Avis du comité (article 64)

Avis 28/2024 relatif à certains aspects de la protection des données liés au traitement de données à caractère personnel dans le contexte des modèles d'IA

Adopté le 17 décembre 2024

Translations proofread by EDPB Members.

This language version has not yet been proofread.

Synthèse

Les technologies de l'IA créent de nombreuses possibilités et de nombreux avantages dans un large éventail de secteurs et d'activités sociales.

En protégeant le droit fondamental à la protection des données, le RGPD soutient ces possibilités et promeut d'autres droits fondamentaux de l'Union, notamment le droit à la liberté de pensée, d'expression et d'information, le droit à l'éducation ou la liberté d'entreprise. Ainsi, le RGPD est un cadre juridique qui encourage l'innovation responsable.

Dans ce contexte, compte tenu des questions de protection des données soulevées par ces technologies, l'AC irlandaise a demandé au comité européen de la protection des données (EDPB) d'émettre un avis sur des questions d'application générale conformément à l'article 64, paragraphe 2, du RGPD. Cette demande concerne le traitement de données à caractère personnel dans le cadre des phases de développement et de déploiement des modèles d'intelligence artificielle («**IA**»). Plus concrètement, la demande posait les questions suivantes: 1) quand et comment un modèle d'IA peut-il être considéré comme «anonyme»? 2) comment les responsables du traitement peuvent-ils démontrer qu'il est approprié de prendre l'intérêt légitime comme base juridique au cours de la phase de développement et 3) de déploiement? et 4) quelles sont les conséquences du traitement illicite de données à caractère personnel pendant la phase de développement d'un modèle d'IA sur le traitement ou le fonctionnement ultérieur du modèle d'IA?

En ce qui concerne la première question, l'avis mentionne que les revendications relatives à l'anonymat d'un modèle d'IA devraient être évaluées au cas par cas par les AC compétentes, étant donné que le comité européen de la protection des données considère que les modèles d'IA entraînés à l'aide de données à caractère personnel ne peuvent pas, et ce dans tous les cas, être considérés comme anonymes. Pour qu'un modèle d'IA soit considéré comme anonyme, (1) la probabilité d'une extraction directe (y compris probabiliste) de données à caractère personnel concernant des individus dont les données à caractère personnel ont été utilisées pour développer le modèle et (2) la probabilité d'obtenir, intentionnellement ou non, de telles données à caractère personnel à la suite de requêtes devraient être négligeables, compte tenu de «*tous les moyens raisonnablement susceptibles d'être utilisés*» par le responsable du traitement ou une autre personne.

Pour procéder à leur évaluation, les AC devraient examiner la documentation fournie par le responsable du traitement afin de démontrer l'anonymat du modèle. À cet égard, l'avis fournit une liste non contraignante et non exhaustive de méthodes qui peuvent être utilisées par les responsables du traitement dans leur démonstration de l'anonymat, et qui donc être prises en considération par les AC lors de l'évaluation de la revendication d'anonymat d'un responsable du traitement. Il s'agit, par exemple, des approches adoptées par les responsables du traitement, au cours de la phase de développement, pour empêcher ou limiter la collecte de données à caractère personnel utilisées à des fins de formation, pour réduire leur caractère identifiable, pour empêcher leur extraction ou pour fournir l'assurance que la résistance aux attaques répond aux normes les plus récentes.

En ce qui concerne les deuxième et troisième questions, l'avis contient des considérations générales dont les AC doivent tenir compte lorsqu'elles évaluent si les responsables du traitement peuvent se fonder sur l'intérêt légitime en tant que base juridique appropriée pour le traitement effectué dans le cadre du développement et du déploiement de modèles d'IA.

L'avis rappelle qu'il n'y a pas de hiérarchie entre les bases juridiques prévues par le RGPD et qu'il appartient aux responsables du traitement d'identifier la base juridique appropriée pour leurs activités de traitement. L'avis rappelle ensuite le test en trois étapes auquel il convient de procéder pour évaluer l'utilisation de l'intérêt légitime en tant que base juridique, à savoir 1) la détermination de l'intérêt légitime poursuivi par le responsable du traitement ou un tiers; 2) l'analyse de la nécessité du traitement aux fins de l'intérêt ou des intérêts légitimes poursuivis (également appelé «critère de nécessité»); et 3) l'évaluation du point de savoir si les intérêts ou les libertés et droits fondamentaux des personnes concernées (également appelée «critère de mise en balance») ne prévalent pas sur le ou les intérêts légitimes.

En ce qui concerne la première étape, l'avis rappelle qu'un intérêt peut être considéré comme légitime si les trois critères cumulatifs suivants sont remplis: 1) l'intérêt est licite; 2) il est clairement et précisément formulé; et 3) il est réel et actuel (c'est-à-dire qu'il ne s'agit pas d'un intérêt spéculatif). Cet intérêt peut porter, par exemple, sur l'élaboration d'un modèle d'IA (développer un service d'agent conversationnel pour aider les utilisateurs), ou sur son déploiement (améliorer la détection des menaces au sein d'un système d'information).

En ce qui concerne la deuxième étape, l'avis rappelle que l'évaluation de la nécessité implique de prendre en compte les éléments suivants: 1) l'activité de traitement permettra-t-elle la poursuite de l'intérêt légitime? et 2) n'existe-t-il pas de moyen moins intrusif de poursuivre cet intérêt? Lorsqu'elles évaluent si la condition de nécessité est remplie, les AC devraient accorder une attention particulière à la quantité de données à caractère personnel traitées et à la proportionnalité de la poursuite de l'intérêt légitime en jeu, y compris à la lumière du principe de minimisation des données.

En ce qui concerne la troisième étape, l'avis rappelle que le test de mise en balance doit être effectué en tenant compte des circonstances spécifiques de chacun des cas. Il donne ensuite un aperçu des éléments que les AC peuvent prendre en considération lorsqu'elles évaluent si les intérêts, les libertés et les droits fondamentaux des personnes concernées prévalent sur l'intérêt d'un responsable du traitement ou d'un tiers.

Dans le cadre de la troisième étape, l'avis met en évidence les risques spécifiques pour les droits fondamentaux qui peuvent apparaître au cours des phases de développement ou de déploiement des modèles d'IA. Il précise également que le traitement de données à caractère personnel qui a lieu au cours des phases de développement et de déploiement des modèles d'IA peut avoir une incidence sur les personnes concernées de différentes manières, ce qui peut être positif ou négatif. Pour évaluer cette incidence, les AC peuvent tenir compte de la nature des données traitées par les modèles, du contexte du traitement et des éventuelles conséquences ultérieures du traitement.

L'avis souligne également le rôle des attentes raisonnables des personnes concernées dans le test de mise en balance. Cela peut être important en raison de la complexité des technologies utilisées dans les modèles d'IA et du fait qu'il peut être difficile pour les personnes concernées de comprendre la variété de leurs utilisations potentielles, ainsi que les différentes activités de traitement concernées. À cet égard, tant les informations fournies aux personnes concernées que le contexte du traitement peuvent faire partie des éléments à prendre en compte pour évaluer si les personnes concernées peuvent raisonnablement s'attendre à ce que leurs données à caractère personnel fassent l'objet d'un traitement. En ce qui concerne le contexte, ces éléments peuvent inclure: la question de savoir si les données à caractère personnel ont été accessibles au public ou non, la nature de la relation entre la personne concernée et le responsable du traitement (et s'il existe un lien entre les deux), la nature du service, le contexte dans lequel les données à caractère personnel ont été collectées, la source à partir de laquelle les données à caractère personnel ont été collectées (c'est-à-dire le site web ou le service

où les données à caractère personnel ont été collectées et les paramètres de confidentialité qu'ils proposent), les autres utilisations potentielles du modèle, ainsi que le point de savoir si les personnes concernées savent effectivement que leurs données à caractère personnel sont en ligne.

L'avis rappelle également que, lorsque les intérêts, les droits et les libertés des personnes concernées semblent prévaloir sur le ou les intérêts légitimes poursuivis par le responsable du traitement ou un tiers, le responsable du traitement peut envisager d'introduire des mesures d'atténuation pour limiter l'incidence du traitement sur ces personnes concernées. Les mesures d'atténuation ne doivent pas être confondues avec les mesures que le responsable du traitement est de toute façon légalement tenu d'adopter pour assurer le respect du RGPD. En outre, les mesures devraient être adaptées aux circonstances de l'espèce et aux caractéristiques du modèle d'IA, y compris son utilisation prévue. À cet égard, l'avis fournit une liste non exhaustive d'exemples de mesures d'atténuation concernant la phase de développement (y compris en ce qui concerne le «*web scraping*») et la phase de déploiement. Les mesures d'atténuation peuvent faire l'objet d'une évolution rapide et devraient être adaptées aux circonstances de l'espèce. Par conséquent, il appartient aux AC d'évaluer la pertinence des mesures d'atténuation mises en œuvre au cas par cas.

En ce qui concerne la quatrième question, l'avis rappelle de manière générale que les AC disposent de pouvoirs d'appréciation discrétionnaires pour évaluer la ou les violations éventuelles et choisir des mesures appropriées, nécessaires et proportionnées, en tenant compte des circonstances de chaque cas d'espèce. L'avis examine ensuite trois scénarios.

Dans le scénario 1, les données à caractère personnel sont conservées dans le modèle d'IA (ce qui signifie que le modèle ne peut pas être considéré comme anonyme, tel qu'indiqué dans la première question) et sont traitées ultérieurement par le même responsable du traitement (par exemple dans le cadre du déploiement du modèle). L'avis affirme que la question de savoir si les phases de développement et de déploiement impliquent des finalités distinctes (et qu'elles constituent donc des activités de traitement distinctes) et dans quelle mesure l'absence de base juridique pour l'activité de traitement initiale a une incidence sur la licéité du traitement ultérieur, devrait être évaluée au cas par cas, en fonction du contexte de l'affaire.

Dans le cadre du scénario 2, les données à caractère personnel sont conservées dans le modèle et sont traitées par un autre responsable du traitement dans le cadre du déploiement du modèle. À cet égard, l'avis indique que les AC doivent tenir compte du fait que le responsable du traitement qui déploie le modèle a effectué une évaluation appropriée, dans le cadre de ses obligations de responsabilité pour démontrer le respect de l'article 5, paragraphe 1, point a), et de l'article 6 du RGPD, afin de s'assurer que le modèle d'IA n'a pas été développé au moyen d'un traitement illégal de données à caractère personnel. Cette évaluation devrait tenir compte, par exemple, de la source des données à caractère personnel et de la question de savoir si le traitement au cours de la phase de développement a fait l'objet d'une constatation d'infraction, en particulier si elle a été déterminée par une AC ou une juridiction, et devrait être moins ou plus détaillée en fonction des risques soulevés par le traitement au cours de la phase de déploiement.

Dans le scénario 3, un responsable du traitement traite de manière illicite des données à caractère personnel pour développer le modèle d'IA, puis veille à ce qu'elles soient anonymisées, avant que le même responsable du traitement ou un autre responsable du traitement n'entament un autre traitement de données à caractère personnel dans le cadre du déploiement. À cet égard, l'avis indique que s'il peut être démontré que le fonctionnement ultérieur du modèle d'IA n'implique pas le traitement de données à caractère personnel, l'EDPB considère que le RGPD ne s'appliquerait pas. Par conséquent, le caractère illicite du traitement initial ne devrait pas avoir d'incidence sur le

fonctionnement ultérieur du modèle. En outre, l'EDPB considère que, lorsque les responsables du traitement traitent ultérieurement des données à caractère personnel collectées au cours de la phase de déploiement, après que le modèle a été anonymisé, le RGPD s'appliquerait à ces opérations de traitement. Dans ces cas, l'avis estime que, en ce qui concerne le RGPD, le caractère illicite du traitement initial ne devrait pas avoir d'incidence sur la licéité du traitement effectué au cours de la phase de déploiement.

Table des matières

Le comité européen de la protection des données

vu l'article 63 et l'article 64, paragraphe 2, du règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (ci-après le «RGPD»),

vu l'accord sur l'Espace économique européen (EEE) et, en particulier, son annexe XI et son protocole 37, tels que modifiés par la décision du comité mixte de l'EEE nº 154/2018 du 6 juillet 2018([1],

vu les articles 10 et 22 de son règlement intérieur,

considérant ce qui suit:

(1) La mission principale du comité européen de la protection des données (ci-après le «**comité**» ou l'«**EDPB**») est de veiller à l'application cohérente du RGPD dans l'ensemble de l'Espace économique européen (ci-après l'«**EEA**»). Conformément à l'article 64, paragraphe 2, du RGPD, toute autorité de contrôle («**AC**»), la présidente de l'EDPB ou la Commission peuvent demander que toute question d'application générale ou produisant des effets dans plusieurs États membres de l'EEE soit examinée par l'EDPB en vue d'obtenir un avis. Le présent avis vise à examiner une question d'application générale ou qui produit des effets dans plusieurs États membres de l'EEE.

(2) L'avis de l'EDPB est adopté conformément à l'article 64, paragraphe 3, du RGPD, lu conjointement avec l'article 10, paragraphe 2, du règlement intérieur de l'EDPB, dans un délai de huit semaines à compter de la date à laquelle la présidente et l'autorité de contrôle compétente ont décidé que le dossier était complet. Sur décision de la présidente, ce délai peut être prolongé de six semaines en fonction de la complexité de la question.

A ADOPTÉ LE PRÉSENT AVIS:

1 Introduction

1.1 Résumé des faits

1. Le 4 septembre 2024, l'autorité de contrôle irlandaise (l'«**AC irlandaise**» ou l'«**AC demandeuse**») a demandé au comité d'émettre un avis conformément à l'article 64, paragraphe 2, du RGPD en ce qui concerne les modèles d'IA et le traitement des données à caractère personnel (la «**demande**»).

2. Le président du conseil d'administration et l'AC irlandaise ont estimé le 13 septembre 2024 que le dossier était complet. Le jour ouvrable suivant, à savoir le 16 septembre 2024, le dossier a été diffusé par le secrétariat de l'EDPB. La présidente de l'EDPB, compte tenu de la complexité de la question, a décidé de prolonger le délai légal conformément à l'article 64, paragraphe 3, du RGPD et à l'article 10, paragraphe 4, du règlement intérieur de l'EDPB.

3. La demande porte sur certains éléments de l'entraînement, de la mise à jour, du développement et de l'exploitation de modèles d'IA dans lesquels des données à caractère personnel font partie du jeu de

[1] Dans l'intégralité du présent avis, on entend par «États membres» les «États membres de l'EEE». Dans l'intégralité du présent avis, on entend par «Union» l'«EEE».

données pertinent. L'AC irlandaise souligne que la demande concerne des questions clés qui ont une forte incidence sur les personnes concernées et les responsables du traitement dans l'EEE et qu'il n'existe pas de position harmonisée à ce stade parmi les autorités de contrôle nationales[2]. La terminologie qui sera utilisée aux fins du présent avis est fournie aux points 2.2 et 2.3 ci-dessous.

4. Les questions suivantes ont été posées par l'AC irlandaise:

Question 1: Le modèle d'IA final, qui a été entraîné à l'aide de données à caractère personnel, est-il considéré dans tous les cas comme ne répondant pas à la définition des données à caractère personnel (telle qu'elle est énoncée à l'article 4, paragraphe 1, du RGPD)?

En cas de réponse affirmative à la première question:

i. À quel stade des opérations de traitement aboutissant à un modèle d'IA les données à caractère personnel ne sont-elles plus traitées?
   a) Comment peut-on démontrer que le modèle d'IA ne traite pas de données à caractère personnel?

ii. Existe-t-il des facteurs qui feraient que le fonctionnement du modèle d'IA final ne serait plus considéré comme anonyme?
   a) Dans l'affirmative, comment les mesures prises pour atténuer ces facteurs, les prévenir ou s'en protéger (de manière à garantir que le modèle d'IA ne traite pas de données à caractère personnel) peuvent-elles être démontrées?

En cas de réponse négative à la première question:

i. Quelles sont les circonstances dans lesquelles cela pourrait se produire?
   a) Dans l'affirmative, comment les mesures prises pour garantir que le modèle d'IA ne traite pas de données à caractère personnel peuvent-elles être démontrées?

Question 2: Lorsqu'un responsable du traitement se fonde sur des intérêts légitimes en tant que base juridique pour le traitement de données à caractère personnel aux fins de la création, de la mise à jour et/ou du développement d'un modèle d'IA, comment ce responsable du traitement devrait-il démontrer le caractère approprié des intérêts légitimes en tant que base juridique, tant en ce qui concerne le traitement de données de tiers que de données de première partie?

i. Quelles considérations ce responsable du traitement doit-il prendre en compte pour s'assurer que les intérêts des personnes concernées, dont les données à caractère personnel sont traitées, sont correctement mis en balance avec les intérêts de ce responsable du traitement dans le cadre de:
   a) Données de tiers

   b) Données de première partie

Question 3: Après la formation, lorsqu'un responsable du traitement se fonde sur des intérêts légitimes en tant que base juridique pour un traitement de données à caractère personnel effectué dans le cadre d'un modèle d'IA, ou d'un système d'IA dont fait partie un modèle d'IA, comment un

[2] Demande, p. 1.

responsable du traitement devrait-il démontrer la pertinence d'intérêts légitimes en tant que base juridique?

