Consultation publique – fiches pratiques sur la constitution de bases de données pour la conception de systèmes d'IA

Synthèse des contributions

Février 2024

La CNIL a lancé, le 11 octobre 2023, une consultation publique sur la constitution de bases de données d'apprentissage des systèmes d'intelligence artificielle.

Les contributions ont nourri les travaux sur **ce premier lot de huit fiches pratiques** en vue de leur publication définitive sur le site.

CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS

La synthèse en chiffres

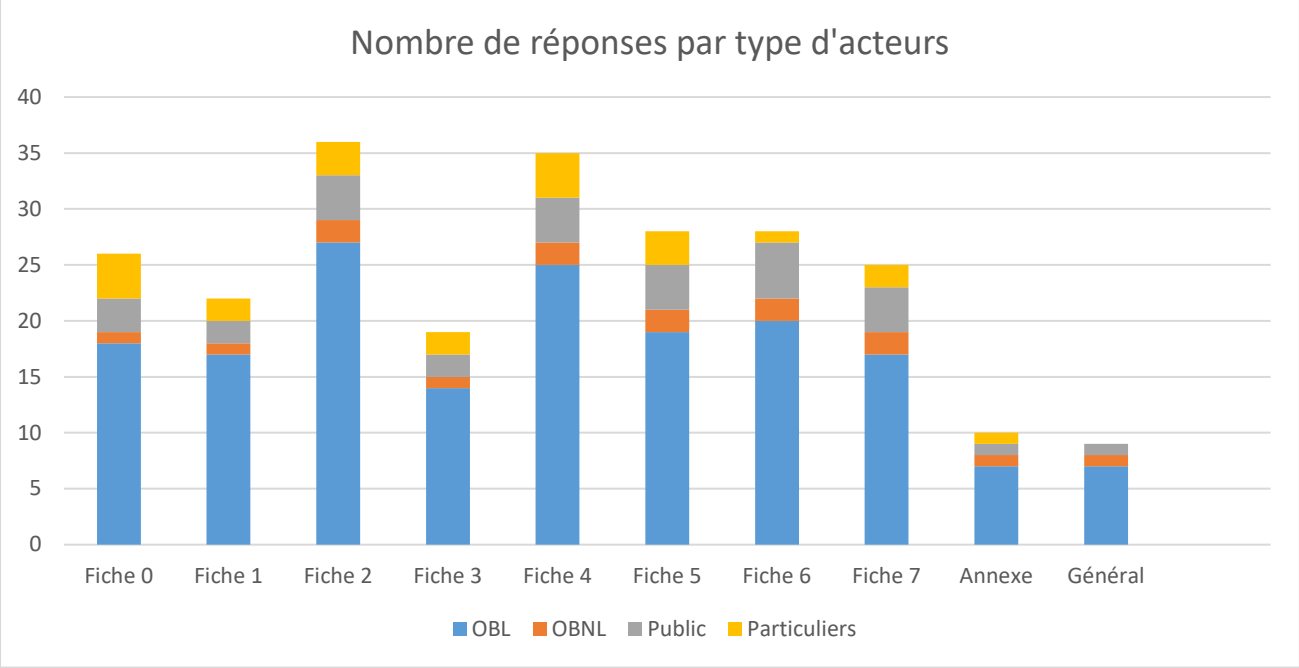
La CNIL a publié, le 11 octobre 2023, une première série de huit fiches pratiques sur la constitution de bases de données d'apprentissage des systèmes d'IA, afin d'accompagner les acteurs du secteur dans leurs démarches de mise en conformité avec la législation sur la protection des données personnelles et de répondre à leurs principales interrogations.

À l'issue de la consultation publique, le 15 décembre 2023, **la CNIL a reçu 42 contributions**, par des contributeurs représentatifs de différents secteurs :

- des organismes à but lucratif de différents secteurs (IA, finance, santé, aéronautique, opérateurs de plateformes en ligne, publicité en ligne, jeux vidéo, audiovisuel, etc.) :
  - 10 organisations représentantes des professionnels ;
  - 18 entreprises privées ;
  - 1 cabinet de conseil

- des organismes à but non lucratif :
  - 1 association représentative de la société civile ;
  - 2 instituts de recherche et 1 groupement d'instituts de recherche publique ;
  - 1 *think tank* indépendant ;
  - 3 syndicats professionnels de salariés.

