou de l'application de nouvelles solutions technologiques ou organisationnelles.</b></p> <p>De même, elles considèrent que l'utilisation d'un système d'IA repose nécessairement sur l'utilisation d'une très grande quantité de données, qui conditionne la représentativité des résultats. Pour autant, elles estiment que cette méthode ne peut pas systématiquement constituer un traitement à grande échelle de données à caractère personnel. Sur ce point, <b>des précisions de la CNIL sur les critères d'appréciation de l'existence d'un traitement à grande échelle de manière générale, et notamment dans le cadre d'un système d'IA, seraient appréciées</b> (tels que des seuils).</p> <p>Enfin, les entreprises de l'Afep s'interrogent sur la coordination entre l'analyse d'impact sur la protection des données évoquée par la CNIL, et l'autoévaluation de conformité (« conformity self assessment ») du fournisseur prévu par l'AI Act. <b>La cohérence entre le RGPD et l'AI Act doit, là encore, être assurée.</b></p>
Fiche 6 : Tenir compte de la protection des données dans les choix de conception du système	<p>Les entreprises de l'Afep accueillent favorablement les bonnes pratiques avancées par la CNIL (i) la conduite d'une étude pilote à petite échelle, comme (ii) l'association d'un comité éthique pour assurer la protection des droits et libertés des personnes dans le cadre du développement d'un système d'IA.</p> <p>Elles partagent également l'idée d'un besoin général de sensibilisation et d'acculturation de la protection des données personnelles au sein des entreprises de toute taille.</p>
Fiche 7 : Tenir compte de la protection des données dans la collecte et la gestion des données	<p>Les entreprises de l'Afep perçoivent une approche plus pragmatique de la CNIL dans l'appréciation de la durée de conservation des données, qui selon elle peut être utile pour les opérations de maintenance ou l'amélioration du système d'IA. <b>Ces hypothèses pourraient être utilement complétées avec la possibilité d'utiliser les mêmes données pour la création et le test d'un nouveau système d'IA, distinct de celui existant, afin de comparer les deux systèmes et conserver le plus performant.</b> Dans une telle hypothèse, l'ensemble des données utilisées initialement devrait être conservé. Une telle solution serait favorable à l'innovation.</p> <p>Sur le point particulier de la détermination de la durée de conservation, l'Afep relève que la CNIL considère que « <i>une durée de conservation de plusieurs années à partir de la date de collecte devra être justifiée</i> ». A ce titre, <b>des précisions de la part de la CNIL sur l'appréciation de la durée seraient appréciées</b> (telles que des durées maximales pour la phase de développement, la maintenance ou l'amélioration du système d'IA).</p> <p>Enfin, l'Afep recommande que la CNIL prenne davantage en compte les techniques de protection de la vie privée et de <b>réduction du risque telles que la pseudonymisation</b> afin de faciliter la constitution des bases de données d'entraînement tout en réduisant le risque pour les personnes.</p>

La CNIL traite les données recueillies à partir de ce formulaire afin d'analyser les observations des participants en vue d'adopter le guide concerné. Les données sont également collectées pour réaliser des statistiques relatives aux contributions et, si nécessaire, pour contacter les contributeurs afin d'approfondir les échanges ou les tenir informés des suites de la consultation. La base légale du traitement est l'exercice de l'autorité publique. Les données sont communiquées aux services de la CNIL en charge de l'analyse des réponses fournies.

**[Vous pouvez accéder à vos données, vous opposer à leur traitement, demander leur rectification ou leur effacement. Vous pouvez également exercer votre droit à la limitation du traitement de vos données.](#)**

[En savoir plus sur la gestion de vos données et vos droits.](#)

Annexe : Fiche descriptive du jeu de données

Les entreprises de l'Afep s'interrogent sur l'opportunité d'intégrer à la fiche descriptive une intervention du comité éthique, visé à la fiche 6, afin que les entreprises encadrent l'utilisation des données par les équipes de développement des systèmes d'IA.

Enfin, les entreprises attendent des précisions sur l'usage attendu par la CNIL de cette fiche par les entreprises, et son caractère obligatoire.

## 5

La CNIL traite les données recueillies à partir de ce formulaire afin d'analyser les observations des participants en vue d'adopter le guide concerné. Les données sont également collectées pour réaliser des statistiques relatives aux contributions et, si nécessaire, pour contacter les contributeurs afin d'approfondir les échanges ou les tenir informés des suites de la consultation. La base légale du traitement est l'exercice de l'autorité publique. Les données sont communiquées aux services de la CNIL en charge de l'analyse des réponses fournies.

Vous pouvez accéder à vos données, vous opposer à leur traitement, demander leur rectification ou leur effacement. Vous pouvez également exercer votre droit à la limitation du traitement de vos données.

[En savoir plus sur la gestion de vos données et vos droits.](#)