Question 4: S'il a été constaté qu'un modèle d'IA a été créé, mis à jour ou développé à l'aide de données à caractère personnel traitées de manière illicite, quelle est l'incidence de cette création, le cas échéant, sur la licéité du traitement ou de l'exploitation continus ou ultérieurs du modèle d'IA, que ce soit en tant que tel ou en tant que partie d'un système d'IA, lorsque:

i. Le modèle d'IA, seul ou en tant que partie d'un système d'IA, traite des données à caractère personnel?
ii. Ni le modèle d'IA, ni le modèle d'IA en tant que partie d'un système d'IA, ne traitent de données à caractère personnel?

1.2 Recevabilité de la demande d'avis au titre de l'article 64, paragraphe 2, du RGPD

5. L'article 64, paragraphe 2, du RGPD, prévoit que toute autorité de contrôle peut demander que toute question d'application générale ou produisant des effets dans plusieurs États membres soit examinée par le comité en vue d'obtenir un avis.

6. L'AC demandeuse a adressé des questions à l'EDPB concernant les aspects relatifs à la protection des données dans le contexte des modèles d'IA. Elle a précisé dans sa demande que, bien que de nombreuses organisations utilisent désormais des modèles d'IA, y compris des grands modèles de langage («**LLM**»), leurs opérations, leur entraînement et leur utilisation soulèvent «*un certain nombre de préoccupations de grande envergure en matière de protection des données*»[3], qui «*ont des incidences sur les personnes concernées dans l'ensemble de l'UE/EEE*»[4].

7. La demande soulève, en substance, des questions sur i) l'application de la notion de données à caractère personnel; ii) le principe de licéité, en ce qui concerne spécifiquement la base juridique de l'intérêt légitime, dans le contexte des modèles d'IA; ainsi que iii) les conséquences du traitement illicite de données à caractère personnel au cours de la phase de développement des modèles d'IA sur le traitement ou l'exploitation ultérieurs du modèle.

8. Le comité estime que la demande concerne une «*question d'application générale*» au sens de l'article 64, paragraphe 2, du RGPD. En particulier, la question porte sur l'interprétation et l'application de l'article 4, paragraphe 1, de l'article 5, paragraphe 1, point a), et de l'article 6 du RGPD en ce qui concerne le traitement des données à caractère personnel dans le cadre du développement et du déploiement de modèles d'intelligence artificielle. Comme l'a souligné l'AC demandeuse, l'application de ces dispositions aux modèles d'IA soulève des questions systémiques, abstraites et nouvelles[5]. Le développement et le déploiement rapides de modèles d'IA par un nombre croissant d'organisations soulèvent des questions spécifiques et, comme le souligne la demande, [6]*«le comité européen de la protection des données bénéficiera grandement de l'adoption d'une position commune sur les questions soulevées par la présente demande, ces questions étant au cœur des travaux du comité planifiés à court et à moyen terme»*. En outre, les technologies de l'IA créent de nombreuses

[3] Demande, p. 1.
[4] Ibidem.
[5] Demande, p. 2.
[6] Demande, p. 1. Comme indiqué dans le programme de travail du CEPD pour la période 2024-2025, adopté le 8 octobre 2024, disponible à l'adresse suivante https://www.edpb.europa.eu/system/files/2024-10/edpb_work_programme_2024-2025_en.pdf, le CEPD prévoit de publier, entre autres, des lignes directrices sur l'anonymisation, la pseudonymisation et le «*scraping*» de données dans le contexte de l'IA générative.

possibilités et avantages dans un large éventail de secteurs et d'activités sociales. En outre, le RGPD est un cadre juridique qui encourage l'innovation responsable. Il s'ensuit qu'il existe un intérêt général à procéder à cette évaluation sous la forme d'un avis de l'EDPB, afin de garantir l'application cohérente de certaines dispositions du RGPD dans le contexte des modèles d'IA.

9. La deuxième condition énoncée à l'article 64, paragraphe 2, du RGPD concerne les questions «*produisant des effets dans plusieurs États membres*». L'EDPB rappelle que le terme «effets» doit être interprété au sens large, et qu'il ne se limite donc pas simplement aux effets juridiques[7]. Étant donné que de plus en plus de modèles d'IA sont entraînés et utilisés par un nombre croissant d'organisations dans l'EEE, ils ont une incidence sur un grand nombre de personnes concernées dans l'ensemble de l'EEE, dont certaines ont déjà fait part de préoccupations à leur AC compétente[8]. Par conséquent, l'EDPB considère que la question soulevée par l'AC demandeuse satisfait également à cette condition.

10. La demande comprend un raisonnement écrit sur le contexte et les motivations de la soumission des questions au Conseil, y compris sur le cadre juridique pertinent. Par conséquent, la commission de recours considère que la demande est motivée conformément à l'article 10, paragraphe 3, du règlement intérieur du CEPD.

11. Conformément à l'article 64, paragraphe 3, du RGPD[9], l'EDPB n'émet pas d'avis s'il a déjà émis un avis sur la question. Le comité européen de la protection des données n'a pas émis d'avis sur le même sujet et n'a pas encore fourni de réponses aux questions soulevées par la demande.

12. Pour ces raisons, le comité estime que la demande est recevable et que les questions qu'elle soulève devraient être analysées dans cet avis (l'«**avis**») adopté conformément à l'article 64, paragraphe 2, du RGPD.

2 Champ d'application et notions clés

2.1 Portée de l'avis

13. Le comité partage l'avis de l'AC demandeuse selon lequel, du point de vue de la protection des données, le développement et le déploiement de modèles d'IA soulèvent des questions fondamentales en matière de protection des données. Les questions portent en particulier sur: i) quand et comment un modèle d'IA peut être considéré comme «anonyme» (question 1 de la demande); ii) comment les responsables du traitement peuvent-ils démontrer le caractère approprié de l'intérêt légitime en tant que base juridique pour les phases de développement (question 2 de la demande) et de déploiement (question 3 de la demande); et iii) si le traitement illicite de données à caractère personnel au cours de la phase de développement a des conséquences sur la licéité du traitement ou du fonctionnement ultérieur du modèle d'IA (question 4 de la demande).

14. L'EDPB rappelle que les AC sont responsables du contrôle de l'application du RGPD et devraient contribuer à son application cohérente dans l'ensemble de l'Union[10]. Il relève donc de la compétence

[7] Document interne 3/2019 de l'EDPB sur les orientations internes relatives à l'article 64, paragraphe 2, du RGPD, adopté le 8 octobre 2019, paragraphe 15, disponible à l'adresse suivante: https://www.edpb.europa.eu/system/files/2022-07/internaledpb_document_201903_art64.2_en.pdf.
[8] Demande, p. 1 à 2.
[9] Article 64, paragraphe 3, du RGPD et article 10, paragraphe 4, du règlement intérieur de l'EDPB.
[10] Article 51, paragraphe 1, du RGPD et article 51, paragraphe 2, du RGPD.

des AC d'enquêter sur des modèles d'IA spécifiques et, ce faisant, de procéder à des évaluations au cas par cas.

15. Le présent avis fournit un cadre permettant aux AC compétentes d'évaluer des cas spécifiques dans lesquels (certaines) des questions soulevées dans la demande se poseraient. Le présent avis ne vise pas à être exhaustif, mais plutôt à fournir des considérations générales sur l'interprétation des dispositions pertinentes, dont les AC compétentes devraient tenir le plus grand compte lorsqu'elles font usage de leurs pouvoirs d'enquête. Bien que le présent avis soit adressé aux AC compétentes et porte sur leurs activités et leurs pouvoirs, il est sans préjudice des obligations qui incombent aux responsables du traitement et aux sous-traitants en vertu du RGPD. En particulier, conformément au principe de responsabilité énoncé à l'article 5, paragraphe 2, du RGPD, les responsables du traitement sont responsables du respect de tous les principes relatifs à leur traitement des données à caractère personnel et sont en mesure de démontrer qu'ils les respectent.

16. Dans certains cas, quelques exemples peuvent être fournis dans l'avis, mais compte tenu du large champ d'application des questions figurant dans la demande, ainsi que des différents types de modèles d'IA couverts par celle-ci, tous les scénarios possibles ne seront pas pris en considération dans le présent avis. Les technologies associées aux modèles d'IA sont sujettes à une évolution rapide; par conséquent, les considérations de l'EDPB dans le présent avis devraient être interprétées à la lumière de ce qui précède.

17. **Le présent avis n'analyse pas les dispositions ci-dessous, qui peuvent néanmoins jouer un rôle important lors de l'évaluation des exigences en matière de protection des données applicables aux modèles d'IA:**

- **Traitement portant sur des catégories particulières de données** L'EDPB rappelle l'interdiction prévue à l'article 9, paragraphe 1, du RGPD concernant le traitement de catégories particulières de données et les exceptions limitées prévues à l'article 9, paragraphe 2, du RGPD[11]. À cet égard, la Cour de justice de l'Union européenne (ci-après la «**CJUE**») a en outre précisé qu'«*il y a lieu de préciser que, dans le cas où un ensemble de données comportant à la fois des données sensibles et des données non sensibles [...] est notamment collecté en bloc sans que les données puissent être dissociées les unes des autres au moment de cette collecte, le traitement de cet jeu de données doit être considéré comme étant interdit, au sens de l'article 9, paragraphe 1, du RGPD dès lors qu'il comporte au moins une donnée sensible et qu'aucune des dérogations visées à l'article 9, paragraphe 2, de ce règlement n'est applicable*»[12]. En outre, la CJUE a également souligné qu'«*aux fins de l'application de l'exception prévue à l'article 9, paragraphe 2, sous e), du RGPD, il importe de vérifier si la personne concernée a entendu, de manière explicite et par un acte positif clair, rendre accessibles au grand public les données à caractère personnel en question*[13]». Ces considérations doivent être prises en compte lorsque le traitement de données

[11] Voir également le rapport du comité européen de la protection des données sur les travaux entrepris par le groupe de travail ChatGPT, adopté le 23 mai 2024, point 18: «En ce qui concerne le traitement de catégories particulières de données à caractère personnel, l'une des exceptions visées à l'article 9, paragraphe 2, doit en outre s'appliquer pour que le traitement soit licite. *En principe, l'une de ces exceptions peut être l'article 9, paragraphe 2, point e), du RGPD. Toutefois, le simple fait que des données à caractère personnel soient accessibles au public n'implique pas que "la personne concernée a manifestement rendu ces données publiques" [...]*».

[12] Arrêt de la CJUE du 4 juillet 2023, affaire C-252/21, *Meta contre Bundeskartellamt*(ECLI:EU:C:2023:537), point 89.

[13] Arrêt de la CJUE du 4 juillet 2023, affaire C-252/21, *Meta contre Bundeskartellamt* (ECLI:EU:C:2023:537), point 77.

à caractère personnel dans le cadre de modèles d'IA implique des catégories spéciales de données.

- **Prise de décision automatisée, y compris le profilage**: Les opérations de traitement effectuées dans le cadre de modèles d'IA peuvent relever du champ d'application de l'article 22 du RGPD, qui impose des obligations supplémentaires aux responsables du traitement et prévoit des garanties supplémentaires pour les personnes concernées. L'EDPB rappelle, à cet égard, ses lignes directrices sur la prise de décision individuelle automatisée et le profilage aux fins du règlement (UE) 2016/679[14].

- **Compatibilité des finalités**: L'article 6, paragraphe 4, du RGPD prévoit, pour certaines bases juridiques, des critères qu'un responsable du traitement doit prendre en compte pour vérifier si le traitement pour une autre finalité est compatible avec la finalité pour laquelle les données à caractère personnel ont été initialement collectées. Cette disposition peut être pertinente dans le contexte du développement et du déploiement de modèles d'IA et son applicabilité devrait être évaluée par les AC.

- **Analyses d'impact relatives à la protection des données** («**AIPD**») (article 35 du RGPD): Les AIPD constituent un élément important de la responsabilité, lorsque le traitement dans le contexte des modèles d'IA est susceptible d'engendrer un risque élevé pour les droits et libertés des personnes physiques[15].

- **Principe de protection des données dès la conception** (article 25, paragraphe 1, du RGPD): La protection des données dès la conception est une garantie essentielle devant être évaluée par les AC dans le contexte du développement et du déploiement d'un modèle d'IA.

2.2 Notions clés

18. À titre liminaire, le comité européen de la protection des données souhaite apporter des précisions sur la terminologie et les concepts qu'il utilise tout au long du présent avis, et uniquement aux fins du présent avis:

- **Les «données de première partie»** désignent les données à caractère personnel que le responsable du traitement a collectées auprès des personnes concernées.

- Par «**données de tiers**», l'on entend des données à caractère personnel que les responsables du traitement n'ont pas obtenues auprès des personnes concernées, mais qui ont été collectées ou reçues d'un tiers, par exemple d'un courtier de données, ou collectées au moyen d'un outil de «*web scraping*» (moissonnage).

- Le «***web scraping***», ou «moissonnage», est une technique couramment utilisée pour collecter des informations à partir de sources en ligne accessibles au public. Les informations extraites, par

[14] Lignes directrices du groupe de travail «article 29» (ci-après le «**GT29**») sur la prise de décision individuelle automatisée et le profilage aux fins du règlement (UE) 2016/679, telles que révisées en dernier lieu et adoptées le 6 février 2018, approuvées par le comité européen de la protection des données le 25 mai 2018. Voir également l'arrêt de la CJUE du 7 décembre 2023, affaire C-634/21, *SCHUFA Holding et autres* (ECLI:EU:C:2023:957).

[15] Voir groupe de travail «Article 29», «Lignes directrices concernant l'analyse d'impact relative à la protection des données (AIPD) et la manière de déterminer si le traitement est "susceptible d'engendrer un risque élevé" aux fins du règlement (UE) 2016/679», modifiées et adoptées le 4 octobre 2017, approuvées par le CEPD le 25 mai 2018:

exemple, de services tels que les organes de presse, les médias sociaux, les forums de discussion et les sites web personnels peuvent contenir des données à caractère personnel.

- La demande fait référence au «**cycle de vie» des modèles d'IA**, ainsi qu'à diverses étapes concernant, entre autres, la «création», le «développement», la «formation», la «mise à jour», le «réglage fin», l'«exploitation» ou le «post-apprentissage» des modèles d'IA. L'EDPB reconnaît que, selon les circonstances, ces étapes peuvent avoir lieu dans le cadre du développement et du déploiement de modèles d'IA et peuvent inclure le traitement de données à caractère personnel pour diverses finalités de traitement. Néanmoins, aux fins du présent avis, l'EDPB considère qu'il est important de rationaliser la catégorisation des étapes susceptibles de se produire. Par conséquent, aux fins du présent avis, l'EDPB fait référence à la «**phase de développement**» et à la «**phase de déploiement**». Le développement d'un modèle d'IA couvre toutes les étapes qui précèdent le déploiement du modèle d'IA et comprend, entre autres, le développement du code, la collecte des données personnelles d'apprentissage, le prétraitement des données personnelles de formation, et l'entraînement. Le déploiement d'un modèle d'IA couvre toutes les étapes relatives à l'utilisation d'un modèle d'IA et peut inclure toutes les opérations menées après la phase de développement. L'EDPB reste conscient de la diversité des cas d'utilisation et de leurs conséquences potentielles en matière de traitement des données à caractère personnel; par conséquent, les AC devraient examiner si les observations formulées dans le présent avis sont pertinentes pour le traitement qu'elles évaluent.
- L'EDPB souligne également que, le cas échéant, le terme «**entraînement**» fait référence à la partie de la phase de développement durant laquelle les modèles d'IA apprennent à partir de données en vue d'exécuter la tâche prévue (comme expliqué dans la section suivante du présent avis).
- La notion et le champ d'application des **modèles d'IA**, tels qu'ils sont compris par l'EDPB aux fins du présent avis, sont précisés dans la section spécifique suivante.

2.3 Les modèles d'IA dans le contexte de l'avis

19. Le règlement de l'UE sur l'intelligence artificielle (le «**règlement sur l'IA**»)[16] définit un «système d'IA» comme étant «*un système automatisé qui est conçu pour fonctionner à différents niveaux d'autonomie et peut faire preuve d'une capacité d'adaptation après son déploiement, et qui, pour des objectifs explicites ou implicites, déduit, à partir des entrées qu'il reçoit, la manière de générer des sorties telles que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels*»[17]. Le considérant (12) du règlement sur l'IA explique en outre la notion de «système d'IA». Une caractéristique essentielle des systèmes d'IA est, en conséquence, leur capacité d'inférence. Les techniques qui permettent de tirer des conclusions lors de la construction d'un système d'IA comprennent l'apprentissage automatique, les approches logiques et les approches fondées sur la connaissance.