- 4 particuliers ;
- 3 établissements publics.



Ces contributions ont permis à la CNIL :

- de faire évoluer ses projets de fiches pratiques en apportant des clarifications supplémentaires et consolider ses analyses au regard des observations formulées dans les contributions ;
- d'apporter des réponses, dans la synthèse ci-dessous, aux préoccupations les plus fréquemment partagées par les contributeurs.

Contributions d'ordre général sur l'ensemble des fiches

Sur la valeur juridique des fiches pratiques

Synthèse des contributions

Plusieurs contributeurs se sont interrogés sur la valeur juridique des fiches ainsi que le caractère contraignant ou non des préconisations qui y sont formulées.

Réponse de la CNIL

La fiche introductive a été complété pour souligner que ces fiches **rappellent les obligations posées par la règlementation** (par exemple, « le responsable du traitement *doit* (…) ») **et formulent des recommandations pour s'y conformer** (par exemple, « *peut* permettre »). Ces recommandations ne sont toutefois pas contraignantes : les responsables de traitement peuvent s'en écarter, à condition de pouvoir justifier leur choix et sous leur responsabilité.

Certaines recommandations sont également formulées à titre de bonnes pratiques (par exemple, « il est recommandé, *à titre de bonnes pratiques* »). Le suivi de ces bonnes pratiques va au-delà de ce que la règlementation impose.

Elle effectue également une distinction plus nette dans les fiches, entre ce qui relève d'un rappel des obligations applicables, de recommandations de mise en conformité ou simplement de bonnes pratiques.

Sur l'articulation avec la proposition de règlement européen sur l'intelligence artificielle

Synthèse des contributions

De nombreux contributeurs jugent qu'il serait utile que les fiches se réfèrent plus explicitement à la proposition de règlement européen sur l'IA (RIA) afin de clarifier leur articulation, au motif que certaines dispositions ou orientations de ce texte seraient déjà stabilisées.

Les problèmes d'articulation identifiés par les acteurs concernent notamment les sujets suivants : la définition de système d'IA, la qualification des acteurs, la définition des risques, les obligations en termes de documentation ainsi que le traitement de données sensibles pour la détection et correction des biais.

Réponse de la CNIL

Dans l'attente de l'adoption définitive du RIA, la CNIL a rappelé, **dans la fiche introductive, que le RGPD s'appliquait déjà indépendamment du RIA** (qui n'a pas le même champ d'application). Elle précise, lorsque cela est pertinent, **certains points d'articulation** (en particulier sur la qualification des acteurs et sur la nécessité d'une AIPD pour les systèmes identifiés comme à haut risque dans le RIA).

Sur le découpage par phases : du développement au déploiement

Synthèse des contributions

Certaines contributions interrogent le choix de découpage proposé entre les phases de développement et de déploiement.

Réponse de la CNIL

Ce découpage permet une séparation chronologique entre les différentes phases de traitement de données personnelles, qui correspondent à des traitements distincts (possiblement mis en œuvre par des organismes différents) et ainsi des ensembles différents de personnes concernées.

La CNIL est consciente qu'une part importante des développements actuellement réalisés en IA peut consister à les mettre en œuvre directement ou les adapter à des cas d'usage spécifiques (par apprentissage par transfert ou fine-tuning par exemple). Ces pratiques complexifient le découpage en deux phases proposées par la CNIL. Toutefois, la CNIL ne peut pas exclure le cas où un fournisseur conçoit un modèle entièrement, que cela soit

afin de développer un système directement utilisé en phase de déploiement ou d'un modèle général (ou de fondation) qui sera adapté par la suite par le fournisseur ou un tiers. Cette étape est cruciale à encadrer car c'est celle qui fait intervenir la plus grande quantité de données (puisqu'elle consiste à initialiser le modèle) présentant donc des risques pour les personnes concernées.

Fiche introductive - Le périmètre des fiches pratiques sur l'IA

Sur le choix de la définition d'IA

Synthèse des contributions

Le périmètre des systèmes concernés entend englober tous les outils utilisant des données personnelles - donc soumis au RGPD - et utilisant des techniques assimilables à de l'IA. Les contributeurs demandent si la définition choisie correspond à celle du RIA.

Réponse de la CNIL

La CNIL propose de clarifier son périmètre en utilisant explicitement la définition du Parlement Européen telle qu'écrite dans la proposition de règlement IA. Une fois le RIA adopté, une mise à jour sera réalisée avec la définition finalement retenue.