20. Les «modèles d'IA», quant à eux, ne sont définis qu'indirectement dans le règlement sur l'IA: «*Bien que les modèles d'IA soient des composants essentiels des systèmes d'IA, ils ne constituent pas en soi des systèmes d'IA. Les modèles d'IA nécessitent l'ajout d'autres composants, tels qu'une interface*

[16] Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle et modifiant les règlements (CE) nº 300/2008, (UE) nº 167/2013, (UE) nº 168/2013, (UE) 2018/858, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828 (règlement sur l'IA)
[17] Article 3, paragraphe 1, du règlement sur l'IA.

utilisateur, pour devenir des systèmes d'IA. Les modèles d'IA sont généralement intégrés dans les systèmes d'IA et en font partie»[18].

21. L'EDPB comprend que la définition d'un modèle d'IA proposée dans la demande est plus étroite que celle énoncée dans le règlement sur l'IA, puisqu'elle fait référence au «modèle d'IA» comme «*englobant le produit résultant des mécanismes d'apprentissage qui sont appliqués à un ensemble de données d'entraînement, dans le contexte de l'intelligence artificielle, de l'apprentissage automatique, de l'apprentissage profond ou d'autres contextes de traitement connexes*» et précise en outre que «*[l]e terme s'applique aux modèles d'IA qui sont destinés à faire l'objet d'un entraînement, d'une mise au point et/ou d'un développement ultérieurs, ainsi qu'aux modèles d'IA qui ne le sont pas*»[19].

22. Sur cette base, l'EDPB a adopté le présent avis en partant du principe qu'un système d'IA s'appuiera sur un modèle d'IA pour atteindre l'objectif visé en intégrant le modèle dans un cadre plus large (par exemple, un système d'IA pour le service à la clientèle pourrait utiliser un modèle d'IA formé à partir de données de conversation historiques pour fournir des réponses aux questions des utilisateurs).

23. En outre, les modèles d'IA (ou «**modèles**») pertinents aux fins du présent avis sont ceux qui sont élaborés dans le cadre d'un processus de formation. Ce processus d'entraînement fait partie de la phase de développement, au cours de laquelle les modèles apprennent à l'aide de données afin d'accomplir la tâche prévue. Par conséquent, le processus d'entraînement nécessite un ensemble de données à partir duquel le modèle identifiera et «apprendra» des schémas. Dans ces cas, le modèle utilisera différentes techniques pour construire une représentation des connaissances extraites du jeu de données d'entraînement. C'est notamment le cas de l'apprentissage automatique.

24. En pratique, tout modèle d'IA est un algorithme, dont le fonctionnement est déterminé par un ensemble d'éléments. Par exemple, les modèles d'apprentissage profond se présentent souvent sous la forme d'un réseau neuronal comportant plusieurs couches composées de nœuds reliés par des bords assortis de pondérations, qui sont ajustés au cours de la formation afin d'apprendre les relations entre les entrées et les sorties. Les caractéristiques d'un modèle d'apprentissage profond simple seraient les suivantes: (i) le type et la taille de chaque couche, (ii) le poids attribué à chaque arête (parfois appelé «paramètres»), (iii) les fonctions d'activation[20] entre les couches, et éventuellement (iv) d'autres opérations susceptibles de se produire entre les couches. Par exemple, lorsqu'il s'agit de former un modèle simple d'apprentissage profond pour la classification d'images, les données d'entrée (les «**pixels d'image»**) seront associées aux résultats, et les pondérations peuvent être ajustées, de manière à produire le plus souvent le résultat juste.

25. Parmi les autres exemples de modèles d'apprentissage profond figurent les LLM et l'IA générative, qui sont utilisés, par exemple, pour générer du contenu de type humain et créer de nouvelles données.

26. **Sur la base des considérations qui précèdent, conformément à la demande, le champ d'application du présent avis ne couvre que le sous-ensemble de modèles d'IA qui résultent d'un entraînement de ces modèles à l'aide de données à caractère personnel.**

[18]Considérant 97 du règlement sur l'IA.

[19] Demande, p. 3.

[20] C'est-à-dire les fonctions qui calculent, sur la base des données d'entrée et des pondérations, les données de sortie d'un nœud neuronal qui seront ensuite envoyées à la couche suivante du réseau neuronal.

3 Sur le fond de la demande

3.1 Sur la nature des modèles d'IA par rapport à la définition des données à caractère personnel

27. L'article 4, paragraphe 1, du RGPD définit les données à caractère personnel comme «*toute information se rapportant à une personne physique identifiée ou identifiable*» (c'est-à-dire la personne concernée); En outre, le considérant 26 du RGPD prévoit que les principes de protection des données ne doivent pas s'appliquer aux informations anonymes, c'est-à-dire aux informations qui ne se rapportent pas à une personne physique identifiée ou identifiable, compte tenu de «*tous les moyens raisonnablement susceptibles d'être utilisés*» par le responsable du traitement ou une autre personne. Cela suppose notamment: i) des données qui n'ont jamais été liées à une personne identifiée ou identifiable; et ii) des données à caractère personnel qui ont été anonymisées de telle manière que la personne concernée n'est pas ou plus identifiable.

28. En conséquence, il est possible[21] de répondre à la question 1 de la demande en déterminant si un modèle d'IA résultant d'un entraînement impliquant le traitement de données à caractère personnel doit, dans tous les cas, être considéré comme anonyme. Sur la base de la formulation de la question, l'EDPB fera référence, dans cette section, au processus d'«entraînement» d'un modèle d'IA.

29. Tout d'abord, l'EDPB souhaite formuler les considérations générales suivantes. Les modèles d'IA, qu'ils soient en traînés ou non à l'aide de données à caractère personnel, sont généralement conçus pour faire des prédictions ou tirer des conclusions, c'est-à-dire qu'ils sont conçus en vue d'effectuer des déductions. En outre, les modèles d'IA entraînés à l'aide de données à caractère personnel sont souvent conçus pour effectuer des déductions sur des personnes qui ne sont pas celles dont les données à caractère personnel ont été utilisées pour entraîner le modèle d'IA. Toutefois, certains modèles d'IA sont spécifiquement conçus pour fournir des données à caractère personnel concernant des personnes physiques dont les données à caractère personnel ont été utilisées pour entraîner le modèle, ou d'une manière ou d'une autre pour mettre ces données à disposition. Dans de tels cas, ces modèles d'IA comprendront intrinsèquement (et, en général, nécessairement) des informations relatives à une personne physique identifiée ou identifiable, et impliqueront donc le traitement de données à caractère personnel. Par conséquent, ces types de modèles d'IA ne peuvent pas être considérés comme anonymes. Ce serait le cas, par exemple, (i) d'un modèle génératif affiné à partir des enregistrements vocaux d'une personne pour imiter sa voix; ou (ii) de tout modèle conçu pour répondre avec des données à caractère personnel issues de la phase d'apprentissage lorsqu'on lui demande des informations concernant une personne spécifique.

30. Sur la base des considérations ci-dessus, en répondant à la question 1 de la demande, l'EDPB se concentre sur la situation des modèles d'IA qui ne sont pas conçus pour fournir des données à caractère personnel liées aux données d'entraînement.

31. Le comité européen de la protection des données considère que, même lorsqu'un modèle d'IA n'a pas été intentionnellement conçu pour produire des informations relatives à une personne physique identifiée ou identifiable à partir des données d'entraînement, les informations du jeu de données d'entraînement, y compris les données à caractère personnel, peuvent néanmoins rester «absorbées» dans les paramètres du modèle, à savoir représentées au moyen d'objets mathématiques. Elles

[21] «*Le modèle d'IA final, qui a été entraîné à l'aide de données à caractère personnel, est-il considéré dans tous les cas comme ne répondant pas à la définition des données à caractère personnel (telle qu'énoncée à l'article 4, paragraphe 1, du RGPD)?*»

peuvent différer des points de données d'entraînement originaux, mais peuvent toujours conserver les informations originales de ces données, qui peuvent finalement être extraites ou obtenues d'une autre manière, directement ou indirectement, à partir du modèle. Lorsque des informations relatives à des personnes identifiées ou identifiables dont les données à caractère personnel ont été utilisées pour entraîner le modèle peuvent être obtenues à partir d'un modèle d'IA par des moyens raisonnablement susceptibles d'être utilisés, on peut conclure que ce modèle n'est pas anonyme.

32. À cet égard, la demande indique que «*[l]es publications scientifiques existantes mettent en évidence certaines vulnérabilités potentielles qui peuvent exister dans les modèles d'IA et qui pourraient entraîner le traitement de données à caractère personnel*[22]*, ainsi que le traitement de données à caractère personnel susceptible de se poursuivre lorsque les modèles sont déployés pour être utilisés avec d'autres données, soit au moyen d'interfaces de programmation d'applications (ci-après les «**API**»), soit au moyen d'interfaces «guidées»*»[23].

33. Dans le même ordre d'idées, les recherches sur la formation à l'extraction de données sont particulièrement dynamiques[24]. Elles montrent qu'il est possible, dans certains cas, d'utiliser des moyens raisonnablement susceptibles d'extraire des données à caractère personnel de certains modèles d'IA, ou simplement d'obtenir accidentellement des données à caractère personnel par le biais d'interactions avec un modèle d'IA (par exemple, dans le cadre d'un système d'IA). Des efforts de recherche continus dans ce domaine contribueront à évaluer plus avant les risques résiduels de régurgitation[25] et d'extraction de données à caractère personnel dans un cas donné.

34. **Sur la base des considérations ci-dessus, l'EDPB estime que les modèles d'IA entraînés à partir de données à caractère personnel ne peuvent pas, dans tous les cas, être considérés comme anonymes. Il convient plutôt, pour déterminer si un modèle d'IA est anonyme, de fonder l'évaluation sur des critères spécifiques, au cas par cas.**

3.2 Sur les circonstances dans lesquelles les modèles d'IA pourraient être considérés comme anonymes, et la démonstration correspondante

[22] Telles que les attaques par inférence d'appartenance (OWASP) et les attaques par inversion de modèle (OWASP & Veale et al, 2018).

[23] Demande, p. 1 et 2.

[24] Voir, à cet égard, par exemple: (i) Veale M., Binns R., Edwards L., 2018, «*Algorithms that remember: model inversion attacks and data protection law*». Phil. Phil. Trans. A 376: 20180083, disponible à l'adresse suivante: http://dx.doi.org/10.1098/rsta.2018.0083 (ii) Brown H., Lee K., Mireshghallah F., Shokri R., et Tramèr F., *What Does it Mean for a Language Model to Preserve Privacy?*, 2022, ACM Digital Library, FAccT' 22, 20 juin 2022, Séoul, République de Corée, disponible à l'adresse suivante: https://dl.acm.org/doi/abs/10.1145/3531146.3534642 iii) Vassilev A., Oprea A., Fordyce A., Anderson H., *Adversarial Machine Learning A Taxonomy and Terminology of Attacks and Mitigations,* janvier 2024, National Institute of Standards and Technology, disponible à l'adresse https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-2e2023.pdf; iv) Carlini N., Tramèr F., Wallace E., Jagielski M., Herbert-Voss A., Lee K., Roberts A., Brown T., Song D., Erlingsson U., Oprea A., Raffel C., *Extracting Training Data from Large Language Models*, arXiv:2012.07805v2 [cs.CR] 15 juin 2021, disponible à l'adresse suivante https://arxiv.org/pdf/2012.07805; v) Fredrikson M., Jha S., Ristenpart T., *Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures,* ACM Digital Library, 12 octobre 2015, disponible à l'adresse suivante https://arxiv.org/pdf/2012.07805; vi) Zhang Y., Jia R., Pei H., Wang W., Li B., Song D., *The Secret Revealer: Generative Model-Inversion Attacks Against Deep Neural Networks*, arXiv:1911.07135v2 [cs.LG] 18 avril 2020, disponible à l'adresse https://arxiv.org/pdf/1911.07135.

[25] Dans le cas d'un système d'IA fondé sur l'IA générative, la régurgitation correspond à la situation dans laquelle les résultats seraient directement liés aux données d'apprentissage.

35. En ce qui concerne la question 1 de la demande[26], il est demandé à l'EDPB de clarifier les circonstances dans lesquelles un modèle d'IA, qui a été entraîné à l'aide de données personnelles, peut être considéré comme anonyme. En ce qui concerne la question 1, point i), lettre a), de la demande[27], l'EDPB est invité à préciser quels éléments de preuve et/ou documents les AC devraient prendre en considération lorsqu'elles évaluent l'anonymat d'un modèle d'IA.

3.2.1 Considération générale concernant l'anonymisation dans le contexte de l'espèce

36. L'utilisation de l'expression «*toute information*» dans la définition de «*données à caractère personnel*» à l'article 4, paragraphe 1, du RGPD reflète l'objectif d'attribuer une large portée à ce concept, qui englobe toutes sortes d'informations à condition qu'elles «*se rapportent*» à la personne concernée, qui est identifiée ou peut être identifiée directement ou indirectement.

37. Les informations peuvent se rapporter à une personne physique même si elles sont techniquement organisées ou codées (par exemple dans un format uniquement lisible par machine, qu'il soit propriétaire ou ouvert) d'une manière qui ne rend pas la relation avec cette personne physique immédiatement apparente. Dans de tels cas, il est possible d'utiliser des applications logicielles pour identifier, reconnaître et extraire facilement des données spécifiques. Cela est particulièrement vrai pour les modèles d'IA dans lesquels les paramètres représentent des relations statistiques entre les données d'entraînement et dans lesquels il pourrait s'avérer possible d'extraire des données à caractère personnel exactes ou inexactes (car déduites de manière statistique), soit directement à partir des relations entre les données incluses dans le modèle, soit en interrogeant ce modèle.

38. Étant donné que les modèles d'IA ne contiennent généralement pas d'archives pouvant être directement isolées ou liées, mais plutôt des paramètres représentant des relations probabilistes entre les données contenues dans le modèle, il peut se révéler possible de déduire[28] des informations à partir

[26] «*Quelles sont les circonstances dans lesquelles cela peut se produire?*»
[27] «*Dans l'affirmative, comment peut-on démontrer que des mesures ont été prises pour garantir que le modèle d'IA ne traite pas de données à caractère personnel?*»
[28] (i) Carlini N., Chien S., Nasr M., Song S., Terzis A., Tramer F., *Membership Inference Attacks From First Principles, arXiv:2112.03570,* disponible à l'adresse suivante https://arxiv.org/abs/2112.03570;
(ii) Creţu A.M, Guépin F., et De Montjoye Y.A, *Correlation inference attacks against machine learning models*. Hung. Adv.10, eadj9260(2024). DOI:10.1126/sciadv.adj9260, disponible à l'adresse suivante https://www.science.org/doi/10.1126/sciadv.adj9260;
(iii) Dana L., Pydi M. S., Chevaleyre Y., *Memorization in Attention-only Transformers* arXiv:2411.10115v1 [cs.AI] 15 novembre 2024, disponible à l'adresse suivante : https://arxiv.org/abs/2411.10115;
(iv) Gehrke M., Liebenow J., Mohammadi E. & Braun T. et al. *Lifting in Support of Privacy-Preserving Probabilistic Inférence (Le lift à l'appui d'inférences probabilistes préservant la vie privée)*. Künstl Intell, 13 juin 2024, disponible à l'adresse suivante: https://doi.org/10.1007/s13218-024-00851-y;
(v) Hu H., *Membership Inference Attacks and Defenses on Machine Learning Models Littérature*, disponible à l'adresse: https://github.com/HongshengHu/membership-inference-machine-learning-literature;
(vi) Nasr M., Carlini N., Hayase J., Jagielski M., Cooper A. F., Ippolito D., Choquette-Choo C. A., Wallace E., Tramèr F., et Lee K., Scalable *Extraction of Training Data from (Production) Language Models*, arXiv:2311.17035, 28 novembre 2023, disponible à l'adresse suivante: https://arxiv.org/abs/2311.17035
(vii) Shokri R., Stronati M., Song C., Shmatikov V., *Membership Inference Attacks against Machine Learning Models* arXiv:1610.05820v2 [cs.CR], 31 mars 2017, disponible à l'adresse suivante:
(viii) Staab R., Vero M., Mislav Balunović, Martin Vechev, 2024, *Beyond Memorization: Violating Privacy Via Inference with Large Language Models*, arXiv:2310.07298v2, 6 mai 2024, disponible à l'adresse suivante:
(ix) Wu F., Cui L., Yao S., Yu S., *InferenceAttacks in Machine Learning as a Service: A Taxonomy, Review, and Promising Directions* arXiv:2406.02027v1 [cs.LG], 27 juin 2024, disponible à l'adresse suivante: https://arxiv.org/abs/2406.02027v1;

du modèle, telles que l'inférence d'appartenance, dans des scénarios réalistes. Par conséquent, pour qu'une AC soit d'accord avec le responsable du traitement sur le fait qu'un modèle d'IA donné peut être considéré comme anonyme, elle devrait au moins vérifier si elle a reçu des preuves suffisantes que, par des moyens raisonnables: i) les données à caractère personnel, liées aux données d'entraînement, ne peuvent être extraites[29] du modèle; et ii) aucun des résultats produits lors de l'interrogation du modèle ne concerne les personnes concernées dont les données à caractère personnel ont été utilisées pour entraîner le modèle.

39. Trois éléments devraient être pris en considération par les AC pour déterminer si ces conditions sont remplies.

40. Premièrement, les AC devraient tenir compte des éléments recensés dans les avis les plus récents du groupe de travail «Article 29» et/ou dans les lignes directrices de l'EDPB en la matière. En ce qui concerne l'anonymisation à la date du présent avis, les AC devraient prendre en considération les éléments inclus dans l'avis 05/2014 du groupe de travail «Article 29» sur les techniques d'anonymisation (ci-après l'«**avis 05/2014 du groupe de travail «Article 29»**), qui indique que s'il n'est pas possible de distinguer, de relier et de déduire des informations à partir du jeu de données prétendument anonyme, les données peuvent être considérées comme anonymes[30]. Il précise également que, «*lorsqu'une proposition ne répond pas à l'un des critères, il convient de procéder à une évaluation approfondie des risques d'identification*»[31]. **Compte tenu de la probabilité d'extraction et d'inférence susmentionnée, l'EDPB estime que les modèles d'IA nécessiteront très probablement ce type d'évaluation approfondie des risques d'identification.**

41. Deuxièmement, cette évaluation doit être effectuée en tenant compte de «*l'ensemble des moyens raisonnablement susceptibles d'être utilisés*» par le responsable du traitement ou une autre personne pour identifier les personnes physiques[32], et la détermination de ces moyens doit être fondée sur des facteurs objectifs, comme l'explique le considérant 26 du RGPD, qui peuvent inclure:
    a. les caractéristiques des données d'entraînement elles-mêmes, le modèle d'IA et la procédure de formation[33];
    b. le contexte dans lequel le modèle d'IA est publié et/ou traité[34];
    c. les informations supplémentaires qui permettraient l'identification et pourraient être mises à la disposition de la personne concernée;

(x) Zhang J., Das D., Kamath G., Tramèr F., *Membership Inference Attacks Cannot Prove that a Model Was Trained On Your Data* arXiv:2409.19798v1, [cs.LG], 29 septembre 2024, disponible à l'adresse suivante: https://arxiv.org/abs/2409.19798;

(xi) Zhou Z., Xiang J., Chen C. et Su S., *Quantifying and Analyzing Entity-Level Memorization in Large Language Models*, arXiv:2308.15727v2 [cs.CL] 5 novembre 2023, disponible à l'adresse: https://arxiv.org/abs/2308.15727.

[29] L'extraction comprend en particulier le cas où des données à caractère personnel sont déduites du modèle d'IA lui-même, avec peu ou pas d'utilisation des interfaces de requête.

[30] Avis 05/2014 du GT29, page 24.

[31] Avis 05/2014 du GT29, page 24.

[32] Arrêt de la CJUE du 19 octobre 2016, affaire C-582/14, *Breyer contre Bundesrepublik Deutschland* (ECLI:EU:C:2016:779), point 43.

[33] Il s'agit notamment de caractéristiques telles que l'unicité des enregistrements dans les données d'entraînement, la précision des informations, l'agrégation, la randomisation et, en particulier, la manière dont elles influent sur la vulnérabilité à l'identification.

[34] Cela inclut des éléments contextuels, tels que la limitation de l'accès à certaines personnes uniquement ou des garanties juridiques.

d. les coûts et le temps dont la personne aurait besoin pour obtenir ces informations supplémentaires (si elles ne sont pas déjà à sa disposition)[35]; et
e. la technologie disponible au moment du traitement, ainsi que les évolutions technologiques[36].

42. Troisièmement, les AC devraient examiner si les responsables du traitement ont évalué le risque d'identification par le responsable du traitement et par différents types d'«*autres personnes*», y compris les tiers non intentionnels accédant au modèle d'IA, en examinant également s'ils peuvent raisonnablement être considérés comme étant en mesure d'accéder aux données en question ou de les traiter.

43. **En résumé, l'EDPB estime que, pour qu'un modèle d'IA soit considéré comme anonyme, en utilisant des moyens raisonnables, tant i) la probabilité d'une extraction directe (y compris probabiliste) de données à caractère personnel concernant des personnes dont les données à caractère personnel ont été utilisées pour entraîner le modèle, que ii) la probabilité d'obtenir, intentionnellement ou non, ces données à caractère personnel à partir de requêtes, devraient être insignifiantes[37] pour toute personne concernée**. **Par défaut, les AC devraient considérer que les modèles d'IA sont susceptibles de nécessiter une évaluation approfondie de la probabilité d'identification afin de parvenir à une conclusion sur leur éventuelle nature anonyme. Cette probabilité devrait être évaluée en tenant compte de «*l'ensemble des moyens raisonnablement susceptibles d'être utilisés*» par le responsable du traitement ou une autre personne, et devrait également tenir compte de la (ré)utilisation ou de la divulgation involontaires du modèle.**

3.2.2 Éléments permettant d'évaluer la probabilité résiduelle d'identification

44. Bien que des mesures puissent être prises tant au stade du développement qu'au stade du déploiement afin de réduire la probabilité d'obtenir des données à caractère personnel à partir d'un modèle d'IA, l'évaluation de l'anonymat d'un modèle d'IA devrait également tenir compte de l'accès direct au modèle.

45. En outre, les AC devraient évaluer, au cas par cas, si les mesures mises en œuvre par le responsable du traitement pour garantir et prouver l'anonymat d'un modèle d'IA sont appropriées et efficaces.

46. En particulier, la conclusion de l'évaluation d'une AC peut varier entre un modèle d'IA accessible au public, qui est accessible à un nombre inconnu de personnes disposant d'un éventail inconnu de méthodes pour essayer d'extraire des données à caractère personnel, et un modèle d'IA interne accessible uniquement aux employés. Si, dans les deux cas, les AC devraient vérifier que les responsables du traitement ont satisfait à leur obligation de responsabilité au titre de l'article 5, paragraphe 2, et de l'article 24 du RGPD, les «*moyens raisonnablement susceptibles d'être utilisés*» par d'autres personnes peuvent avoir une incidence sur l'étendue et la nature des scénarios possibles à envisager. Par conséquent, en fonction du contexte de développement et de déploiement du modèle, les AC peuvent envisager différents niveaux d'essai et de résistance aux attaques.

47. À cet égard, l'EDPB fournit ci-après une liste non prescriptive et non exhaustive d'éléments susceptibles d'être pris en considération par les AC lors de l'évaluation de l'allégation d'anonymat d'un

[35] Arrêt de la CJUE du 7 mars 2024, affaire C-479/22 P, *OC contre Commission européenne* (ECLI:EU:C:2024:215), point 50.
[36] Arrêt de la CJUE du 7 mars 2024, affaire C-479/22 P, *OC contre Commission européenne* (ECLI:EU:C:2024:215), point 50.
[37] Arrêt de la CJUE du 19 octobre 2016, affaire C-582/14, *Breyer contre Bundesrepublik Deutschland* (ECLI:EU:C:2016:779), point 46, et arrêt de la CJUE du 7 mars 2024, affaire C-479/22 P, *OC contre Commission européenne* (ECLI:EU:C:2024:215), point 51.

responsable du traitement. D'autres approches peuvent être possibles si elles offrent un niveau de protection équivalent, en particulier si l'on tient compte de l'état de la technique.

48. La présence ou l'absence des éléments énumérés ci-dessous n'est pas un critère concluant pour évaluer l'anonymat d'un modèle d'IA.

3.2.2.1 Conception d'un modèle d'IA

49. En ce qui concerne la conception des modèles d'IA, les AC devraient évaluer les approches adoptées par les responsables du traitement au cours de la phase de développement. L'application et l'efficacité de quatre domaines clés (identifiés ci-dessous) devraient être prises en considération à cet égard.

Sélection des sources

50. Le premier domaine d'évaluation consiste à examiner la sélection des sources utilisées pour entraîner le modèle d'IA. Cela inclut une évaluation, par les AC, de toutes les mesures prises pour éviter ou limiter la collecte de données à caractère personnel, y compris, entre autres, i) le caractère approprié des critères de sélection, ii) la pertinence et l'adéquation des sources choisies compte tenu de la ou des finalités prévues; et iii) la question de savoir si des sources inappropriées ont été exclues.

Préparation et minimisation des données

51. Le deuxième domaine d'évaluation concerne la préparation des données en vue de la phase de formation. Les AC devraient en particulier examiner: (i) si l'utilisation de données anonymes et/ou de données à caractère personnel ayant fait l'objet d'une pseudonymisation a été envisagée; et ii) s'il a été décidé de ne pas recourir à de telles mesures, les raisons de cette décision, compte tenu de la finalité poursuivie; iii) les stratégies et techniques de minimisation des données employées pour limiter le volume de données à caractère personnel incluses dans le processus d'entraînement; et iv) tout processus de filtrage des données mis en œuvre avant l'entraînement du modèle et destiné à supprimer les données à caractère personnel non pertinentes.

Choix méthodologiques relatifs à l'entraînement

52. Le troisième domaine d'évaluation concerne la sélection de méthodes solides dans le développement de modèles d'IA. Les AC devraient évaluer les choix méthodologiques susceptibles de réduire ou d'éliminer de manière significative le caractère identifiable, y compris en examinant, entre autres: (i) si cette méthodologie utilise des méthodes de régularisation pour améliorer la généralisation du modèle et réduire les ajustements excessifs; et, surtout, (ii) si le responsable du traitement a mis en œuvre des techniques appropriées et efficaces de préservation de la vie privée (par exemple, le respect différentiel de la vie privée).

Mesures relatives aux résultats du modèle

53. Le dernier domaine de l'évaluation concerne toutes les méthodes ou mesures ajoutées au modèle d'IA lui-même qui ne peuvent avoir d'incidence sur le risque d'extraction directe de données à caractère personnel pour le modèle par toute personne accédant directement à celui-ci, mais qui pourraient réduire la probabilité d'obtenir des données à caractère personnel liées à des données d'entraînement à la suite de requêtes.

3.2.2.2 Analyse des modèles d'IA

54. Pour permettre aux AC d'évaluer la solidité du modèle d'IA conçu au regard de l'anonymisation, une première étape consiste à s'assurer que la conception a été développée comme prévu et qu'elle fait l'objet d'une gouvernance d'ingénierie efficace. Les AC devraient évaluer si les responsables du traitement ont réalisé des audits (internes ou externes) fondés sur des documents qui comprennent une évaluation des mesures choisies et de leur incidence afin de limiter la probabilité d'identification. Cela pourrait inclure l'analyse de rapports de réexamen des codes, ainsi qu'une analyse théorique

documentant le caractère approprié des mesures choisies pour réduire la probabilité de ré-identification du modèle concerné.

3.2.2.3 Essais des modèles d'IA et résistance aux attaques

55. Enfin, les AC devraient tenir compte de la portée, de la fréquence, de la quantité et de la qualité des essais que le responsable du traitement a effectués sur le modèle. En particulier, les AC devraient tenir compte du fait que des essais réussis couvrant des attaques largement connues et à la pointe de la technologie ne peuvent que constituer une preuve de la résistance à ces attaques. À la date du présent avis, cela pourrait inclure, entre autres, des tests structurés contre: i) les inférences d'appartenance et de membres; ii) l'exfiltration; iii) la régurgitation des données d'entraînement; iv) l'inversion de modèle; ou v) les attaques de reconstruction.

3.2.2.4 Documentation

56. Les articles 5, 24, 25 et 30 du RGPD et, en cas de risque élevé probable pour les droits et libertés des personnes concernées, l'article 35 du RGPD, exigent des responsables du traitement qu'ils documentent de manière adéquate leurs opérations de traitement. Cela s'applique également à tout traitement qui inclurait l'entraînement d'un modèle d'IA, même si l'objectif du traitement est l'anonymisation. Les AC devraient tenir compte de cette documentation ainsi que de toute évaluation régulière des risques qui en découlent pour le traitement effectué par les responsables du traitement, car il s'agit d'étapes fondamentales pour démontrer que les données à caractère personnel ne font pas l'objet d'un traitement.

57. **L'EDPB estime que les AC devraient tenir compte de la documentation chaque fois qu'une revendication d'anonymat concernant un modèle d'IA donné doit être évaluée. L'EDPB note que, si une AC n'est pas en mesure de confirmer, après avoir évalué la demande d'anonymat, y compris à la lumière de la documentation, que des mesures efficaces ont été prises pour anonymiser le modèle d'IA, l'AC serait en mesure de considérer que le responsable du traitement n'a pas respecté ses obligations en matière de responsabilité au titre de l'article 5, paragraphe 2, du RGPD. Par conséquent, le respect d'autres dispositions du RGPD devrait également être pris en considération.**

58. Idéalement, les AC devraient vérifier si la documentation du responsable du traitement comprend:

    a. toute information relative aux AIPD, y compris les évaluations et les décisions qui ont déterminé qu'une AIPD n'était pas nécessaire;

    b. tout conseil ou retour d'information fourni par le délégué à la protection des données («**DPD**») (lorsqu'un DPD a été - ou aurait dû être - désigné);

    c. des informations sur les mesures techniques et organisationnelles prises lors de la conception du modèle d'IA afin de réduire la probabilité de l'identification, y compris le modèle de menace et les évaluations des risques sur lesquels ces mesures sont fondées. Elle doit comprendre les mesures spécifiques à chaque source des jeux de données d'entraînement, y compris les URL des sources pertinentes et les descriptions des mesures prises (ou déjà prises par les tiers fournisseurs de jeux de données);

    d. les mesures techniques et organisationnelles prises à toutes les étapes du cycle de vie du modèle, qui, soit ont contribué à l'absence de données à caractère personnel dans le modèle, soit ont permis de vérifier l'absence de données à caractère personnel dans le modèle;

    e. la documentation démontrant la résistance théorique du modèle d'IA aux techniques de ré-identification, ainsi que les contrôles conçus pour limiter ou évaluer le succès et l'impact des attaques principales (régurgitation, attaques par inférence d'appartenance, exfiltration, etc.).

Cela peut notamment inclure: (i) le rapport entre la quantité de données d'apprentissage et le nombre de paramètres dans le modèle, y compris l'analyse de son incidence sur le modèle[38]; (ii) des mesures sur la probabilité de ré-identification sur la base de l'état actuel de la technique; (iii) des rapports sur la manière dont le modèle a été testé (par qui, quand, comment et dans quelle mesure) et (iv) les résultats des tests;

f. la documentation fournie au(x) responsable(s) du traitement déployant le modèle et/ou aux personnes concernées, en particulier la documentation relative aux mesures prises pour réduire la probabilité d'identification et concernant les éventuels risques résiduels.

3.3 Sur le caractère approprié de l'intérêt légitime en tant que base juridique pour le traitement des données à caractère personnel dans le contexte du développement et du déploiement de modèles d'IA

59. Pour répondre aux questions 2 et 3 de la demande, l'EDPB formulera d'abord des observations générales sur certains aspects importants que les AC devraient prendre en compte, quelle que soit la base juridique du traitement, lorsqu'elles évaluent la manière dont les responsables du traitement peuvent démontrer la conformité avec le RGPD dans le contexte des modèles d'IA. L'EDPB, en s'appuyant sur les lignes directrices 1/2024 relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD[39], examinera ensuite les trois étapes requises par l'évaluation de l'intérêt légitime dans le contexte du développement et du déploiement de modèles d'IA.

3.3.1 Observations générales

60. L'EDPB rappelle que le RGPD n'établit aucune hiérarchie entre les différentes bases juridiques prévues à l'article 6, paragraphe 1, du RGPD[40].

61. L'article 5 du RGPD fixe les principes relatifs au traitement des données à caractère personnel. L'EPDB met en évidence ceux qui sont importants pour le présent avis et devraient au moins être pris en considération par les AC lors de l'évaluation de modèles d'IA spécifiques, ainsi que les exigences les plus pertinentes découlant d'autres dispositions du RGPD, compte tenu du champ d'application du présent avis.