Sur l'inclusion de l'ajustement (fine tuning) dans le périmètre

Question soulevée

Certains contributeurs demandent à inclure l'ajustement (*fine-tuning*) dans le périmètre des fiches, voire à découper les phases de traitement en fonction de l'utilisation de modèles préexistants.

Réponse de la CNIL

La CNIL a inclus l'ajustement ou encore l'apprentissage par transfert dans son périmètre : ces traitements sont assimilables à des traitements en phase de développement.

Si le *fine-tuning* d'un modèle existant est un cas courant d'utilisation des modèles d'IA, **il ne remet toutefois pas en question le découpage en deux phases de développement et déploiement d'un système d'IA.** En effet, le *fine-tuning*, au même titre que le développement d'un modèle à partir de zéro, nécessite la réalisation de choix de conception, la collecte d'une base de données, ainsi qu'une phase d'apprentissage. L'organisme réalisant le *fine-tuning* aura un rôle de fournisseur de système pour le modèle qu'il aura conçu et sera tenu par les mêmes obligations.

Une différence est à noter en ce qui concerne le modèle préexistant utilisé comme base pour le *fine-tuning* : ce modèle sera issu d'une phase de développement antérieure à l'issue de laquelle il aura été transmis ou mis à disposition (par exemple en open source). En l'absence de spécificités au stade du développement, les questionnements relatifs à la transmission ou à la mise à disposition du modèle pré-entraîné ne seront traitées qu'à un stade ultérieur.

Sur les risques de présence des données personnelles

Synthèse des contributions

Certains contributeurs demandent des clarifications sur la manière d'identifier les risques de présence de données personnelles dans un jeu de données.

Réponse de la CNIL

La présence de données personnelles constitue un risque pour les personnes concernées qui n'est pas assimilable à la notion de risque tel que défini par le RIA. Si les indices permettant d'évaluer la présence de données personnelles sont nombreux, et qu'il n'est pas possible d'en faire la liste exhaustive, des précisions ont toutefois été apportées sur la distinction entre les bases de données contenant possiblement des données à caractère personnel, et celles en contenant très probablement. La CNIL reconnait ainsi que, les volumes

mobilisés pour le développement de systèmes d'IA étant potentiellement très important et les données utilisées complexes (texte non structuré, audio, vidéo, etc.), il est dans de nombreux cas très difficile d'être certain qu'aucune donnée personnelle ne figure dans un jeu de données.

Sur la distinction entre « modèle d'IA » et « système d'IA »

Synthèse des contributions

Plusieurs contributions ont souligné la nécessité de distinguer les notions de « modèle » et de « système ».

Réponse de la CNIL

Cette distinction, qui recoupe celle entre « modèles de fondation » et « systèmes d'IA à usage général » dégagée notamment au cours des négociations sur le RIA, a été reprise lorsque cela était pertinent. Ainsi, la CNIL considère le modèle comme le produit de l'entrainement réalisé à partir des données d'apprentissage et le système comme l'intégration logicielle du modèle qui pourra ensuite faire l'objet d'un déploiement ou d'un ajustement (*fine-tuning*).

La CNIL considère toutefois que cette distinction n'est pas structurante au regard de la conformité au RGPD des traitements de données personnelles en phase de développement.

Fiche 1 – Déterminer le régime juridique applicable

Synthèse des contributions

De nombreuses contributions témoignent d'une **incompréhension de l'objectif et du périmètre de la fiche**, notamment au regard des éléments suivants :

- l'absence de définition de « régime juridique » ;
- l'absence d'indications quant à l'applicabilité du RGPD ou de la directive « police-justice » ;
- une confusion entre les différents cas présentés.

Réponse de la CNIL

L'objectif de cette fiche est d'aider le responsable du traitement à déterminer quelle est la réglementation applicable en matière de protection des données (RGPD, directive « police-justice » ou traitements intéressant la défense nationale ou la sûreté de l'État) lors du développement de systèmes d'IA.