62. **Principe de responsabilité** (article 5, paragraphe 2, du RGPD) - Ce principe prévoit que le responsable du traitement est responsable du respect du RGPD et est en mesure de le démontrer. À cet égard, les rôles et responsabilités des parties qui traitent des données à caractère personnel dans le cadre du développement ou du déploiement d'un modèle d'IA devraient être évalués avant que le traitement n'ait lieu, afin de définir d'emblée les obligations des responsables du traitement ou des responsables conjoints du traitement et des sous-traitants (le cas échéant).

[38] Ricciato F., *A Cautionary Reflection on (Pseudo-)Synthetic Data from Deep Learning on Personal Data*, Privacy in Statistical Databases conference (PSD 2024), Antibes, France, septembre 2024, diapositives disponibles à l'adresse suivante: https://cros.ec.europa.eu/system/files/2024-10/20240926_PSD2024_Ricciato_v6_1.pdf et Belkin M., Hsu D., Ma S., & Mandal S. (2019), *Reconciling modern machine-learning practice and the classical bias-variance trade-off*. Proceedings of the National Academy of Sciences, 24 juillet 2019, 116(32) 15849-15854, disponible à l'adresse suivante: https://www.pnas.org/doi/10.1073/pnas.1903070116

[39] Voir les lignes directrices 1/2024 de l'EDPB sur le traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024.

[40] Ibidem, point 1.

63. **Principes de licéité, d'équité et de transparence** [article 5, paragraphe 1, point a), du RGPD] - Lors de l'évaluation de la licéité du traitement dans le contexte des modèles d'IA, à la lumière de l'article 6, paragraphe 1, du RGPD, l'EDPB estime qu'il est utile de distinguer les différentes étapes du traitement des données à caractère personnel[41]. Le principe d'équité, qui est étroitement lié au principe de transparence, exige que les données à caractère personnel ne soient pas traitées par des méthodes déloyales ou par tromperie, ou d'une manière qui soit «*injustifiablement préjudiciable ou illégalement discriminatoire, inattendue ou trompeuse pour la personne concernée*»[42]. Compte tenu de la complexité des technologies concernées, les informations sur le traitement des données à caractère personnel dans le cadre des modèles d'IA devraient donc être fournies de manière accessible, compréhensible et conviviale[43]. La transparence concernant le traitement des données à caractère personnel inclut, en particulier, le respect des obligations d'information énoncées aux articles 12 à 14 du RGPD[44], qui exigent également, en cas de prise de décision automatisée, y compris le profilage, des informations utiles relatives à la logique sous-jacente, ainsi que l'importance et les conséquences envisagées du traitement pour la personne concernée[45]. Sachant que les phases de développement des modèles d'IA peuvent impliquer la collecte de grandes quantités de données à partir de sources accessibles au public (par exemple, au moyen de techniques de «*web scraping*»), le recours à l'exception prévue à l'article 14, paragraphe 5, point b), du RGPD est strictement limité aux cas où les exigences de cette disposition sont pleinement remplies[46].

64. **Principes de limitation des finalités et de minimisation des données** [article 5, paragraphe 1, points b) et c), du RGPD] — Conformément au principe de minimisation des données, le développement et le déploiement de modèles d'IA exigent que les données à caractère personnel soient adéquates, pertinentes et nécessaires au regard de la finalité. Cela peut inclure le traitement de données à caractère personnel destiné à éviter les risques de biais et d'erreurs potentiels lorsque cela est clairement et spécifiquement identifié dans le cadre de la finalité, et que les données à caractère personnel sont nécessaires à cette fin (par exemple, lorsqu'elles ne peuvent être obtenues efficacement par le traitement d'autres données, y compris des données synthétiques ou anonymisées)[47]. Le groupe de travail «article 29» a déjà souligné que «*la finalité de la collecte doit être clairement et spécifiquement identifiée* [...]»[48]. Pour déterminer si la finalité poursuivie est légitime, spécifique et explicite, et si le traitement est conforme au principe de minimisation des données, il

[41] Rapport du comité européen de la protection des données sur les travaux entrepris par le groupe de travail ChatGPT, adopté le 23 mai 2024, point 14.

[42] Rapport du comité européen de la protection des données sur les travaux entrepris par le groupe de travail ChatGPT, adopté le 23 mai 2024, point 23; lignes directrices 4/2019 de l'EDPB relatives à l'article 25, Protection des données dès la conception et protection des données par défaut, version 2.0, adoptées le 20 octobre 2020, point 69; lignes directrices du groupe de travail «Article 29» sur la transparence en vertu du règlement 2016/679, version révisée et adoptée le 11 avril 2018, approuvée par l'EDPB le 25 mai 2018, point 2.

[43] Lignes directrices du groupe de travail «Article 29» sur la transparence au titre du règlement (UE) 2016/679, version révisée et adoptée le 11 avril 2018, approuvée par l'EDPB le 25 mai 2018, point 5.

[44] Soit également le considérant 39 du GDPR, qui stipule que «*[l]e fait que des données à caractère personnel concernant des personnes physiques sont collectées, utilisées, consultées ou traitées d'une autre manière et la mesure dans laquelle ces données sont ou seront traitées devraient être transparents à l'égard des personnes physiques concernées [...]*».

[45] Article 13, paragraphe 2, point f), du RGPD et article 14, paragraphe 2, point g), du RGPD.

[46] Rapport du comité européen de la protection des données sur les travaux entrepris par le groupe de travail ChatGPT, adopté le 23 mai 2024, point 27.

[47] En outre, l'article 10, paragraphe 5, du règlement sur l'IA prévoit des règles spécifiques pour le traitement de catégories particulières de données à caractère personnel en rapport avec les systèmes d'IA à haut risque, afin de garantir la détection et la correction des biais.

[48] Avis 03/2013 du groupe de travail «Article 29» sur la limitation de la finalité (WP203), pages 15 et 16.

convient d'identifier d'abord l'activité de traitement en cause. Les différentes étapes des phases de développement ou de déploiement peuvent notamment constituer des activités de traitement identiques ou différentes et impliquer des responsables du traitement successifs ou des responsables conjoints du traitement. Dans certains cas, il est possible de déterminer la finalité qui sera poursuivie pendant le déploiement du modèle d'IA à un stade précoce de son développement. Même lorsque tel n'est pas le cas, le contexte de ce déploiement devrait déjà être clair et, par conséquent, la manière dont ce contexte contribue à la finalité du développement devrait être prise en considération. Lors de l'examen de la finalité du traitement à un stade donné de développement, les AC devraient s'attendre à un certain degré de détail de la part du ou des responsables du traitement, ainsi qu'à une explication de la manière dont ces détails éclairent la finalité du traitement. Il peut s'agir, par exemple, d'informations sur le type de modèle d'IA développé, ses fonctionnalités attendues et tout autre contexte pertinent déjà connu à ce stade. Le contexte du déploiement pourrait également inclure, par exemple, la question de savoir si un modèle est en cours d'élaboration pour un déploiement interne, si le responsable du traitement a l'intention de vendre ou de distribuer le modèle à des tiers après son développement, et notamment si le modèle est principalement destiné à être déployé à des fins de recherche ou à des fins commerciales.

65. **Droits des personnes concernées** (chapitre III du RGPD) — Nonobstant la nécessité pour les AC de veiller à ce que tous les droits des personnes concernées soient respectés lorsque des modèles d'IA sont élaborés et déployés par les responsables du traitement, l'EDPB rappelle que chaque fois que l'intérêt légitime est invoqué en tant que base juridique par un responsable du traitement, le droit d'opposition au titre de l'article 21 du RGPD s'applique et devrait être garanti[49].

3.3.2 Considérations relatives aux trois étapes de l'évaluation de l'intérêt légitime dans le contexte du développement et du déploiement de modèles d'IA

66. Afin de déterminer si un traitement donné de données à caractère personnel peut être fondé sur l'article 6, paragraphe 1, point f), du RGPD, les AC devraient vérifier que les responsables du traitement ont soigneusement évalué et documenté si les trois conditions cumulatives suivantes sont remplies: (i) le responsable du traitement, ou un tiers, poursuit un intérêt légitime; (ii) le traitement est nécessaire à la poursuite de l'intérêt légitime; et (iii) les intérêts ou les libertés et droits fondamentaux des personnes concernées ne prévalent pas sur l'intérêt légitime[50].

[49] Conformément à l'article 21 du RGPD, si une personne concernée s'oppose, pour des motifs liés à sa situation particulière, au traitement de données à caractère personnel la concernant, le responsable du traitement ne traite plus les données à caractère personnel, à moins que le responsable du traitement ne démontre l'existence de motifs légitimes et impérieux justifiant le traitement, qui prévalent sur les intérêts, droits et libertés de la personne concernée ou pour la constatation, l'exercice ou la défense de droits en justice. Par conséquent, les deux aspects à prendre en considération par les AC sont: le responsable du traitement est-il en mesure de démontrer l'existence de ces motifs légitimes et impérieux, et le droit d'opposition peut-il être exercé?

[50] CJUE, arrêt du 4 juillet 2023, affaire C-252/21, *Meta contre Bundeskartellamt* (ECLI:EU:C:2023:537), point 106; CJUE, arrêt du 11 décembre 2019, affaire C-708/18, *Asociaţia de Proprietari bloc M5A-ScaraA* (ECLI:EU:C:2019:1064), point 40. Voir également les lignes directrices 1/2024 du CEPD relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, points 12 et suivants. Comme rappelé dans lesdites lignes directrices, cette «*évaluation devrait être effectuée dès le début du traitement, avec la participation du délégué à la protection des données (DPD) (s'il est désigné), et devrait être documentée par le responsable du traitement conformément au principe de responsabilité énoncé à l'article 5, paragraphe 2, du RGPD*».

3.3.2.1 *Première étape - Poursuite d'un intérêt légitime par le responsable du traitement ou par un tiers*

67. L'intérêt est l'enjeu ou le bénéfice plus large qu'un responsable du traitement ou un tiers peut tirer de l'exercice d'une activité de traitement spécifique[51]. Bien que le RGPD et la CJUE aient reconnu plusieurs intérêts comme étant légitimes[52], l'évaluation de la légitimité d'un intérêt donné devrait être le résultat d'une analyse au cas par cas.

68. Comme le rappelle l'EDPB dans ses lignes directrices relatives à l'intérêt légitime[53], un intérêt peut être considéré comme légitime si les trois critères cumulatifs suivants sont remplis:

    a. L'intérêt est licite[54];

    b. L'intérêt est formulé de manière claire et précise; et

    c. L'intérêt est réel et présent, et non spéculatif.

69. Sous réserve des deux autres étapes requises par l'évaluation de l'intérêt légitime, les exemples suivants peuvent constituer un intérêt légitime dans le contexte des modèles d'IA: i) développement d'un service d'agent conversationnel pour aider les utilisateurs; ii) développement d'un système d'IA pour détecter les contenus ou comportements frauduleux; et iii) amélioration de la détection des menaces dans un système d'information.

3.3.2.2 *Deuxième étape – Analyse de la nécessité du traitement aux fins de la poursuite de l'intérêt légitime*

70. La deuxième étape de l'évaluation consiste à déterminer si le traitement des données à caractère personnel est nécessaire aux fins des intérêts légitimes poursuivis ([55]) («test de nécessité»).

71. Le considérant 39 du RGPD précise que «*[l]es données à caractère personnel ne devraient être traitées que si la finalité du traitement ne peut être raisonnablement atteinte par d'autres moyens*». Selon la CJUE et les orientations antérieures de l'EDPB, la condition relative à la nécessité du traitement doit être examinée à la lumière des libertés et droits fondamentaux des personnes concernées et en liaison avec le principe de minimisation des données énoncé à l'article 5, paragraphe 1, point c), du RGPD[56].

---

[51] Lignes directrices 1/2024 du CEPD relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, point 14.

[52] Lignes directrices 1/2024 du CEPD relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, point 16.

[53] Lignes directrices 1/2024 du CEPD relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, point 17.

[54] CJUE, arrêt du 4 octobre 2024, affaire C-621/22, *Koninklijke Nederlandse Lawn Tennisbond* (ECLI:EU:C:2024:857), point 49, dans lequel la CJUE a souligné qu'un intérêt légitime ne saurait être contraire au droit. À cet égard, l'EDPB souligne que, le cas échéant, les cadres législatifs devraient être pris en compte lors de l'appréciation de la licéité d'un intérêt donné. Voir par exemple: Article 26, paragraphe 3, et article 28 du règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (loi sur les services numériques) («DSA») sur la publicité ciblée interdite aux mineurs; article 5, paragraphes 1 et 2, du règlement sur l'IA sur les pratiques interdites en matière d'IA (pratiques manipulatrices et subliminales); le traitement en violation des droits de propriété intellectuelle et des dispositions de la directive (UE) 2019/790 sur le droit d'auteur et les droits voisins dans le marché unique numérique.

[55] Lignes directrices 1/2024 du CEPD sur le traitement des données à caractère personnel fondées sur l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, points 28 à 30.

[56] CJUE, arrêt du 4 juillet 2023, affaire C-252/21, *Meta contre Bundeskartellamt* (ECLI:EU:C:2023:537), points 108 et 109, renvoyant également à la CJUE, arrêt du 11 décembre 2019, affaire C-708/18, *Asociaţia de Proprietari*

72. La méthodologie mentionnée par la CJUE prend en compte le contexte du traitement, ainsi que les effets sur le responsable du traitement et sur les personnes concernées. L'appréciation de la nécessité comporte donc deux éléments: i) la question de savoir si l'activité de traitement permettra la poursuite de la finalité[57]; et ii) la question de savoir s'il n'existe pas de manière moins intrusive de poursuivre cette finalité[58].

73. Par exemple, et selon le cas, le volume prévu de données à caractère personnel impliqué dans le modèle d'IA doit être évalué à la lumière d'alternatives moins intrusives qui peuvent raisonnablement être disponibles pour atteindre tout aussi efficacement l'objectif de l'intérêt légitime poursuivi. Si la poursuite de la finalité est également possible au moyen d'un modèle d'IA qui n'implique pas le traitement de données à caractère personnel, le traitement de données à caractère personnel devrait être considéré comme non nécessaire. Ceci est particulièrement important pour le développement de modèles d'intelligence artificielle. Lorsqu'elles évaluent si la condition de nécessité est remplie, les AC devraient accorder une attention particulière à la quantité de données à caractère personnel traitées et à la proportionnalité de la poursuite de l'intérêt légitime en jeu, y compris à la lumière du principe de minimisation des données.

74. L'évaluation de la nécessité devrait également tenir compte du contexte plus large du traitement envisagé des données à caractère personnel. L'existence de moyens moins intrusifs pour les libertés et droits fondamentaux des personnes concernées peut varier selon que le responsable du traitement a une relation directe avec les personnes concernées (données de première partie) ou non (données de tierce partie). La CJUE a formulé certaines considérations à prendre en compte lors de l'analyse de

bloc M5A-ScaraA (ECLI:EU:C:2019:1064), point 48; CJUE, arrêt du 9 novembre 2010, affaires jointes C-92/09 et C-93/09, Volker und Markus Schecke (ECLI:EU:C:2010:662), points 85 et 86; CJUE, arrêt du 22 juin 2021, affaire C-439/19, *Latvijas Republikas Saeima* (ECLI:EU:C:2021:504), points 98, 109, 110, 113. Voir, aussi, par exemple: Lignes directrices 3/2019 de l'EDPB relatives au traitement des données à caractère personnel au moyen de dispositifs vidéo, version 2.0, adoptée le 29 janvier 2020, paragraphes 24-26 et 73; lignes directrices 2/2019 de l'EDPB relatives au traitement des données à caractère personnel en vertu de l'article 6, paragraphe 1, point b), du RGPD dans le cadre de la fourniture de services en ligne aux personnes concernées, version 2.0, adoptée le 8 octobre 2019, points 23-25; avis 11/2024 de l'EDPB sur l'utilisation de la reconnaissance faciale pour rationaliser le flux des passagers dans les aéroports, version 1.1, adoptée le 23 mai 2024, point 27.

[57] Voir CJUE, arrêt du 16 décembre 2008, affaire C-524/06, *Heinz Huber contre Bundesrepublik Deutschland* (ECLI:EU:C:2008:724), point 66. Dans la même affaire, voir également les conclusions de l'avocat général Poiares Maduro dans l'affaire C-524/06, *Heinz Huber contre Bundesrepublik Deutschland* (ECLI:EU:C:2008:194), point 16, dans lesquelles il est indiqué: «*le critère approprié en l'occurrence est un critère d'efficacité et c'est à la juridiction nationale qu'il appartient de l'appliquer. La question qu'elle doit se poser consiste à s'interroger quant à l'existence d'autres modes de traitement des données qui permettraient aux autorités en charge de l'immigration d'assurer le respect des règles du statut de résidence. Si la juridiction de renvoi répond à cette question par l'affirmative, la conservation et le traitement centralisés des données relatives aux ressortissants de l'Union doivent être déclarés illégaux. Il n'est pas nécessaire que le système substitué soit le plus efficace ou le plus approprié; il suffit qu'il soit en mesure de fonctionner de manière adéquate. En d'autres termes, même si le registre central est plus efficace, plus commode ou plus facile à utiliser que les solutions de substitution (tels des registres décentralisés, locaux), ce substitut doit être préféré s'il permet de fournir les données relatives au statut de résidence des ressortissants de l'Union*».

[58] Voir CJUE, arrêt du 27 septembre 2017, affaire C-73/16, *Peter Puškár* (ECLI:EU:C:2017:725), point 113: «*Il appartient ainsi à la juridiction de renvoi de vérifier si l'établissement de la liste litigieuse et l'inscription sur celle-ci du nom des personnes concernées sont propres à réaliser les objectifs poursuivis par ceux-ci et s'il n'existe pas d'autres moyens moins contraignants afin d'atteindre ces objectifs*»; voir aussi, par exemple, les conclusions de l'avocat général Rantos dans l'affaire C-252/21, *Meta contre Bundeskartellamt*, ECLI:EU:C:2022:704, point 61, dans lesquelles il est indiqué: *«[...] Il faut donc qu'il y ait un lien étroit entre le traitement et l'intérêt poursuivi, en l'absence d'alternatives plus respectueuses de la protection des données à caractère personnel, car il ne suffit pas que le traitement relève d'une simple utilité pour le responsable du traitement»*.

la nécessité du traitement des données de première partie aux fins du ou des intérêts légitimes poursuivis (bien qu'elles concernent le cadre de la divulgation de ces données à des tiers)[59].

75. La mise en œuvre de garanties techniques visant à protéger les données à caractère personnel peut également contribuer à satisfaire au critère de nécessité. Il pourrait s'agir, par exemple, la mise en œuvre de mesures telles que celles mentionnées à la section 3.2.2 de telle sorte que l'anonymisation ne soit réalisée, mais en réduisant néanmoins la facilité avec laquelle les personnes concernées peuvent être identifiées. L'EDPB note que certaines de ces mesures, lorsqu'elles ne sont pas nécessaires pour se conformer au RGPD, peuvent constituer des garanties supplémentaires, comme cela est analysé plus en détail dans la sous-section «mesures d'atténuation», à la section 3.3.2.3[60].

3.3.2.3 Troisième étape - Test de mise en balance

76. La troisième étape de l'évaluation de l'intérêt légitime est l'«**exercice de mise en balance**» (également désigné dans le présent document par le terme «**test de mise en balance**»)[61]. Cette étape consiste à identifier et à décrire les différents droits et intérêts opposés en jeu[62], c'est-à-dire, d'une part, les intérêts, les libertés et droits fondamentaux des personnes concernées et, d'autre part, les intérêts du responsable du traitement ou d'un tiers. Les circonstances spécifiques de l'affaire devraient alors être prises en considération pour démontrer que l'intérêt légitime constitue une base juridique appropriée pour les activités de traitement en cause[63].

Intérêts, droits fondamentaux et libertés des personnes concernées

77. L'article 6, paragraphe 1, point f), du RGPD prévoit que, lors de l'évaluation des différents éléments dans le cadre du test de mise en balance, le responsable du traitement doit tenir compte des intérêts, des droits fondamentaux et des libertés des personnes concernées. Les intérêts des personnes concernées sont ceux qui peuvent être affectés par le traitement en cause. Dans le contexte de la phase de développement d'un modèle d'IA, ceux-ci peuvent inclure, sans s'y limiter, l'intérêt à l'autodétermination et au maintien du contrôle sur ses propres données à caractère personnel (par exemple, les données recueillies pour le développement du modèle). Dans le contexte du déploiement d'un modèle d'IA, les intérêts des personnes concernées peuvent inclure, sans s'y limiter, des intérêts à garder le contrôle de leurs propres données à caractère personnel (par exemple, les données traitées une fois le modèle déployé), des intérêts financiers (par exemple, lorsqu'un modèle d'IA est utilisé par la personne concernée pour générer des recettes, ou est utilisé par une personne dans le cadre de son activité professionnelle), des avantages personnels (par exemple, lorsqu'un modèle d'IA est utilisé pour améliorer l'accessibilité à certains services) ou des intérêts socio-économiques (par exemple, lorsqu'un modèle d'IA permet l'accès à de meilleurs soins de santé ou facilite l'exercice d'un droit fondamental tel que l'accès à l'éducation)[64].

---

[59] CJUE, arrêt du 4 octobre 2024, affaire C-621/22, *Koninklijke Nederlandse Lawn Tennisbond* (ECLI:EU:C:2024:857), points 51-53.

[60] Voir les lignes directrices de l'EDPB 1/2024 relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, point 57.

[61] Voir les lignes directrices 1/2024 de l'EDPB relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, point 31 à 60.

[62] Voir les lignes directrices de l'EDPB 1/2024 relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, point 32.

[63] Voir les lignes directrices 1/2024 de l'EDPB relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, point 32, faisant également référence à la CJUE, arrêt du 4 juillet 2023, affaire C-252/21, *Meta contre Bundeskartellamt* (ECLI:EU:C:2023:537), point 110.

[64] Voir les lignes directrices de l'EDPB 1/2024 relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, point 38.

78. Plus la définition d'un intérêt est précise à la lumière de la finalité du traitement, plus elle permettra d'appréhender clairement la réalité des avantages et des risques à prendre en compte dans le cadre du test de mise en balance.

79. En ce qui concerne les droits et libertés fondamentaux des personnes concernées, le développement et le déploiement de modèles d'IA peuvent présenter des risques sérieux pour les droits protégés par la charte des droits fondamentaux de l'Union européenne (la «**Charte de l'UE**»), y compris, mais sans s'y limiter, le droit à la vie privée et familiale (article 7 de la Charte de l'UE) et le droit à la protection des données à caractère personnel (article 8 de la Charte de l'UE). Ces risques peuvent survenir pendant la phase de développement, par exemple lorsque des données à caractère personnel sont supprimées contre les souhaits des personnes concernées ou à leur insu. Ces risques peuvent également survenir au cours de la phase de déploiement, par exemple lorsque des données à caractère personnel sont traitées par le modèle (ou dans le cadre de celui-ci) d'une manière qui va à l'encontre des droits des personnes concernées, ou lorsqu'il est possible de déduire, accidentellement ou par des attaques (par exemple, déduction de l'appartenance, extraction ou inversion du modèle), quelles données à caractère personnel sont contenues dans la base de données d'apprentissage. De telles situations présentent un risque pour la vie privée des personnes concernées dont les données pourraient apparaître au cours de la phase de déploiement du système d'IA (par exemple, risque de réputation, usurpation d'identité ou fraude, risque de sécurité en fonction de la nature des données).

80. En fonction de l'affaire en cause, il peut également y avoir des risques pour d'autres droits fondamentaux. Par exemple, la collecte de données à grande échelle et indifférenciée par des modèles d'IA au cours de la phase de développement peut créer un sentiment de surveillance chez les personnes concernées, compte tenu notamment des difficultés à empêcher que des données publiques ne soient effacées. Cela peut conduire les individus à s'autocensurer et présenter des risques d'atteinte à leur liberté d'expression (article 11 de la Charte de l'Union européenne). Au cours de la phase de déploiement, des risques pour la liberté d'expression sont également présents lorsque des modèles d'IA sont utilisés pour bloquer la publication de contenus par les personnes concernées. En outre, un modèle d'IA recommandant des contenus inappropriés à des personnes vulnérables peut présenter des risques pour leur santé mentale (article 3, paragraphe 1, de la Charte de l'UE). Dans d'autres cas, le déploiement de modèles d'IA peut également avoir des conséquences négatives sur le droit de l'individu à exercer un emploi (article 15 de la Charte de l'UE), par exemple lorsque des demandes d'emploi sont présélectionnées à l'aide d'un modèle d'IA. De la même manière, un modèle d'IA pourrait présenter des risques pour le droit à la non-discrimination (article 21 de la Charte de l'UE), s'il discrimine les individus sur la base de certaines caractéristiques personnelles (telles que la nationalité ou le sexe). En outre, le déploiement de modèles d'IA peut également présenter des risques pour la sécurité et la sûreté de l'individu (par exemple, lorsque le modèle d'IA est utilisé avec une intention malveillante), ainsi que des risques pour leur intégrité physique et mentale[65].

81. Le déploiement de modèles d'IA peut également avoir un impact positif sur certains droits fondamentaux, par exemple le modèle peut soutenir le droit à l'intégrité mentale de la personne (article 3 de la Charte), par exemple lorsqu'un modèle d'IA est utilisé pour identifier des contenus préjudiciables en ligne; ou alors, le modèle peut faciliter l'accès à certains services essentiels ou faciliter l'exercice de droits fondamentaux, tels que l'accès à l'information (article 11 de la Charte de l'UE) ou l'accès à l'éducation (article 14 de la Charte de l'UE).

[65] Lignes directrices 1/2024 sur le traitement des données à caractère personnel fondées sur l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, paragraphe 46.

Impact du traitement sur les personnes concernées

82. Le traitement des données à caractère personnel effectué au cours du développement et du déploiement des modèles d'IA peut avoir sur les personnes concernées différentes incidences, qui peuvent être positives ou négatives[66]. Par exemple, si une activité de traitement comporte des avantages pour la personne concernée, ceux-ci peuvent être pris en compte dans le test de mise en balance. Si l'existence de tels avantages peut conduire une AC à conclure que les intérêts, les libertés et droits fondamentaux des personnes concernées ne prévalent pas sur les intérêts du responsable du traitement ou d'un tiers, cette conclusion ne peut résulter que d'une analyse au cas par cas tenant compte de tous les facteurs appropriés.

83. L'impact du traitement sur les personnes concernées peut être influencé par (i) la nature des données traitées par les modèles; (ii) le contexte du traitement; et (iii) les conséquences ultérieures que le traitement peut avoir[67].

84. En ce qui concerne la **nature des données traitées**, il convient de rappeler que, hormis les catégories particulières de données à caractère personnel et de données relatives à des condamnations pénales et à des infractions, qui bénéficient respectivement d'une protection supplémentaire au titre des articles 9 et 10 du RGPD, le traitement de certaines autres catégories de données à caractère personnel peut entraîner des conséquences importantes pour les personnes concernées. Dans ce contexte, le traitement de certains types de données à caractère personnel qui révèlent des informations hautement confidentielles (par exemple, des données financières ou des données de localisation) aux fins du développement et du déploiement d'un modèle d'IA devrait être considéré comme susceptible d'avoir une incidence grave sur les personnes concernées. Au cours de la phase de déploiement, les conséquences de ce traitement pour les personnes concernées peuvent être, par exemple, d'ordre économique (par exemple, la discrimination dans le contexte de l'emploi) et/ou en termes de réputation (par exemple, la diffamation).

85. En ce qui concerne le **contexte du traitement,** il est tout d'abord nécessaire d'identifier les éléments susceptibles de créer des risques pour les personnes concernées (par exemple, la manière dont le modèle a été développé, la manière dont le modèle peut être déployé et/ou si les mesures de sécurité utilisées pour protéger les données à caractère personnel sont appropriées). La nature du modèle et les utilisations opérationnelles prévues jouent un rôle clé dans l'identification de ces causes potentielles.

86. Il est également nécessaire d'évaluer la gravité de ces risques pour les personnes concernées. Il peut être tenu compte, entre autres, de la manière dont les données à caractère personnel sont traitées (par exemple, si elles sont combinées avec d'autres jeux de données), de l'ampleur du traitement et de la quantité de données à caractère personnel traitées[68] (par exemple, le volume global de données, le volume de données par personne concernée, le nombre de personnes concernées)[69], du statut de la personne concernée (par exemple, des enfants ou d'autres personnes vulnérables) et de sa relation avec le responsable du traitement (par exemple, si la personne concernée est un client). Par exemple, l'utilisation d'un outil de «*web scraping*» au cours de la phase de développement peut, en l'absence

[66] Voir les lignes directrices 1/2024 de l'EDPB relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, point 39.
[67] Voir les lignes directrices 1/2024 de l'EDPB relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, point 32.
[68] Voir les lignes directrices de l'EDPB 1/2024 relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, point 43.
[69] CJUE, arrêt du 4 juillet 2023, affaire C-252/21, *Meta contre Bundeskartellamt* (ECLI:EU:C:2023:537), paragraphe 116.

de garanties suffisantes, avoir des incidences significatives sur les personnes, en raison du volume important de données collectées, du grand nombre de personnes concernées et de la collecte indifférenciée de données à caractère personnel.

87. Les **autres conséquences** que le traitement peut avoir devraient également être prises en considération lors de l'évaluation de l'incidence du traitement sur les personnes concernées. Elles devraient être évaluées par les AC au cas par cas, en tenant compte des faits spécifiques en l'espèce.

88. Ces conséquences peuvent inclure (mais ne se limitent pas à) des risques de violation des droits fondamentaux des personnes concernées, tels que décrits dans la sous-section précédente[70]. Les risques peuvent varier en termes de probabilité et de gravité, et peuvent résulter d'un traitement de données à caractère personnel susceptible d'entraîner des dommages physiques, matériels ou non matériels, en particulier lorsque le traitement peut donner lieu à des discriminations[71].

89. Lorsque le déploiement d'un modèle d'IA implique le traitement de données à caractère personnel tant i) des personnes concernées dont les données à caractère personnel sont incluses dans le jeu de données utilisé au cours de la phase de développement, que ii) des personnes concernées dont les données à caractère personnel sont traitées au cours de la phase de déploiement, les AC devraient distinguer et prendre en considération les risques affectant les intérêts, les droits et les libertés de chacune de ces catégories de personnes concernées lors de la vérification du test de mise en balance effectué par un responsable du traitement.

90. **Enfin, l'analyse des éventuelles conséquences ultérieures du traitement devrait également tenir compte de la probabilité que ces conséquences supplémentaires se matérialisent**. L'évaluation de cette probabilité devrait être effectuée en tenant compte des mesures techniques et organisationnelles en place et des circonstances spécifiques du cas d'espèce. Par exemple, les AC peuvent examiner si des mesures ont été mises en œuvre pour éviter une utilisation abusive potentielle du modèle d'IA. Pour les modèles d'IA qui peuvent être déployés à des fins diverses, comme l'IA générative, il peut s'agir de contrôles limitant autant que possible leur utilisation pour des pratiques préjudiciables, par exemple: la création de «deepfakes» (hypertrucage); les «chatbots» (agents conversationnels) utilisés pour la désinformation, le «phishing» ainsi que d'autres types de fraude; et l'IA ou les agents d'IA manipulateurs (en particulier lorsqu'ils sont anthropomorphes ou fournissent des informations trompeuses).

Attentes raisonnables des personnes concernées

91. Aux termes du considérant 47 du RGPD, «*[e]n tout état de cause, l'existence d'un intérêt légitime devrait faire l'objet d'une évaluation attentive, notamment afin de déterminer si une personne concernée peut raisonnablement s'attendre, au moment et dans le cadre de la collecte des données à caractère personnel, à ce que celles-ci fassent l'objet d'un traitement à une fin donnée*». *Les intérêts et droits fondamentaux de la personne concernée pourraient, en particulier, prévaloir sur l'intérêt du responsable du traitement lorsque des données à caractère personnel sont traitées dans des circonstances où les personnes concernées ne s'attendent raisonnablement pas à un traitement ultérieur*[72]».

[70] Voir la sous-section «Intérêts, libertés et droits fondamentaux des personnes concernées» ci-dessus.

[71] Voir section 2.3 des lignes directrices 1/2024 du comité européen de la protection des données relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024. Voir également le considérant 75 du RGPD pour d'autres exemples.

[72] Voir également CJUE, arrêt du 4 juillet 2023, affaire C-252/21, *Meta contre Bundeskartellamt* (ECLI:EU:C:2023:537), point 112; CJUE, arrêt du 11 décembre 2019, affaire C-708/18, *Asociaţia de Proprietari*

92. Les attentes raisonnables jouent un rôle clé dans le test de mise en balance, notamment en raison de la complexité de la technologie utilisée dans les modèles d'IA, et du fait qu'il peut être difficile pour les personnes concernées de comprendre la diversité des utilisations potentielles d'un modèle d'IA et le traitement des données concerné[73]. À cette fin, les informations fournies aux personnes concernées peuvent être prises en compte pour évaluer si les personnes concernées peuvent raisonnablement s'attendre à ce que leurs données à caractère personnel soient traitées. Toutefois, si l'omission d'informations peut contribuer à ce que les personnes concernées ne s'attendent pas à un certain traitement, le simple respect des exigences de transparence énoncées dans le RGPD n'est pas suffisant en soi pour considérer que les personnes concernées peuvent raisonnablement s'attendre à un certain traitement[74]. En outre, le simple fait que des informations relatives à la phase de développement d'un modèle d'IA soient incluses dans la politique de confidentialité du responsable du traitement ne signifie pas nécessairement que les personnes concernées peuvent raisonnablement s'attendre à ce que cela se produise; au contraire, cela devrait être analysé par les AC en fonction des circonstances spécifiques de l'affaire et en tenant compte de tous les facteurs pertinents.