En réponse à ces observations, **la CNIL a clarifié la notion de « régime juridique » et intégré des illustrations supplémentaires s'agissant des systèmes d'IA à usage général.**

Fiche 2 – Définir une finalité

Sur les critères de définition de la finalité pour les systèmes d'IA à usage général

Synthèse des contributions

Plusieurs contributeurs se sont interrogés sur la pertinence des critères de définition de la finalité pour les systèmes d'IA à usage général :

- une partie des contributeurs considèrent que ces critères sont trop flexibles et qu'ils ne permettent pas de répondre effectivement au critère de précision de la finalité, ce qui pourrait porter préjudice au respect des principes qui en découlent (notamment la minimisation des données et la limitation des finalités) ;
- une autre partie des contributeurs considèrent que ces critères sont trop prescriptifs, compte tenu notamment de l'impossibilité de prévoir les usages des systèmes d'IA dès la phase de développement.

Réponse de la CNIL

La CNIL a bien identifié les complexités liées à la définition d'une finalité suffisamment précise.

En réponse, elle rappelle, dans la version définitive de la fiche 2, des critères permettant de prendre en compte les difficultés pour le responsable du traitement de 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.

Pour assurer la sécurité juridique des acteurs, **la CNIL a également clarifié la distinction entre les préconisations qui relèvent ou non de l'ordre de la bonne pratique**.

Sur la réutilisation des données de recherche scientifique à d'autres fins

Synthèse des contributions

Plusieurs contributeurs ont appelé à une clarification de la notion de recherche scientifique ainsi que **des conditions dans lesquelles il est envisageable de réutiliser les données d'une recherche scientifique à d'autres fins, notamment commerciales.**

Réponse de la CNIL

La réutilisation de données anonymisées ou de modèle n'ayant pas mémorisé de données personnelles, même en dehors d'une finalité de recherche, ne pose pas de difficultés.

En revanche, la CNIL rappelle, dans la version définitive de la fiche 4, le principe selon lequel **la réutilisation de données personnelles initialement traitées à des fins de recherche pour des finalités hors recherche n'est légale que pour des finalités jugées compatibles**. Ce nouveau traitement devra se conformer à l'ensemble des principes posés par le RGPD (information des personnes, respect des droits, identification d'une nouvelle exception pour le traitement de données sensibles le cas échéant, etc.).

Fiche 3 – Déterminer la qualification juridique des fournisseurs de systèmes d'IA

Sur l'articulation avec la proposition de règlement sur l'IA

Synthèse des contributions

Plusieurs contributions ont appelé à une clarification plus explicite des rôles et qualifications au sens du RGPD en articulation avec la proposition de règlement sur l'IA encore en cours d'élaboration.

Réponse de la CNIL

Bien que le RIA n'ait pas encore été adopté, des développements ont été rajoutés pour clarifier les qualifications qu'un « fournisseur de système d'IA » pourrait endosser au sens du RGPD. Par ailleurs, un exemple illustre désormais le cas particulier de l'ajustement d'un modèle (*fine-tuning*) qui aurait mémorisé des données personnelles.

En effet, **la diffusion, la conservation ou encore la maintenance d'un tel modèle sont considérés comme des traitements de données personnelles**. Cela entraîne des conséquences en termes de responsabilités, notamment pour le fournisseur de tels modèles d'IA.

Sur la valeur des critères et des exemples donnés

Synthèse des contributions

Certaines contributions ont estimé que les critères et les exemples de qualification étaient trop prescriptifs.

Réponse de la CNIL

L**a qualification juridique des fournisseurs de systèmes d'IA doit se faire au cas par cas**. Cette fiche n'a pas vocation à créer de nouveaux critères, mais bien à éclairer les indices permettant de mener cette analyse, déjà évoqués dans les lignes directrices 07/2020 du Comité européen de protection des données concernant les notions de responsable du traitement et de sous-traitant.

Fiche 4 – Assurer que le traitement est licite

La base légale de l'obligation légale

Synthèse des contributions

Plusieurs contributions s'interrogent sur la possibilité de mobiliser la base légale de l'obligation légale (article 6.1.c du RGPD) pour le développement et l'utilisation de systèmes d'IA pour certaines finalités. Plusieurs exemples sont donnés (lutte contre le blanchiment et financement du terrorisme (LCB-FT) ou encore de modération de contenus).

Réponse de la CNIL

La CNIL a complété la fiche pour préciser les limites de l'obligation légale **comme base légale pertinente pour le développement de modèle.** Cela n'interdit toutefois pas de recourir à un système d'IA déjà développé pour répondre à une obligation légale lorsque les conditions de mobilisation de la base juridique sont remplies.

Sur le traitement de données sensibles de manière incidente

Synthèse des contributions

Plusieurs contributeurs s'interrogent sur la **présence incidente de données sensibles dans les bases de données d'apprentissage**. Ils soulignent la difficulté voire l'impossibilité, dans certains cas, de garantir l'absence de données sensibles, notamment lorsque la base de données est constituée par la collecte de données en ligne.

Réponse de la CNIL

Le traitement de données sensibles n'est possible, en principe, que sur le fondement de l'une des exceptions limitativement listées à l'article 9.1 du RGPD.

À noter : une précision est ajoutée sur les conditions dégagées par la jurisprudence récente (CJUE, 4 juillet 2023, affaire C-252/21) pour mobiliser l'exception relative à la collecte des données « manifestement rendues publiques ».

La CNIL a précisé les règles applicables en cas de collecte incidente de données sensibles lors de l'utilisation d'outils de moissonnage (*web scraping*) pour la constitution de bases de données d'IA :

- Le responsable du traitement est tenu de mettre en œuvre toutes les mesures permettant d'exclure automatiquement la collecte de données sensibles non pertinentes. Il doit notamment appliquer des filtres empêchant la collecte de certaines catégories de données et/ou s'abstenir de réaliser la collecte sur certains sites comportant des données sensibles par nature.
- Si, malgré les mesures prises, l'organisme traite de manière incidente et résiduelle des données sensibles qu'il n'avait pas cherché à collecter, cela n'est pas considéré comme illégal. 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.

Sur les vérifications supplémentaires en cas de réutilisation de bases de données librement accessibles

Synthèse des contributions

Certains contributeurs considèrent que les préconisations relatives aux vérifications à opérer pour s'assurer qu'une base de données n'a pas fait l'objet d'une décision de justice interdisant sa réutilisation sont trop contraignantes. Plusieurs contributeurs demandent également de clarifier la question relative à la preuve de

l'absence de « doutes flagrants » sur la licéité de la base de données et de préciser les procédures détaillées pour la vérification de la légalité des bases de données réutilisées.

Réponse de la CNIL

L'illicéité manifeste doit s'apprécier au cas par cas. De ce fait, la CNIL considère qu'il appartiendra au responsable du traitement d'effectuer les vérifications nécessaires selon le cas de figure.

Fiche 5 – Réaliser une analyse d'impact si nécessaire

Sur l'articulation avec la proposition de règlement européen sur l'IA

Synthèse des contributions

Plusieurs contributeurs ont souligné que l'approche par les risques de la proposition du RIA n'était pas la même que celle conduisant à la réalisation d'une AIPD. En particulier, il a été demandé de clarifier si une AIPD était obligatoire pour les systèmes à haut risque dont le développement nécessite de traiter des données personnelles.

Réponse de la CNIL

Conformément à la position tenue par la CNIL et ses homologues dans l'opinion publiée conjointement avec le contrôleur européen à la protection des données, il est précisé qu'**une AIPD serait obligatoire pour les systèmes classés à haut risque par la proposition de RIA**.

Des précisions supplémentaires sur la nécessité d'une AIPD pour les modèles de fondation et sur les systèmes à usage général ont par ailleurs été apportées.

Enfin, en réponse aux interrogations de plusieurs contributeurs, des précisions ont été apportées concernant **l'articulation entre les exigences de documentation de la proposition de RIA et la réalisation d'une AIPD**, de nombreux éléments pouvant être communs aux deux productions.

Sur l'évaluation des risques sur les modèles provenant de tiers

Synthèse des contributions

Certaines contributions ont soulevé qu'il pouvait être difficile de réaliser une AIPD pour les concepteurs de systèmes reposant sur des modèles tiers pré-entraînés (qu'ils adaptent à leurs besoins par ajustement, ou *fine-tuning* par exemple).

Réponse de la CNIL

La CNIL considère que la recommandation, formulée à l'intention des concepteurs de modèles, de réaliser une AIPD afin de la transmettre aux réutilisateurs, et notamment à ceux souhaitant les intégrer dans leur propre phase de développement (notamment pour l'ajustement) est **une mesure suffisante pour permettre aux utilisateurs de modèles pré-entrainés de réaliser leur propre AIPD**.