93. Lors de l'évaluation des attentes raisonnables des personnes concernées en ce qui concerne le traitement effectué pendant la phase de développement, il est important de se référer aux éléments mentionnés dans les lignes directrices de l'EDPB sur l'intérêt légitime[75]. En outre, dans le cadre de l'objet du présent avis, il est important de prendre en considération le contexte plus large du traitement.- Il peut s'agir, entre autres, de la question de savoir si les données à caractère personnel étaient ou non accessibles au public, de la nature de la relation entre la personne concernée et le responsable du traitement (et s'il existe un lien entre les deux), de la nature du service, du contexte dans lequel les données à caractère personnel ont été collectées, de la source à partir de laquelle les données ont été collectées (par exemple, le site web ou le service sur lequel les données à caractère personnel ont été collectées et les paramètres de confidentialité qu'il propose), des utilisations ultérieures potentielles du modèle et de la question de savoir si les personnes concernées sont réellement conscientes que leurs données à caractère personnel se trouvent en ligne.

94. Au cours de la phase de développement du modèle, les attentes raisonnables des personnes concernées peuvent varier selon que les données traitées pour développer le modèle sont ou non rendues publiques par les personnes concernées. En outre, les attentes raisonnables peuvent également varier selon qu'elles ont directement fourni les données au responsable du traitement (par exemple, dans le cadre de leur utilisation du service) ou si le responsable du traitement les a obtenues auprès d'une autre source (par exemple, par l'intermédiaire d'un tiers, ou par *scraping*). Dans les deux cas, les mesures prises pour informer les personnes concernées des activités de traitement devraient être prises en considération lors de l'évaluation des attentes raisonnables.

---

bloc M5A-ScaraA (ECLI:EU:C:2019:1064), point 58; CJUE, arrêt du 4 octobre 2024, affaire C-621/22, *Koninklijke Nederlandse Lawn Tennisbond* (ECLI:EU:C:2024:857), point 55.

[73] Par exemple, dans l'arrêt du 4 juillet 2023, affaire C-252/21, *Meta contre Bundeskartellamt* (ECLI:EU:C:2023:537), point 123, si la CJUE a conclu que l'«amélioration du produit» ne saurait en principe être exclue en tant qu'intérêt légitime, elle a également conclu qu'il «*apparaît douteux que [...] l'objectif visant l'amélioration du produit puisse, compte tenu de l'ampleur de ce traitement et de l'impact important de celui-ci sur l'utilisateur, ainsi que de la circonstance que ce dernier ne saurait raisonnablement s'attendre à ce que ces données soient traitées [...], prévaloir sur les intérêts et les droits fondamentaux d'un tel utilisateur, d'autant plus dans l'hypothèse où celui-ci est un enfant*».

[74] Lignes directrices 1/2024 sur le traitement des données à caractère personnel fondées sur l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, point 53.

[75] Lignes directrices 1/2024 relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, points 50 à 54.

95. Lors de la phase de déploiement du modèle d'IA, il est tout aussi important de tenir compte des attentes raisonnables des personnes concernées dans le contexte des capacités spécifiques du modèle. Par exemple, pour les modèles d'IA capables de s'adapter en fonction des données d'entrée fournies, il peut être utile d'examiner si les personnes concernées savaient qu'elles avaient fourni des données à caractère personnel afin que le modèle d'IA puisse adapter ses réponses à leurs besoins et qu'elles puissent obtenir des services sur mesure. En outre, il peut également être utile d'examiner si cette activité de traitement n'aura une incidence que sur le service fourni aux personnes concernées (par exemple, la personnalisation du contenu pour un utilisateur spécifique) ou si elle sera utilisée pour modifier le service fourni à tous les clients (par exemple, pour améliorer le modèle d'une manière générale). Comme dans la phase de développement, il peut également être particulièrement pertinent d'examiner s'il existe un lien direct entre les personnes concernées et le responsable du traitement. Un tel lien direct peut, par exemple, permettre au responsable du traitement de fournir facilement aux personnes concernées des informations sur l'activité de traitement et le modèle, ce qui pourrait alors influencer les attentes raisonnables de ces personnes.

Mesures d'atténuation

96. Lorsque les intérêts, droits et libertés des personnes concernées semblent prévaloir sur le ou les intérêts légitimes poursuivis par le responsable du traitement ou par un tiers, le responsable du traitement peut envisager d'introduire des mesures d'atténuation visant à limiter l'incidence du traitement sur ces personnes concernées. Les mesures d'atténuation sont des garanties qui devraient être adaptées aux circonstances de l'espèce et dépendent de différents facteurs, y compris de l'utilisation prévue du modèle d'IA. Ces mesures d'atténuation viseraient à garantir que les intérêts du responsable du traitement ou du tiers ne seront pas écartés, de sorte que le responsable du traitement puisse s'appuyer sur cette base juridique.

97. Comme rappelé dans les lignes directrices de l'EDPB relatives à l'intérêt légitime, les mesures d'atténuation ne devraient pas être confondues avec les mesures que le responsable du traitement est légalement tenu d'adopter de toute manière pour garantir le respect du RGPD, indépendamment de la question de savoir si le traitement est fondé sur l'article 6, paragraphe 1, point f), du RGPD[76]. Ceci est particulièrement important pour les mesures qui, par exemple, doivent se conformer aux principes du RGPD, tels que le principe de minimisation des données.

98. La liste de mesures fournie ci-dessous est non exhaustive et non prescriptive et la mise en œuvre des mesures doit être envisagée au cas par cas. Si, en fonction des circonstances, certaines des mesures ci-dessous peuvent être requises pour se conformer à des obligations spécifiques du RGPD, elles peuvent, lorsque ce n'est pas le cas, être prises en compte en tant que garanties supplémentaires. En outre, certaines des mesures mentionnées ci-dessous concernent des domaines qui connaissent une évolution rapide et de nouveaux développements, et devraient être prises en considération par les AC lors du traitement d'un cas spécifique.

99. **En ce qui concerne la phase de développement des modèles d'IA**, plusieurs mesures peuvent être prises pour atténuer les risques posés par le traitement des données de première partie et de tiers (y compris pour atténuer les risques liés aux pratiques de «*web scraping*»). Sur la base de ce qui précède, l'EDPB fournit quelques exemples de mesures qui peuvent être mises en œuvre pour atténuer les risques identifiés dans le cadre du test de mise en balance et qui devraient être prises en compte par les AC lors de l'évaluation de modèles d'IA spécifiques au cas par cas.

[76] Lignes directrices 1/2024 relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, point 57.

100. **Mesures techniques:**

a. les mesures mentionnées à la section 3.2.2 qui sont appropriées pour atténuer les risques en jeu, lorsque ces mesures n'entraînent pas l'anonymisation du modèle et ne sont pas nécessaires pour se conformer à d'autres obligations du RGPD ou dans le cadre du test de nécessité (deuxième étape de l'évaluation de l'intérêt légitime).

101. Outre ces mesures, d'autres mesures pertinentes peuvent aussi inclure:

b. Des mesures de pseudonymisation: il peut s'agir par exemple de mesures visant à empêcher toute combinaison de données basées sur des identifiants individuels. Ces mesures peuvent ne pas être appropriées lorsque l'AC estime que le responsable du traitement a démontré la nécessité raisonnable de recueillir différentes données au sujet d'une personne particulière aux fins du développement du système ou du modèle d'IA en question.

c. Mesures visant à masquer les données à caractère personnel ou à les remplacer par de fausses données à caractère personnel dans le jeu d'apprentissage (par exemple, le remplacement des noms et des adresses électroniques par de faux noms et de fausses adresses électroniques). Cette mesure peut être particulièrement appropriée lorsque le contenu substantiel réel des données n'est pas pertinent pour l'ensemble du traitement (par exemple, dans le cadre de l'apprentissage par LLM).

102. **Mesures qui facilitent l'exercice des droits des personnes:**

a. Observation d'un délai raisonnable entre la collecte d'un jeu de données d'entraînement et son utilisation. Cette garantie supplémentaire peut permettre aux personnes concernées d'exercer leurs droits pendant cette période, le délai raisonnable étant apprécié en fonction des circonstances de chaque cas.

b. Proposer un «opt-out» inconditionnel dès le départ, par exemple en prévoyant un droit d'opposition discrétionnaire pour les personnes concernées avant que le traitement n'ait lieu, afin de renforcer le contrôle des personnes sur leurs données, ce qui va au-delà des conditions de l'article 21 du RGPD[77].

c. Permettre aux personnes concernées d'exercer leur droit à l'effacement même lorsque les motifs spécifiques énumérés à l'article 17, paragraphe 1, du RGPD ne s'appliquent pas[78].

d. Permettre aux personnes concernées de présenter des réclamations concernant la régurgitation ou la mémorisation de données à caractère personnel, ainsi que les circonstances et les moyens par lesquels ces réclamations peuvent être reproduites, afin de permettre aux responsables du traitement de reproduire et d'évaluer les techniques de désapprentissage pertinentes pour répondre à ces réclamations.

103. **Mesures de transparence:** dans certains cas, les mesures d'atténuation pourraient inclure des mesures visant à assurer une plus grande transparence en ce qui concerne le développement du modèle d'IA. Certaines mesures, en plus du respect des obligations du RGPD, peuvent aider à surmonter l'asymétrie de l'information et permettre aux personnes concernées de mieux comprendre le traitement impliqué dans la phase de développement:

a. Publication de communications publiques et facilement accessibles qui vont au-delà des informations requises au titre de l'article 13 ou de l'article 14 du RGPD, par exemple en

[77] Ibidem.
[78] Ibidem.

fournissant des détails supplémentaires sur les critères de collecte et tous les jeux de données utilisés, en tenant compte de la protection particulière dont bénéficient les enfants et les personnes vulnérables.

b. Autres formes d'information des personnes concernées, par exemple: campagnes médiatiques auprès de différents médias pour informer les personnes concernées, campagne d'information par courrier électronique, utilisation de la visualisation graphique, questions fréquemment posées, labels de transparence et cartes modèles dont la systématisation pourrait structurer la présentation des informations sur les modèles d'IA, et rapports annuels de transparence sur une base volontaire.

104. **Mesures d'atténuation spécifiques dans le contexte du «*web scraping*» (moissonnage):** Étant donné que, comme indiqué ci-dessus, le «*web scraping*», ou moissonnage, présente des risques spécifiques[79], des mesures d'atténuation spécifiques pourraient être définies dans ce contexte. Le cas échéant, elles peuvent être prises en considération par les AC, en plus des mesures d'atténuation mentionnées plus haut, lorsqu'elles enquêtent sur les responsables du traitement qui procèdent au «*web scraping*».

105. Des mesures spécifiques, lorsqu'elles ne sont pas nécessaires dans le cadre de la deuxième étape de l'évaluation de l'intérêt légitime, peuvent s'avérer utiles pour atténuer le risque dans le contexte du *web scraping*. Il peut s'agir de **mesures techniques**, telles que:

a. l'exclusion du contenu de publications qui pourraient contenir des données à caractère personnel présentant des risques pour des personnes ou des groupes de personnes particuliers (par exemple, des personnes qui pourraient faire l'objet d'abus, de préjugés ou même de dommages physiques si les informations étaient rendues publiques).

b. veiller à ce que certaines catégories de données ne soient pas collectées ou à ce que certaines sources soient exclues de la collecte de données; cela pourrait inclure, par exemple, certains sites web qui sont particulièrement intrusifs en raison de la sensibilité de leur objet.

c. L'exclusion de la collecte à partir de sites web (ou de sections de sites web) qui s'opposent clairement au *web scraping* et à la réutilisation de leur contenu aux fins de la construction de bases de données d'entraînement d'IA (par exemple, en respectant les fichiers robots.txt ou ai.txt ou tout autre mécanisme reconnu pour signifier l'exclusion du *crawling* ou du *scraping* automatisés).

d. Imposer d'autres limites pertinentes à la collecte, en incluant éventuellement des critères basés sur des périodes de temps.

106. Dans le contexte du *web scraping*, les exemples de mesures spécifiques **facilitant l'exercice des droits des personnes et la transparence** peuvent inclure: la création d'une liste d'exclusion, gérée par le responsable du traitement et permettant aux personnes concernées de s'opposer à la collecte de leurs

[79] Ces pratiques peuvent également soulever d'autres questions qui ne sont pas couvertes par le présent avis, voir par exemple Pagallo U., Ciani Sciolla J., *Anatomy of web data scraping: ethics, standards, and the troubles of the law (Anatomie de l'extraction de données sur le web: éthique, normes et problèmes juridiques).* European Journal of Privacy Law & Technologies, (2023) 2 p. 1 - 19, disponible à l'adresse: https://doi.org/10.57230/EJPLT232PS.

données sur certains sites web ou plateformes en ligne en fournissant des informations qui les identifient sur ces sites web, y compris avant que la collecte des données ne soit effectuée[80].

107. **Considérations spécifiques concernant les mesures d'atténuation au cours de la phase de déploiement:** Bien que certaines des mesures susmentionnées puissent également être pertinentes pour la phase de déploiement, en fonction des circonstances, l'EDPB fournit ci-dessous une liste non exhaustive de mesures de soutien supplémentaires qui peuvent être mises en œuvre et qui devraient être évaluées par les AC au cas par cas.

    a. **Des mesures techniques** peuvent, par exemple, être mises en place pour empêcher le stockage, la regurgitation ou la production de données à caractère personnel, en particulier dans le contexte des modèles d'IA générative (tels que les filtres de sortie), et/ou pour atténuer le risque de réutilisation illicite par des modèles d'IA à usage général (par exemple, le repérage numérique des résultats générés par l'IA).

    b. **Mesures qui facilitent ou accélèrent l'exercice des droits des personnes** lors de la phase de déploiement, au-delà de ce qui est exigé par la loi, en ce qui concerne en particulier, et sans s'y limiter, l'exercice du droit à l'effacement des données à caractère personnel à partir des données de sortie du modèle ou de la déduplication, et les techniques de post-formation qui tentent de supprimer ou d'éliminer les données à caractère personnel.

108. Lorsqu'elles enquêtent sur le déploiement d'un modèle d'IA spécifique, les AC devraient examiner si le responsable du traitement a publié le test de mise en balance qu'il a effectué, étant donné que cela peut accroître la transparence et l'équité. Comme indiqué dans les lignes directrices de l'EDPB sur l'intérêt légitime, d'autres mesures peuvent être envisagées pour fournir aux personnes concernées des informations provenant du test de mise en balance avant toute collecte de données à caractère personnel[81]. Le comité[82] rappelle également qu'un élément à prendre en considération est la question de savoir si le responsable du traitement a associé le DPD, le cas échéant.

3.4 Sur l'impact possible d'un traitement illicite dans le cadre du développement d'un modèle d'IA sur la licéité du traitement ou de l'exploitation ultérieurs du modèle d'IA

109. La présente section de l'avis traite de la question 4 de la demande. Cette question vise à obtenir des éclaircissements sur l'incidence éventuelle d'un traitement illicite au cours de la phase de développement sur le traitement ultérieur (par exemple lors de la phase de déploiement du modèle d'IA) ou sur l'exploitation du modèle. La question vise à traiter à la fois la situation dans laquelle un tel modèle d'IA traite des données à caractère personnel qui sont conservées dans le modèle [question 4, point i), de la demande], ainsi que la situation dans laquelle aucun traitement de données à caractère personnel n'intervient plus dans le déploiement du modèle d'IA (c'est-à-dire que le modèle est anonyme) [question 4, point ii) de la demande].

[80] Le responsable du traitement ne traite plus les données à caractère personnel, à moins qu'il ne démontre qu'il existe des motifs légitimes et impérieux pour le traitement qui prévalent sur les intérêts et les droits et libertés de la personne concernée, ou pour la constatation, l'exercice ou la défense de droits en justice.
[81] Lignes directrices 1/2024 du CEPD relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, point 68.
[82] Lignes directrices 1/2024 de l'EDPB relatives au traitement des données à caractère personnel sur la base de l'article 6, paragraphe 1, point f), du RGPD, version 1.0, adoptées le 8 octobre 2024, point 12.