Par ailleurs, la recommandation de la CNIL invitant les concepteurs de modèles à rendre publiques leurs AIPD **doit apporter des garanties sur la transmission des informations nécessaires aux utilisateurs pour la réalisation de leur propre AIPD dans le cas des modèles publiés en source ouverte**. Pour être conformes aux attentes de la CNIL, ces AIPD doivent être suffisamment complètes afin de permettre aux utilisateurs d'évaluer les risques liés à l'utilisation du modèle au sein de leur traitement.

Sur les critères nécessitant la réalisation d'une AIPD

Synthèse des contributions

Plusieurs contributeurs ont soulevé que ces critères, parmi lesquels figurent les « usages innovants » et les « traitements à grande échelle », pouvaient être difficiles à apprécier sans un seuil exact. Ils invitent la CNIL à apporter ces précisions.

Réponse de la CNIL

La CNIL considère que l'identification de ces seuils ne peut se faire d'une manière générale pour l'ensemble des traitements concernés. Elle relève de **l'appréciation des responsables de traitement**, qui devront tenir compte du contexte propre à leur traitement.

Fiche 6 – Tenir compte de la protection des données dans la conception du système

Sur le respect du principe de minimisation

Synthèse des contributions

Les contributeurs relèvent que le **principe de minimisation** semble difficile à articuler avec le développement d'un système d'IA.

En particulier, les contributeurs signalent qu'il peut être difficile **d'anticiper quelle sera l'architecture la plus adéquate** avant d'avoir testé le système sur des données, par exemple en phase pilote.

Réponse de la CNIL

Si le principe de minimisation demande que la méthode choisie pour atteindre un objectif soit la plus économe possible en données, il ne prévoit pas de **seuil explicite** et n'interdit pas la collecte de grands ensembles de données. En revanche, la CNIL appelle à anticiper le mieux possible la collecte de données et à identifier les données nécessaires avant de se lancer dans la collecte, afin de permettre de ne traiter que ce qui est **strictement nécessaire** à la conception du système d'IA. Les conditions du respect du principe de minimisation ont à cet égard été clarifiées : les méthodes d'apprentissage profond doivent être réservées aux cas où **aucune alternative plus économe n'existe**, et doivent être justifiées. De la même manière, les solutions nécessitant de recourir à des types de données particulièrement identifiantes comme des vidéos ou des photos doit être nécessaire à l'atteinte de l'objectif.

Sur le rôle du Comité éthique

Synthèse des contributions

Plusieurs contributions demandent des précisions concernant le **rôle d'un comité éthique**.

Réponse de la CNIL

La constitution et la consultation d'un tel comité est une **bonne pratique** à associer à la validation d'un projet de développement d'IA qui ne saurait être assimilée à de la gouvernance, ni se substituer aux autorités compétentes en la matière. Même s'il peut intégrer des membres externes, son rôle est de donner un avis interne sur la pertinence de la poursuite d'un projet d'IA.

Etant donné que la constitution d'un comité éthique dépend de la taille et des moyens de la structure dont il dépend, une bonne pratique alternative peut aussi être de consulter ou de mobiliser un « référent IA ».

Fiche 7 – Tenir compte de la protection des données dans la collecte et la gestion des données

Sur la conservation des données à des fins d'audit

Synthèse des contributions

Les contributions ont souligné que la limitation de la durée de conservation pour les données d'apprentissage pouvait être **un frein à la réalisation d'audit du système d'IA et en particulier à la mesure des biais**.

Réponse de la CNIL

D'un point de vue technique, ces analyses sont en effet souvent facilitées par un accès aux données d'apprentissage. La possibilité de conduire ces audits est **d'une importance cruciale à la sécurité des systèmes en phase de déploiement**, en particulier au regard du risque de discrimination qu'ils comportent.

S'il était déjà précisé dans la fiche que des audits sont prévus dans la phase de maintenance qui peut justifier une conservation des données, la CNIL a souhaité clarifier sa position. La fiche a ainsi été modifiée de sorte à :

- préciser **qu'il est possible, après une phase de tri, de conserver les données d'apprentissage à des fins d'audit** ;
- indiquer que **cette conservation nécessite la mise en œuvre de certaines garanties de sécurité** portant notamment sur la restriction des accès aux données, leur chiffrement et leur pseudonymisation ou anonymisation dès que cela est possible.

En revanche, la conservation des données ne peut être justifiée par anticipation qu'un éventuel contentieux nécessite l'accès aux données pour sa résolution.