110. Avant d'aborder certains scénarios spécifiques, le comité formule les considérations générales suivantes.

111. Tout d'abord, les éclaircissements fournis dans la présente section porteront sur le traitement de données à caractère personnel effectué au cours de la phase de développement sans respecter le principe de licéité énoncé à l'article 5, paragraphe 1, point a), du RGPD et à l'article 6 du RGPD en particulier (ci-après le «**caractère illicite**»)[83]. Dans le même ordre d'idées, les considérations de l'EDPB se concentreront sur l'impact du caractère illicite du traitement au cours de la phase de développement sur la licéité (c'est-à-dire le respect de l'article 5, paragraphe 1, point a), du RGPD et de l'article 6 du RGPD) du traitement ou de l'exploitation ultérieurs du modèle. Toutefois, l'EDPB souligne que le traitement effectué au cours de la phase de développement peut également entraîner des violations d'autres dispositions du RGPD, telles que le manque de transparence à l'égard des personnes concernées ou la protection des données dès la conception et/ou par défaut, qui ne font pas l'objet d'une analyse dans le présent avis.

112. Deuxièmement, lorsqu'on aborde cette question, le principe de responsabilité, qui exige que les responsables du traitement soient responsables du respect, entre autres, de l'article 5, paragraphe 1, du RGPD et l'article 6 du RGPD[84], et démontrent qu'ils respectent ceux-ci, joue un rôle clé. Il en va de même pour la nécessité d'évaluer quelle organisation est responsable du traitement pour l'activité de traitement en cause et si des situations de responsabilité conjointe du traitement surviennent (étant donné qu'elles peuvent être inextricablement liées)[85]. Compte tenu de l'importance des circonstances factuelles de chaque cas, y compris en ce qui concerne le rôle joué par chaque partie intervenant dans le traitement, les considérations du comité européen de la protection des données devraient être comprises comme des observations générales qui devraient être appréciées au cas par cas par les AC.

113. Troisièmement, l'EDPB souligne que, conformément à l'article 51, paragraphe 1, du RGPD, les AC sont «*chargées de surveiller l'application [du RGPD], afin de protéger les libertés et droits fondamentaux des personnes physiques à l'égard du traitement et de faciliter le libre flux des données à caractère personnel au sein de l'Union*». Il appartient donc aux AC d'évaluer la licéité du traitement et d'exercer les pouvoirs qui leur sont conférés par le RGPD conformément à leur cadre national[86]. Dans de tels cas, les AC disposent d'un pouvoir discrétionnaire pour évaluer la ou les violations éventuelles et choisir des mesures appropriées, nécessaires et proportionnées, parmi celles mentionnées à l'article 58 du RGPD, en tenant compte des circonstances de chaque cas d'espèce[87].

114. **Lorsqu'une infraction est constatée, les AC peuvent imposer des mesures correctives, consistant par exemple à ordonner aux responsables du traitement, en tenant compte des circonstances de chaque cas, de prendre des mesures pour remédier au caractère illicite du traitement initial.** Il peut s'agir, par exemple, d'infliger une amende, d'imposer une limitation temporaire du traitement, d'effacer une partie du jeu de données qui a été traitée de manière illicite ou, lorsque cela n'est pas possible, en

[83] CJUE, arrêt du 4 mai 2023, affaire C-60/22, *Bundesrepublik Deutschland* (ECLI:EU:C:2023:373), points 55-57.
[84] CJUE, arrêt du 4 mai 2023, affaire C-60/22, *Bundesrepublik Deutschland* (ECLI:EU:C:2023:373), point 53.
[85] Lignes directrices 07/2020 de l'EDPB concernant les notions de responsable du traitement et de sous-traitant dans le RGPD, version 2.1, adoptées le 7 juillet 2021, point 55.
[86] Il se peut que des règles nationales spécifiques doivent être prises en compte. Voir par exemple l'article 2-decies du code italien de protection des données (décret législatif 196/2003) qui établit que les données traitées en violation des règles de protection des données ne peuvent pas être utilisées. Ceci est sans préjudice d'autres cadres juridiques nationaux, tels que les lois pénales.
[87] Voir à cet égard le considérant 129 du RGPD, ainsi que CJUE, arrêt du 26 septembre 2024, affaire C-768-21, T*R contre Land Hessen* (ECLI:EU:C:2024:785), point 37; CJUE, arrêt du 7 décembre 2023, dans les affaires jointes C-26/22 et C-64/22, SCHUFA Holding (Libération de reliquat de dette) (ECLI:EU:C:2023:958), point 57; et CJUE, arrêt du 14 mars 2024, affaire C-46/23, *Újpesti Polgármesteri Hivatal* (ECLI:EU:C:2024:239), point 34.

fonction des faits en cause, compte tenu de la proportionnalité de la mesure, d'ordonner l'effacement du jeu de données utilisé pour développer le modèle d'IA et/ou du modèle d'IA lui-même. Lorsqu'elles évaluent la proportionnalité de la mesure envisagée, les AC peuvent tenir compte des mesures qui peuvent être appliquées par le responsable du traitement pour remédier au caractère illicite du traitement initial (par exemple, la reconversion professionnelle).

115. L'EDPB souligne également que, lorsque des données à caractère personnel sont traitées de manière illicite, les personnes concernées peuvent demander la suppression de leurs données à caractère personnel, sous réserve des conditions énoncées à l'article 17 du RGPD, et que les AC peuvent ordonner l'effacement d'office des données à caractère personnel[88].

116. Lorsqu'elles évaluent si une mesure est appropriée, nécessaire et proportionnée, les AC peuvent tenir compte, entre autres éléments, des risques encourus pour les personnes concernées, de la gravité de l'infraction, de la faisabilité technique et financière de la mesure, ainsi que du volume de données à caractère personnel concerné.

117. Enfin, l'EDPB rappelle que les mesures prises par les AC au titre du RGPD sont sans préjudice de celles prises par les autorités compétentes en vertu de la loi sur l'IA et/ou d'autres cadres juridiques applicables (par exemple, la législation en matière de responsabilité civile).

118. Dans les sections suivantes, le comité européen de la protection des données examinera trois scénarios couverts par la question 4 de la demande, dans lesquels les différences résident dans la question de savoir si les données à caractère personnel traitées aux fins de l'élaboration du modèle sont conservées dans le modèle et/ou si le traitement ultérieur est effectué par le même responsable du traitement ou par un autre responsable du traitement.

3.4.1 Scénario 1 Un responsable du traitement traite de manière illicite des données à caractère personnel pour développer le modèle, les données à caractère personnel sont conservées dans le modèle et sont ensuite traitées par le même responsable du traitement (par exemple dans le cadre du déploiement du modèle)

119. Ce scénario se rapporte à la question 4(i) de la demande, dans la situation où un responsable du traitement traite de manière illicite des données à caractère personnel [c'est-à-dire sans respecter l'article 5, paragraphe 1, point a), du RGPD et l'article 6 du RGPD] pour développer un modèle d'IA, le modèle d'IA conserve des informations relatives à une personne physique identifiée ou identifiable et, par conséquent, n'est pas anonyme. Les données à caractère personnel sont ensuite traitées par le même responsable du traitement (par exemple dans le cadre du déploiement du modèle). En ce qui concerne ce scénario, le comité européen de la protection des données présente les considérations suivantes.

120. Le pouvoir de l'autorité de contrôle d'imposer des mesures correctives au traitement initial (comme expliqué aux points 113, 114 et 115 ci-dessus) aurait en principe une incidence sur le traitement ultérieur (par exemple, si l'AC ordonne au responsable du traitement de supprimer les données à caractère personnel qui ont été traitées de manière illicite, de telles mesures correctives ne

[88] À cet égard, l'avis 39/2021 de l'EDPB sur la question de savoir si l'article 58, paragraphe 2, point g), du RGPD pourrait servir de base juridique à une autorité de contrôle pour ordonner d'office l'effacement de données à caractère personnel dans une situation où une telle demande n'a pas été présentée par la personne concernée, point 28. Voir également, à cet égard, CJUE, arrêt du 14 mars 2024, affaire C-46/23, *Újpesti Polgármesteri Hivatal* (ECLI:EU:C:2024:239), point 42.

permettraient pas à ce dernier de traiter ultérieurement les données à caractère personnel faisant l'objet des mesures).

121. En ce qui concerne plus particulièrement l'incidence du traitement illicite au cours de la phase de développement sur le traitement ultérieur (par exemple lors de la phase de déploiement), le CEPD rappelle qu'il appartient aux AC de procéder à une analyse au cas par cas qui tienne compte des circonstances propres à chaque cas.

122. **La question de savoir si les phases de développement et de déploiement impliquent des finalités distinctes (constituant ainsi des activités de traitement distinctes) et dans quelle mesure l'absence de base juridique pour l'activité de traitement initiale a une incidence sur la licéité du traitement ultérieur devrait être évaluée au cas par cas, en fonction du contexte de l'affaire.**

123. Par exemple, en ce qui concerne la base juridique de l'article 6, paragraphe 1, point f), du RGPD, lorsque le traitement ultérieur est fondé sur l'intérêt légitime, le fait que le traitement initial était illicite devrait être pris en compte dans l'évaluation de l'intérêt légitime (par exemple, en ce qui concerne les risques pour les personnes concernées ou le fait que les personnes concernées peuvent ne pas s'attendre à un tel traitement ultérieur). Dans ces cas, le caractère illicite du traitement au cours de la phase de développement peut avoir une incidence sur la licéité du traitement ultérieur.

3.4.2 Scénario 2 Un responsable du traitement traite de manière illicite des données à caractère personnel pour développer le modèle, les données à caractère personnel sont conservées dans le modèle et sont traitées par un autre responsable du traitement dans le cadre du déploiement du modèle

124. Ce scénario se rapporte à la question 4, point i), de la demande. Il diffère du scénario 1 (voir le point 3.4.1 du présent avis) car les données à caractère personnel sont traitées ultérieurement par un autre responsable du traitement dans le cadre du déploiement du modèle d'IA.

125. L'EDPB rappelle que la vérification des rôles assignés à ces différents acteurs dans le cadre de la protection des données est une étape essentielle pour déterminer quelles obligations au titre du RGPD s'appliquent et qui est responsable de ces obligations, et que les situations de responsabilité conjointe devraient également être prises en considération lors de l'évaluation des responsabilités de chaque partie au titre du RGPD. Par conséquent, les observations ci-dessous devraient être considérées comme des éléments généraux devant être pris en considération par les AC, le cas échéant. En ce qui concerne ce scénario 2, le comité européen de la protection des données présente les considérations suivantes.

126. Tout d'abord, il convient de rappeler que, conformément à l'article 5, paragraphe 1, point a), du RGPD, lu à la lumière de l'article 5, paragraphe 2, du RGPD, chaque responsable du traitement doit s'assurer de la licéité du traitement qu'il effectue et être en mesure de la démontrer. Par conséquent, les AC devraient évaluer la licéité du traitement effectué i) par le responsable du traitement qui a initialement développé le modèle d'IA; et ii) par le responsable du traitement qui a acquis le modèle d'IA et traite les données à caractère personnel par lui-même.

127. Deuxièmement, les considérations formulées aux points 113, 114 et 115 ci-dessus sont pertinentes en l'espèce, en ce qui concerne le pouvoir des AC d'intervenir dans le cadre du traitement initial. L'article 17, paragraphe 1, point d), du RGPD (effacement des données traitées de manière illicite) et l'article 19 du RGPD (obligation de notification concernant la rectification ou l'effacement des données à caractère personnel ou la limitation du traitement) peuvent, en fonction des circonstances de l'espèce, être également pertinents dans ce contexte, par exemple en ce qui concerne la notification

que le responsable du traitement qui élabore le modèle doit adresser au responsable du traitement qui déploie le modèle.

128. Troisièmement, en ce qui concerne l'incidence possible du caractère illicite du traitement initial sur le traitement ultérieur effectué par un autre responsable du traitement, une telle appréciation devrait être effectuée par les AC au cas par cas.

129. **Les AC devraient tenir compte du fait que le responsable du traitement qui déploie le modèle a effectué une évaluation appropriée, dans le cadre de ses obligations de responsabilité [89] pour démontrer le respect de l'article 5, paragraphe 1, point a), et de l'article 6, du RGPD afin de s'assurer que le modèle d'IA n'a pas été mis au point en traitant de manière illicite des données à caractère personnel.** Cette évaluation par les AC devrait tenir compte de la question de savoir si le responsable du traitement a évalué certains critères non exhaustifs, tels que la source des données et si le modèle d'IA est le résultat d'une violation du RGPD, en particulier si celle-ci a été déterminée par une AC ou une juridiction, de sorte que le responsable du traitement qui a déployé le modèle ne pouvait ignorer que le traitement initial était illicite.

130. Le responsable du traitement devrait examiner, par exemple, si les données proviennent d'une violation de données à caractère personnel ou si le traitement a fait l'objet d'une constatation de violation par une AC ou une juridiction. **Le degré d'évaluation du responsable du traitement et le niveau de détail attendu par les AC peuvent varier en fonction de divers facteurs, y compris le type et le degré de risques soulevés par le traitement dans le modèle d'IA au cours de son déploiement en ce qui concerne les personnes concernées dont les données ont été utilisées pour développer le modèle.**

131. Le comité européen de la protection des données note que la législation sur l'IA exige des fournisseurs de systèmes d'IA à haut risque qu'ils établissent une déclaration de conformité de l'UE[90], et que cette déclaration contient une déclaration attestant que le système d'IA concerné est conforme à la législation de l'UE en matière de protection des données[91]. L'EDPB note qu'une telle déclaration sur l'honneur peut ne pas constituer une conclusion concluante de conformité au titre du RGPD. Elle peut néanmoins être prise en considération par les AC lors de l'examen d'un modèle d'IA spécifique.

132. Les mêmes considérations formulées au point 123 ci-dessus sont également pertinentes en l'espèce. Lorsque les AC vérifient si et comment le responsable du traitement a évalué le caractère approprié de l'intérêt légitime en tant que base juridique du traitement qu'il effectue, le caractère illicite du traitement initial devrait être pris en considération dans le cadre de l'évaluation de l'intérêt légitime, par exemple en évaluant les risques potentiels susceptibles de survenir pour les personnes concernées dont les données à caractère personnel ont été traitées de manière illicite afin d'élaborer le modèle. Différents aspects, qu'ils soient de nature technique (par exemple, l'existence de filtres ou de limitations d'accès au cours du développement du modèle, que le responsable du traitement ultérieur ne peut contourner ou influencer, et qui pourraient empêcher l'accès à des données à caractère personnel ou la divulgation de ces données), ou de nature juridique (par exemple, la nature et la gravité du caractère illicite du traitement initial), doivent être dûment pris en considération dans le cadre du test de mise en balance.

[89] Article 5, paragraphe 2, et article 24 du RGPD.
[90] Article 16, point g), et article 47 de la législation sur l'IA.
[91] Annexe V, point 5 de la loi sur l'IA.

3.4.3 Scénario 3 Un responsable du traitement traite de manière illicite des données à caractère personnel pour développer le modèle, puis veille à ce que le modèle soit anonymisé, avant que le même responsable du traitement ou un autre responsable du traitement n'entame un autre traitement de données à caractère personnel dans le cadre du déploiement

133. Ce scénario se rapporte à la question 4, point (ii) de la demande et fait référence à un cas où un responsable du traitement traite de manière illicite des données à caractère personnel pour développer le modèle d'IA, mais le fait d'une manière qui garantit que les données à caractère personnel sont anonymisées, avant que le même responsable du traitement ou un autre responsable du traitement n'entame un autre traitement de données à caractère personnel dans le contexte du déploiement. Premièrement, l'EDPB rappelle que les AC sont compétentes et ont le pouvoir d'intervenir en ce qui concerne le traitement lié à l'anonymisation du modèle, ainsi que le traitement effectué au cours de la phase de développement. Par conséquent, les AC peuvent, en fonction des circonstances spécifiques du cas d'espèce, imposer des mesures correctives à ce traitement initial (comme expliqué aux points 113, 114 et 115 ci-dessus).

134. S'il peut être démontré que l'exploitation ultérieure du modèle d'IA n'implique pas le traitement de données à caractère personnel, l'EDPB considère que le RGPD ne s'applique pas[92]. Par conséquent, le caractère illicite du traitement initial ne devrait pas avoir d'incidence sur le fonctionnement ultérieur du modèle. Toutefois, l'EDPB souligne qu'une simple affirmation de l'anonymat du modèle ne suffit pas pour l'exempter de l'application du RGPD, et note que les AC devraient l'évaluer en tenant compte, au cas par cas, des considérations fournies par l'EDPB pour répondre à la question 1 de la demande.

135. **Lorsque les responsables du traitement traitent ultérieurement des données à caractère personnel collectées au cours de la phase de déploiement, après que le modèle a été anonymisé, le RGPD s'appliquerait à ces activités de traitement. Dans ces cas, en ce qui concerne le RGPD, le caractère illicite du traitement initial ne devrait pas avoir d'incidence sur la licéité du traitement effectué au cours de la phase de déploiement.**

4 Observations finales

136. Le présent avis est adressé à toutes les AC et sera publié conformément à l'article 64, paragraphe 5, point b), du RGPD.

Pour le comité européen de la protection des données

La présidente

Anu Talus

[92] Considérant 26 du RGPD.